NEURAL SEMIOTICS

NEURAL SEMIOTICS:

A BIOLOGICAL FOUNDATION FOR BELIEF

By:

ROBYN BLUHM, B.Sc., B.A.

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfilment of the Requirements

For the Degree

Master of Arts

McMaster University

MASTER OF ARTS (2000)
(Philosophy)

McMaster University
Hamilton, Ontario

TITLE: Neural Semiotics: A Biological Foundation for Belief

AUTHOR: Robyn Bluhm, B.Sc. B.A. (Laurentian University)

SUPERVISOR: Professor Barry Allen

NUMBER OF PAGES: iv, 91

INTRODUCTION:

We are all folk psychologists. Every time we explain our behavior by describing our beliefs, desires, fears and expectations, or predict what someone else will do based on the beliefs we think they have, we are practicing folk psychology.

There are also other ways of predicting and explaining, behavior and these may be in conflict with folk psychological explanations. Psychologists and neuroscientists draw on a wealth of experimental and clinical evidence to provide these alternative explanations. Some philosophers, including Paul Churchland, have suggested that the philosophy of mind should accept these latter scientific explanations rather than those of commonsense psychology. Furthermore, since the beliefs and desires invoked in folk psychological explanations are unlikely to be reducible to the theoretical entities of neuropsychology, we will ultimately be able to dispense with beliefs altogether. This position is known as eliminative materialism.

Other philosophers believe that there is room for folk psychology in the philosophy of mind. Jerry Fodor, for example, suggests that the categories found in folk psychology are similar to those of scientific psychology. Like Churchland, he acknowledges that psychological explanations are unlikely to reduce smoothly to physiological ones but he claims that this irreducibility is evidence that psychology is an autonomous discipline.

Like Fodor, Lynne Rudder Baker believes that folk psychology is a legitimate explanation of behavior; it is a social practice and is used to explain other social

practices. Because Baker emphasizes the interpersonal aspect of folk psychology, she claims that it is best explained through an analysis, not of brain or of mind, but of language.

In this thesis, I examine these three approaches to folk psychology. I argue that each of them can provide only a limited explanation of folk psychology because each author is committed to a particular, limited methodology. Instead, I suggest that Terrence Deacon's multidisciplinary theory of brain-language co-evolution provides a better framework from which to examine our beliefs and desires.

Deacon claims that language is the outward expression of a cognitive ability that is unique to human beings, the ability to use symbolic reference. He appropriates C.S. Peirce's semiotic categories of icon, index and symbol to show the qualitative difference between human cognition and that of other animals. Our evolution as "the symbolic species" means that we are able to think in abstract terms; human behavior is mediated in large part by this competence with symbols.

While Deacon does not address the issue of folk psychology, I argue that his theory is directly applicable to the question of the reality of beliefs. Churchland insists that folk psychological descriptions of behavior must be justified as theoretical descriptions. I believe that a philosophy of mind must explain our ability to use talk of beliefs and desires to make sense of human behavior. Deacon's approach allows us to explain folk psychology in terms of the biology that makes it possible for us to have, and discuss, beliefs and desires. At the same time, it leaves space for philosophical theories of mind that do not need to be reduced to biological theories.

CHAPTER 1:

The status of folk psychology is contentious in the philosophy of mind. Our everyday talk is full of explanations of behaviour in terms of propositional attitudes, but the philosophical justification of this practice is problematic. Some argue that folk psychology is a theory, and that it is a good theory at that. Others say that it is such a bad theory that it will inevitably be replaced by a more adequate conceptual framework. Both of these camps view folk psychology as a science, or proto-science of the mind. A third option is to assert that folk psychology is not a theory at all, but a social practice that gives us valuable information about the people we encounter, and about ourselves. These theorists claim that the mind (or the brain) is not the best place to look for an understanding of propositional attitudes; rather we should look outward to the social practices that are given meaning by terms such as "belief" and "desire".

In this chapter, I will examine the views of Jerry Fodor, Paul Churchland and Lynne Rudder Baker. These authors disagree about they way in which folk psychology ought to be examined and about the kind of results this examination will offer. It is hard to separate the choice of methodology from the results predicted, and it is also difficult to see whether the understanding of the problem directs the methodology used, or the methodology of choice dictates the scope of the problem. By making explicit the link between methodology and conclusions, I will argue that each philosopher has grasped only a part of the underlying problem. Like the parable of the blind men and the

elephant, the full story of folk psychology must incorporate elements of all three approaches.

Fodor and Functionalism

It has been suggested that, for Jerry Fodor, there is no real distinction between psychology and philosophy of mind. Fodor's approach is functional: a good theory will posit whatever processes necessary to explain our behaviour. He suggests that, at least in the initial stages of the development of a theory, psychologists should adopt an "ontological agnosticism" as to the existence of psychological entities (drives, motivations, etc.), similar to the attitude of early geneticists toward the actual nature of genes (1968, p. 13).

There are, however, limits to Fodor's functionalism. He sets two constraints on the autonomy of psychological explanation. First, Fodor is committed to an approach to psychology that reflects our common experiences and everyday questions about these experiences. In fact, "[i]t is just such questions that psychological theories are, in the first instance, constructed in order to answer" (1968. p. xxi). Following from this commitment, Fodor also claims that it is desirable to use psychological terms that are continuous with everyday language. In his view, the terminology of folk psychology should be an important part of psychological explanation. Fodor seems to be suggesting that a psychological theory should provide us with explanations of our behaviour that are more powerful and accurate than our everyday explanations, and that make reference to

the causal structures underlying behaviour. However, these explanations should be intimately connected with "folk" explanations.

There are limits to this limited functionalism. Fodor claims that, insofar as psychological explanations are scientific, "mental state terms are explicitly taken to designate theoretical entities, and the explanations in which such terms figure are explicitly subject to conditions of simplicity and rigor" (1968. p. 93). Related to this view of mental state terms is Fodor's conception of the scope of psychological explanation. "Insofar as it seeks to account for behaviour, a psychological theory may be thought of as a function that maps an infinite set of possible inputs to an organism onto an infinite set of possible outputs" (1968. p. 29). Fodor adopts a position of "methodological solipsism": the input and output are considered solely with respect to the internal states of the organism. The relation between these states and the world is not part of the domain of psychological explanation. "Methodological solipsism is the doctrine that psychological states are individuated *without respect to their semantic evaluation*"; it is not concerned with "how the state corresponds to the world" (1987, p. 42). Psychological explanation differs from common sense explanation with regard to (1) the role of mental state terms and (2) the adoption of methodological solipsism. The two kinds of explanation do (or should), however, share a vocabulary and a common view of the questions that should be asked in an explanation of human behaviour.

The autonomy of psychology is limited, for Fodor, on another front as well. Psychological theories must conform to what is biologically - especially

neurobiologically - possible. If we are to avoid lapsing into Cartesian dualism, psychological explanations must pay due regard to the way in which the brain works. While, initially, psychologists must be free to posit whatever psychological entities account for their data, these entities must eventually be tested against biochemical and neurological explanations. This second phase of psychological explanation is concerned with "the problem of whether the nervous system has subsystems whose functional characteristics correspond with those of antecedently plausible psychological theories" (1968, p. 110).

Despite this limitation on psychology's functionalist theories, Fodor is *not* claiming that a psychological theory must ultimately be reducible to a neurological level of explanation. In fact, he argues extensively against the necessity, and the desirability, of reducing psychology to physiology. Rather, the two levels of explanation are complementary, and develop in a mutually reinforcing fashion. "To say that the goals of physiological psychology are set by the attempt to find mechanisms that correspond to certain functions is to say that it is the psychological theory that articulates these functions that determines the principle of individuation for neurological mechanisms" (ibid.). Without a well-developed psychological theory, we would have no way of understanding what individual brain states, or classes of brain states, meant. Like psychology, neurology depends on a common-sense basis that provides at least an initial description of the events to be explained.

Each of the three kinds of theory - folk, functional and physiological - are concerned with the explanation of the same phenomena, however, each has its own methodology and internal structure. Although the two scientific theories must both conform to a more rigorous structure than common sense explanation, the two theories are "equal" in a sense not accepted by those committed to reductionism. "Every science implies a taxonomy of the events in its universe of discourse. In particular, every science employs a descriptive vocabulary of theoretical and observation predicates such that the events fall under the laws of the science by virtue of satisfying those predicates. Patently, not every true description of an event is a description in such a vocabulary" (1975, pp. 13-14). Fodor accepts that, for the most part, psychology and neurology have overlapping "universes of discourse". He does not feel, though, that this leads automatically to the claim that "for every psychological kind predicate there is a coextensive neurological kind predicate, and the generalization which states this conclusion is a law" (ibid. p. 17). "There are no firm data for any but the grossest correspondence between types of psychological states and types of neurological states, and it is entirely possible that the nervous system of higher organisms characteristically achieves a given psychological end by a wide variety of neurological means" (ibid.).

Fodor's argument against reductionism, then, is directly related to his functionalist view of mental states. Both of these factors, in turn, are influenced by Fodor's insistence that psychology should be rooted in common sense questions about our experiences. Any theory of mind must reflect our mental lives, including the experience of having beliefs, desires and other propositional attitudes. The influence of

his views on the scope and function of psychology is obvious throughout Fodor's description of the language of thought.

Fodor suggests that his description of the language of thought is "not a theory but a theory schema" (1975, p. 29). He does not set out to relate particular inputs to particular outputs, because "a serious theory of the way an organism behaves presupposes extensive information about what the organism knows and values" (ibid.). Instead, Fodor argues for a "representational system of some complexity in which mental processes are carried out" (ibid. p. 29), and outlines the structure that such a system must have.

According to Fodor, "[t]he only psychological models of cognitive processes that seem even remotely plausible represent such processes as computational" (ibid., p. 27). Computation, however, requires a medium – in this case a representational system. In order to account for the variety of behaviours that can be produced by an organism, as well as the ability of an organism to respond to novel stimuli, the representational system must be of a certain level of complexity. For Fodor, nothing less than the complexity of a language would be sufficient; computation requires a language of thought that "must share a number of the characteristic features of real languages" (ibid., p. 31), including the ability to manipulate an infinity of distinct representations (ibid.), the use of formulae that exhibit "such familiar semantic properties as truth and reference" (ibid., p. 32), and a mechanism for expressing intentional properties "that can distinguish between possible, nonactual states of affairs" (ibid.).

Fodor acknowledges that his model is "highly idealized. We do not always contemplate each (or, indeed, any) of the behavioural options we believe to be available to us in a given situation" (ibid., p. 29). However, the mechanisms involved in computation remain the same whether or not the organism makes each representation in the computational process explicit. In fact, this kind of computation underlies "not only considered action, but also learning and perception" (ibid., p. 34). For Fodor, the language of thought goes all the way down.

Intrinsic to Fodor's description of the language of thought is a theory of the structure of propositional attitudes and their role in cognition. Specifically, "having a propositional attitude is being in some *computational* relation to an internal representation" (ibid., p. 198). This relation is further described in a 1984 article "Why There Still has to be a Language of Thought". Here, Fodor notes that "LOT wants to construe propositional attitude tokens as relations to symbol tokens" (p. 136). Fodor uses the metaphor of a box to describe the difference between various attitudes toward the same proposition; the proposition itself does not change, but it can be placed into different boxes, and depending on its placement, it will exert different effects on behaviour. "So, for example, suppose I intend to raise my left hand (I intend to make true the proposition that I raise my left hand). Then what I do is, I put into my intention box a token of a mental symbol that means 'I raise my left hand.' And then, after suitable churning and gurgling and computing and causing, my left hand goes up" (ibid., p. 136). However, the LOT story also insists that "what I put into the intention box has to be

something like a *sentence*, in the present case it has to be a formula which contains, inter alia, an expression that denotes me and an expression that denotes my left hand" (p. 137).

The language of thought hypothesis, then, requires that mental states "have a *combinatorial* semantics: the kind of semantics in which there are (relatively) complex expressions whose content is determined, in some regular way, by the content of their (relatively) simple parts" (ibid., p. 138). The notion of combinatorial semantics, for Fodor, is equivalent to that of syntax. "To a metaphorical real first approximation, we can think of the syntactic structure of a symbol as an abstract feature of its shape. Because, to all intents and purposes, syntax reduces to shape, and because the shape of a symbol is a potential determinant of its causal role, it is fairly easy to see how there could be environments in which the causal role of a symbol correlates with its syntax. It's easy, that is to say, to imagine symbol tokens interacting causally *in virtue of* their syntactic structures" (1984, p. 18).

This remark suggests that, for Fodor, syntax is the primary determinant of the causal powers of mental token, i.e. of propositions in the language of thought. His methodological solipsism is a necessary assumption of this view. Only a psychology that minimizes the content and the meaningfulness of "inputs" allows Fodor to emphasize the process of computation to that great an extent.

This assumption also leads Fodor to suggest that there is little difference between human and animal cognition; he begins his book Psychosemantics with an anecdote about his "strikingly intelligent" cat, who "has, and acts out of, beliefs and desires" (p. x).

Arising from this depiction is the suggestion that natural languages do not make much difference to our cognitive processes. "Attitudes to propositions are...reduced to attitudes to formulae, though the formulae are couched in a proprietary inner code" (1975, p. 198). From this, as Dennett points out, it follows that learning a predicate in a natural language allows one to "augment one's inner code with a synonym, as it were, of the natural language predicate and henceforth use this non-native inner word as an abbreviation for the cumbersome truth-functional molecule of native mentalese" (Dennett, 1987, p. 93).

The benefits of (natural) language use are merely those of efficiency of expression, and communication of the cognitive processes occurring in the language of thought. Because natural language use is intimately tied to the language of thought, however, we are able to accurately communicate what is going on in our heads. When we explain behaviour (our own or someone else's), we are expressing a possibly simplified version of our own cognitive processes. Even though "[o]n even the most optimistic estimate, it's a long way from the intuitive belief/desire explanations that common sense gives us to the rigorous and explicit intentional psychology that is our scientific goal" (1987, p. xii), Fodor would say that folk psychology does give the gist of cognition. Moreover, it does a much better job of explanation and prediction than many sciences. "If you want to know where my physical body will be next Thursday, mechanics – our best science of middle-sized objects after all, and reputed to be pretty good in its field – *is no use to you at all*. Far the best way to find out (usually, in practice, the *only* way to find out) is: *ask me*!" (ibid., p. 6).

Despite its utility, folk psychology is limited in the type of explanations it is able to provide, and this is something that Fodor readily admits. While propositional attitudes are explained, according to Fodor's theory, as the outcome of a computational process, it is possible that even some of those attitudes are not the result of symbolic computation. "Sometimes...such states may be appropriately represented as the causal consequences of subterranean processes of inference" (1975, p. 200). Nor can we explain " 'creative' things like writing poems, discovering laws, or generally, having good ideas" (ibid. p. 201). These states and behaviours, though, are not part of the purview of cognitive psychology. Fodor describes cognitive psychology – and, thus, the language of thought – as being "about how rationality is structured, viz. how mental states are contingent upon each other" (ibid., p. 202). He thus admits the possibility that "highly valued mental states are sometimes the effects of (literally) nonrational causes" (p. 202). Ultimately, for Fodor, these limitations do not prevent either folk- or cognitive psychology from being the best explanation we have for their domain of interest.

Churchland's Eliminative Materialism

Paul Churchland is among those for whom "nonreductive psychology" as a theory is "disappointing in the modesty of its ambitions" (Fodor, 1975, p. 203). As with Fodor, Churchland's opinion as to the validity of folk psychology is due as much to his preferred methodology as to his theory of mind. While Fodor insists that cognitive psychology provides a level of explanation that is related to , but autonomous from, physiology, Churchland requires that a valid psychological theory be ultimately reducible to

neurological terms. Our theory of brain states will ultimately determine the kinds of mental state terms we are willing to recognize. The theoretical terms of folk psychology (the beliefs, desires, and other propositional attitudes) are unlikely to correspond to what is actually going on in the brain, and so will ultimately be eliminated from the vocabulary of a mature theory of mind.

In his 1984 book Matter and Consciousness, Churchland outlines the distinctions between the kind of "cognitive/computational approach" favoured by Fodor, and the methodological materialism that he prefers. He describes the former as "the 'top-down approach', because one starts with our current understanding of what intelligent creatures do and then asks what sort of underlying operations could possibly produce or account for such cognitive activities" (p. 96). Churchland espouses the "bottom-up" approach, which takes as its starting point the idea that, since "cognitive activities are ultimately just activities of the nervous system" (ibid.), these activities can best be understood by studying the nervous system itself. Of course, Churchland acknowledges that the complexity of the nervous system "defeats any ready understanding, and we have only just begun to unravel it" (ibid., p. 97). Even the most optimistic advocates of this approach recognize that it will be a long time before we can understand complex behaviours, such as cognition, in purely neuroscientific terms. Ultimately, however, understanding "the behavior of neurons, and especially systems of neurons" will set us "on our way toward understanding everything there is to know about natural intelligence" (p. 97).

Churchland recognizes, of course, that we use the terms of folk psychology to explain our behaviour, but claims that the utility of beliefs and desires on a common sense level does not guarantee them a place in a neuroscientific theory of mind. "If the thumb-worn categories of folk psychology (belief, desire, consciousness, and so on) really do possess objective integrity, then the bottom-up approach will eventually lead us back to them. And if they do not, then the bottom-up approach, being so closely tied to the empirical brain, offers the best hope for constructing a new and more adequate set of concepts with which to understand our inner life" (ibid., p. 97). The status of folk psychology, then, is an empirical question.

Yet Churchland makes more than an empirical claim; he argues for eliminative materialism in ontological terms as well (ibid. pp. 43-49). From this point of view, eliminative materialism becomes a response to the mind-body problem. Our common sense view of mind is, simply, "a false and radically misleading conception of the causes of human behavior and the nature of cognitive activity" (ibid., p. 43, original in italics). Ultimately, this framework "will simply be eliminated, rather than be reduced, by a matured neuroscience" (ibid.).

Churchland gives three reasons why folk psychology must ultimately be eliminated. In the first place, the scope of folk psychology is severely limited; it cannot explain, for example, sleep, memory, or mental illness (ibid., pp. 45-6). Of course, it must be said in defense of folk psychology that it does not attempt to do so. Fodor, too, acknowledges the limited domain of folk psychology, and of its scientific relative,

cognitive psychology. Churchland, however, insists that folk psychology, as a theory of mind, should address even such patently non-cognitive behaviours as sleep.

The second argument draws an analogy between the future elimination of folk psychology and the history of other "folk" theories. Churchland points out that traditional, pre-scientific theories of motion, astronomy and the vital essence of life were all eventually eliminated in favour of more sophisticated theories. Given these past examples, he suggests, "it would be a *miracle* if we had got [psychology] right the first time, when we fell down so badly on all the others" (ibid., p. 46).

Finally, Churchland offers an argument that "attempts to find an *a priori* advantage for eliminative materialism over identity theory and functionalism" (ibid.). While all three theories share a materialist view of the mind, the latter two types of theory expect that "the concepts of folk psychology will find vindicating match-ups in a matured neuroscience" (ibid., p. 47). Against this expectation, the eliminativist points out that "the requirements on a reduction are rather demanding" (ibid.), with the new theory entailing principles and concepts that closely mirror the theory to be reduced. "And the fact is, there are vastly many more ways of being an explanatorily successful neuroscience while *not* mirroring the structure of folk psychology, than there are ways of being an explanatorily successful neuroscience while also *mirroring* the very specific structure of folk psychology" (ibid.). Despite our common intuitions to the contrary, eliminative materialism is thus *a priori* more probable than either identity theories or functionalist accounts of mental states. Churchland then provides a caveat, noting that

the advantage of eliminative materialism would, in fact, be reduced if there were independent grounds on which to believe that folk psychology is true. However, he claims, his two previous arguments have already weakened the theoretical status of folk psychology.

All three of these arguments are interrelated, and all rely on Churchland's views on the nature and desirability of intertheoretic reduction. Some unpacking of these three arguments, combined with Churchland's own statements on reduction, will show the basic principles upon which the case for eliminative materialism is built.

The push toward reductionism originates with the assumptions of a materialist position; there is nothing in the world but matter and its relations. If we accept "the purely physical origins and ostensibly physical constitution of each individual human" (ibid., p. 27), then it is relatively easy to claim "the neural dependence of all known mental phenomena" (ibid., p. 28). If mental states are nothing more than brain states, then it will be possible, once an adequate vocabulary of brain states is achieved, to translate our descriptions of mental states into corresponding brain-state terms. This description of reduction, however, uses the top-down approach, in which we begin with a framework that is used to describe mental states, and move toward a new framework constructed at a lower level of explanation. This view is generally sympathetic to the mental level of description, and aims to discover how our mental terms are cashed out at a neuronal level. However, because Churchland uses an exclusively bottom-up approach, brain states become, in a sense, the primary data of a theory of mind. The theoretical

entities of both folk psychology and scientific psychology have no value except as placeholders for - as yet undiscovered – theoretical entities of neuroscience.

Unlike Churchland, Fodor draws a distinction between the two concepts of "the unity of science" and "the generality of physics". He claims that we can accept the latter (which is "roughly, the view that all events which fall under the laws of any science are physical events and hence fall under the laws of physics" (Fodor, 1975, p. 10), without having to accept the former. Fodor emphasizes that "reductionism is an empirical doctrine" (ibid.), but notes that "it is intended to play a regulative role in scientific practice. Reducibility to physics is taken to be a *constraint* upon the acceptability of theories in the special sciences, with the curious consequence that the more the special sciences succeed, the more they ought to disappear" (ibid.).

Churchland, by contrast, accepts both "the generality of physics" and the reductionist principle implied by "the unity of science". He describes functionalists as tending to be strongly antireductionist (1989, p. 46). "They deny that there can be any general characterization of what makes something a *thinker* that is expressible in the language of any of the physical sciences" (ibid.). Functionalists consider the type identity of mental states to be due to their causal relations (to input, other mental states, and output), rather than to a shared underlying physical mechanism. Causal relations stipulate "what is to be a belief, a desire, a pain and so forth" (ibid., p. 45). Because of this, it is all too easy for a functionalist "to ignore or devalue the bottom-up approach" and end up with a methodology that is "dangerously conservative and one-sided" (ibid.,

p. 46). The question that Churchland does not seem to want to consider is whether staunch adherence to the bottom-up approach might not be dangerously radical and one-sided.

Churchland's arguments for eliminative materialism appear to be making what *should* be an empirical question into a methodological point of pride. The first argument – that folk psychology is severely limited in its explanatory domain – is motivated by the "unity of science" hypothesis. The second, that folk theories have been historically inaccurate, ignores the power and utility of folk psychology in the explanation of rational behavior and implies, again, that a successful theory is one that will be found to reduce to a lower level of explanation. The third argument claims an *a priori* advantage of eliminative materialism over identity theories or functionalist views. He notes, however, that this argument rests, at least partly, on the power of his first two arguments to weaken the "presumption in favor of the truth of folk psychology" (1984, p. 47). If we have independent reasons to believe that folk psychology is true, the *a priori* advantage of eliminitivism is weakened. Ultimately, then, Churchland's eliminative materialism must be seen as a prediction that arises from his bottom-up approach to theory construction, and from his interest in neuroscientific explanations of behaviour.

For Churchland, the place to begin constructing a theory of mind is at the level of "neurons and systems of neurons". Of course, it is a long way from understanding the behaviour of neurons to understanding the behaviour of intelligent organisms, and Churchland acknowledges this fact. However, he stresses that we can learn much about

normal behaviour from clinical cases, from "the victims of chemical, physical, or degenerative abnormalities" (1984, p. 97). More importantly, for the bottom-up perspective, we can extrapolate from our extensive knowledge of the nervous systems of simple creatures to the functioning of more complex animals. The basic physiology and biochemistry is similar in all animals, and this similarity provides clues to the cellular mechanisms that underlie complex mental phenomena. These mechanisms can also be studied using artificial "cognitive" systems. Computer simulations of cognitive activity can also allow us to theorize about the way cognition occurs in the brain. According to Churchland, the best examples of computer simulation of cognition are those provided by connectionist or parallel distributed processing (PDP) models. I will outline both the neurophysiological and the computer theories, and show how the similarities between these approaches give Churchland a model of cognition that is radically different from Fodor's language of thought.

The best way to understand the activity of the nervous system, according to the bottom-up approach, is to begin with "the nervous system itself, to discover the structure and behavior of the tiniest elements, their interconnections and interactivity, their development over time, and their collective control of behavior" (1984, p. 96). Intelligent behaviour is the result of the interconnections among neurons, and the way in which these connections change over time. This study is the domain of neurophysiology.

A neuron consists of three basic segments. The soma, or cell body, contains the metabolic machinery of the cell, including the cell's genome. The dendrites are

branching processes that receive most of the cell's input from other neurons, while the single axon is responsible for the neuron's output. At its distal end, the axon branches and forms multiple points of contact (or synapses) with the dendrites, soma, or axons of other brain cells.[1] The synapse allows a transfer of chemicals from the axon of the first neuron to the membrane of the second. This results in metabolic and/or microstructural changes to receptive, or postsynaptic, neurons. These changes may be only temporary, though repeated synaptic contact between neurons *permanently* changes the structure of the cells. Although they are not physically joined, the cells become connected more strongly, in that the firing of the first cell is more likely to lead to a response from the second.

Thus the behaviour of single neurons is related to that of systems of neurons; patterns of synaptic firing create links among cells. Churchland outlines the work done in the discipline of cognitive neurobiology, which explores how particular patterns of neural behaviour are related to particular kinds of cognitive behaviour. He describes the two central questions of cognitive neurobiology as: "How does the brain *represent* the world? And how does the brain perform *computations* over these representations?" (ibid., p. 146). These questions are similar to those posed by Fodor, but Churchland's focus on the brain rather than the mind leads him to a very different set of answers.

The representation of sensory stimuli by the brain can be described as the creation of "a pattern of stimulation across the [various] kinds of sensory cells" (ibid., p. 147).

[1] In the case of motor neurons, the axon actually synapses onto a muscle cell. The other two classes of neuron, sensory (which receive information from specialized receptor cells at the periphery of the body)

For each kind of sensation, there are different modalities of receptors: the sense of taste has four, colour has three, and so on. The relative activation of each type of receptor is combined into a "sensory coding vector" with a discrete magnitude for each of the kinds of receptor. Thus, in the case of taste, "it is a pattern of spiking frequencies across the four neural channels that convey news of these activity levels away from the mouth and to the rest of the brain"(ibid.). This can be mathematically represented as a four dimensional array in a "taste-space". Churchland notes that "[w]hat is interesting is that subjectively similar tastes turn out to have very similar coding vectors" (p. 148). Similar descriptions hold for the behavior of the motor neurons; a motor vector is "a set of simultaneous activity levels in all of the motor neurons, neurons that convey messages from the brain to the body's musculature" (ibid., p. 151). The chain of neural activity from stimulus to response can be modeled by a set of mathematical transformations from vector to vector.

The next question Churchland faces is, of course, whether or not cognitive behaviour, or higher level computation on these sorts of representations, follows the same rule. So far, the answer seems likely to be yes, since the physiology of sensory and motor neurons and interneurons is in many respects the same. Not only are stimulation vectors "a beautifully effective means of representing things as various as taste, faces, and complex limb positions"; they also solve the problem of how high-speed computing can take place "so that the inputs [to the system] are in some way *guiding* or *producing* the outputs" (ibid.) The problem that arises when we try to explain how the brain performs

---

and interneurons (which receive information from other neurons) synapse onto other neurons.

high level cognitive activities is that we do not know enough yet about which parts of the brain are responsible for, for example, the recognition of objects, the understanding of spoken or written language, or the way we classify the world into categories.

The second major source of empirical support for Churchland's theory of mind can do exactly these kinds of tasks. Connectionist, or PDP networks use the same sorts of vector coding, and vector-to-vector transformations, to model various kinds of cognitive behaviour. Churchland discusses an example of this kind of system at length in a 1989 article, and shows how connectionist networks can give some insight into "the functional opacity of the biological brain" (1989, p. 257).

Like real neurons, the neurons of a connectionist system are connected in networks. "In the brain, neurons frequently constitute a population all of which send their axons to the site of a second population of neurons, where each arriving axon divides into terminal end branches in order to make many different synaptic connections within the target population" (ibid., p. 259). A similar architecture is followed by connectionist systems. The initial layer, or input layer, consists of a varying number of neurons, depending on the system, each of which are connected to all of the neurons in the next layer (the "hidden units"), which correspond to the interneurons of the brain. Depending on the system, the number of units in the hidden layer, and the number of hidden layers may vary. If there is more than one hidden layer, all of the units of the second layer project to the third, and so on. The final layer of output units corresponds roughly to the motor neurons of the brain. These networks work, on a mathematical

level, in a fashion similar to the brain. "The assembled set of simultaneous activation levels in all of the input units is the network's *representation* of the input stimulus" (ibid., p. 260). Once the input has been coded, the strength of the connections between the units is altered to create a mathematical relationship between the input and the output.

Initially, each of the connections made by the units of the input layer is assigned an arbitrary connection weight, which determines the way in which the activation vector of the input units affect the various hidden units of the next layer. Next, "[t]he activation vector at the hidden layer is propagated upwards to the output (topmost) layer of units, where an *output vector* is produced" (ibid., p. 261). This output vector is the result of the activation vector of the (topmost) hidden layer, and the value of the connection weights of these hidden units. It is at the output layer that the output of the system is coded. An example will make this process clearer.

Churchland illustrates the functioning of a connectionist system by describing a sonar system designed by Gorman and Sejnowski (1988) that "learned" to distinguish "between the sonar echoes returned from explosive mines, such as might lie on the bottom of sensitive waterways during wartime, and the sonar echoes returned from rocks of comparable size that dot the same underwater landscape" (Churchland, 1989, p. 262). This particular network is quite complex; it has thirteen units at the input layer. "An example of a sonar echo is run through a frequency analyzer and is sampled for its relative energy levels at thirteen frequencies" (ibid., pp. 262-3). Each of these values is then assigned to a single input neuron. These units then activate the units at the hidden

level, to which they are linked by a randomly weighted connection, and these, in turn, excite or inhibit the output neurons. Because the purpose of the network is to give one of two possible responses (corresponding to "mine" or "rock"), there are only two output units.

The next stage is to train the network. This is done by procuring "a large set of recorded samples of various (genuine) mine echoes, from mines of various sizes and orientations, and a comparable set of genuine rock echoes" (p. 264). The actual output of the system for each echo is compared with the desired (correct) output, and corrected by adjusting the weights on the connections between each pair of units. "The idea is to identify those weights most responsible for the error, and then to nudge their values in a direction that would at least reduce the amount by which the output vector is in error" (ibid.) After having been given several thousand presentations of echoes, the program was able to reliably distinguish between rocks and mines, and was able to generalize its "knowledge" to new samples.

Of course, this laborious procedure does not resemble normal human learning, and Churchland admits this limitation of the model. (In this case, however, the computer eventually outperformed human sonar operators (ibid., pp. 255-6)). The architecture of the system, too, is in several important ways distinct from human neuronal architecture: connections from a single connectionist unit can be either excitatory or inhibitory; "real axons have terminal end bulbs that are uniformly inhibitory or uniformly excitatory, depending on the type of neuron...nor do we find connections changing their sign during

learning, as is the case in the model" (ibid., p. 281). The brain also contains many "horizontal" connections among the neurons of a single layer, while the computer models are strictly hierarchical. And, of course, the modification of the connections among the neurons in a connectionist system is highly artificial: there is nothing in the brain that "knows" the correct answer to problems and changes neuronal connections accordingly.

The point of these systems, for Churchland, is that they are able to make discriminations *without* conceptualization. The system described above had no notion of mines or of rocks; its distinction between the two classes of echoes was purely mathematical. This suggests that "it is activity vectors that form the most important kind of representation within the brain. And it is vector-to-vector transformations that form the most important kind of computation" (1984, p. 165). Computation, then, does not occur over symbols, as Fodor suggests, but over a lower level of representation. This suggests that "[t]he elements of cognition have a character unfamiliar to common sense", and that the concepts of folk psychology "need not capture the dynamically significant states and activities of the mind" (ibid.). The behaviour of connectionist systems suggests to Churchland that human behaviour can also be reduced to vector-to-vector transformation.

Folk Psychology as Social Practice

Other theorists claim, however, that beliefs are *a priori* irreducible to brain states, regardless of whether brain states turn out to be vector-coded or operations on symbols. According to such theorists as Lynne Rudder Baker, the proper place to look for an

explanation of belief is not in the mind or the brain of the subject, but in the way in which these beliefs cause the subject to behave. Folk psychology, then, is an account of our social practices rather than our mental states.

Baker defends propositional attitudes from what she describes as "the Standard View", which encompasses both Fodor's and Churchland's theories. The Standard View claims that "the attitudes, if there are any, are (or are constituted by, or are realized in) particular brain states" (Baker, 1995, p. 5). She describes her own contrasting view as "Practical Realism", which sees a belief as "a global state of a whole person, not of any proper part of the person, such as the brain" (ibid., p. 153). While "[h]aving certain neural states is, presumably, necessary for people to have beliefs, it does not follow that for a person to have a particular belief, there is a neural state that constitutes that belief" (ibid., pp. 153-4).

This description of mental states suggests that Rudder Baker is the philosophical heir of Wittgenstein and Ryle, though she herself does not stress these links. One philosopher who holds a similar view to that of Baker, and is explicit about his methodological heritage, is Elmer Sprague. I will briefly describe Sprague's account of what "philosophy of mind would be like if Wittgenstein and Ryle were taken seriously" (Sprague, 1999, p. ix), in order to make evident the account of beliefs that results from Baker's approach.

Sprague describes his own approach as "personism". Personists "refuse to regard the mind as any kind of object", but rather "find the mind in people's capacities to say

and do many different things, and in the way in which people conduct their saying and doing" (ibid.). Sprague notes the influences of Wittgenstein on personism. The meaning of a term is shown by making clear its use in the context of the language game in which it appears. While Sprague is most concerned to make this point with regard to the term "mind", the same approach is also taken to "belief", "desire" and other terms of folk psychology. A second point of Wittgenstein's is also relevant to both personism and Practical Realism: the description of language games is not a causal explanation, and it is this description that is the proper topic of philosophical theory of mind. "In that area philosophy cannot be replaced by a natural science of hypotheses and hypothesis testing" (ibid., p. 45).

Ryle's philosophy of mind is also an important influence on Baker's Practical Realism as it is on Sprague's personism. Sprague outlines Ryle's thesis that "everything philosophers want to say by speaking of minds is only truly said by speaking of persons" (ibid., p. 67). This is reflected in Ryle's distinction between dispositional and causal explanations. The former "explains people's doings by citing their knowing how to do something coupled with a motive for doing it" (ibid., p. 69). In this way Ryle, like Wittgenstein, "puts the mind 'outside', in the publicly observable doings of people as they tie knots, read aloud, and move pencils in making calculations" (ibid., p.91). The terms of folk psychology, then, name "neither an event nor an action, but a disposition that issues in a range of appropriate doings" (ibid., p. 70). The similarities of this account with the kind of analysis offered by Practical Realism will become evident below.

Practical Realism, like Sprague's personism, emphasizes the role of social activity in theories of mind. While Sprague consistently emphasizes the agency of persons who participate in diverse language games, Baker suggests that folk psychology is the social practice by which we explain social practices.

Unlike Sprague (or, for that matter, Wittgenstein or Ryle), Baker explicitly describes her position as a metaphysical one.[2] She "uses the term 'practical' not to denote a theory of truth or meaning, but to denote a method in metaphysics", specifically that "[t]he *nature* of the attitudes is best revealed by their operation in our practices" (ibid., p. 20, emphasis mine). The attitudes are not features of either the brain or an immaterial mind-substance, but are attributes of whole persons. "Ontologically speaking, Practical Realism recognizes persons with intentional states, medium-sized objects, natural and artificial, conventions and institutions – everything presupposed by successful explanatory practices in science and in everyday life" (ibid., p. 21).

Certainly this approach is radically different from the materialist theories of Churchland and Fodor. Baker, however, is concerned to show that the explanations upon which we rely on everyday life are not of a lesser order than those provided by scientific theories. The Standard View is committed to "a particular conception of causal explanation: Genuine causal explanations must appeal to physical states or entities" (ibid., p. 17). This is not, according to Practical Realism, the case: beliefs *do* causally explain behavior, but are not simply physical states. Moreover, Baker insists that "even

[2] In this she resembles Strawson, though her ontology is somewhat different

if beliefs actually were brain states, it would not be by virtue of being brain states that they are causally explanatory" (ibid., p. 30).

Instead, Practical Realism emphasizes that beliefs explain behaviour only within the context of our everyday life practices, and that, conversely, these practices depend on the existence of the propositional attitudes. "Nothing would be a contract or an invitation to dinner or an election or a death sentence in the absence of beliefs, desires or intentions" (ibid., p. 4). In opposition to the view of causal explanation given under the Standard View, Baker suggests a wider view of causal explanation that is able to encompass the causal role of beliefs. "The basic idea is that we know that we have an adequate causal explanation when it affords control over phenomena of the type explained. When we can produce or prevent a phenomenon at will, we know that we have found a cause" (ibid., pp. 121-2).

Of course, Baker acknowledges that this definition provides "only a sufficient condition for a causal explanation" and that "many causal explanations will elude the test" (ibid., p. 121). However, it should be noted that many scientific theories predict and explain using accounts of sufficient causes; while the ideal scientific explanation offers an "if and only if" necessary cause, the practice of science is more lax.

This description also implies that folk psychology is not a theory, in the scientific sense of the word. Thus, it "is not in competition with scientific theories in any wholesale way", though it "may be corrected by science" (ibid., p. 240). Again, this does not suggest that explanations within the context of practical realism are subordinate to

those offered by science, rather "Practical Realism offers a way to do justice to common sense and science at the same time" (ibid.).

The analysis of beliefs, according to Practical Realism, should be done by a specification of the counterfactuals relevant to that belief. "Whether a person *S* has a particular belief...is determined by what *S* does, says and thinks, and what *S would* do, say and think in various circumstances, where 'what *S* would do' may itself be specified intentionally" (ibid. pp. 154-5). Thus, "one may have a belief even if one never manifests it in overt behavior...even if there are not relevant nonactual circumstances in which one would manifest it in overt behavior" (ibid., p. 155). According to Baker, we know *which* counterfactuals are relevant by understanding "the 'that' clause of the attribution" (ibid., p. 156), generalizations about how people behave and the circumstances, and other attitudes of the believer. The relation between the "that" clause and the relevant counterfactuals illustrates "the stubbornly nonreductive character of belief, on the Practical Realist's view" (ibid., p. 157).

I have spent less time outlining Baker's view of folk psychology, and particularly the theories that arise from it, because I believe that the questions Baker raises are more important for my own purposes than the answers she provides. The conceptual analysis of what it means to have a particular belief simply puts the content of the belief into the

context of a web of intentional attitudes: it is a philosophy of language, not a philosophy of mind.[3]

Baker does, however, diagnose a problem faced by both Fodor and Churchland. Folk-psychology is concerned with explaining our behaviour, and that behaviour cannot be understood without reference to the social context in which it takes place. Fodor's methodological solipsism is simply not adequate. Yet eliminativism is not the answer either: without some conception of the mental activities we wish to explain, our theory of mind will be barren.

From Fodor and Baker, then, I wish to take the suggestion that the questions we want philosophy of mind to answer must be rooted in our everyday conception of the mind, and in the way in which that view of the mind leads us to act in the social world. Baker emphasizes that cognition is not, and cannot be, a purely internal phenomenon. We are linked to the world, and our mental processes cannot be considered in isolation from that link. Yet Baker's Practical Realism does not allow us to focus on what I consider to be the most compelling questions. The mind, it seems, is more than the sum of its social practices. The mere fact that we can engage in any number of discrete language games and still experience our lives as continuous suggests that there is something more in our heads than a set of discrete dispositions. We must ask, then, what it is about the human mind that allows us to act in a flexible manner, what allows us, in Baker's own terms "to produce or prevent a phenomenon at will" (ibid., p. 122).

---

[3] Fodor also acknowledges the overlapping domains of the two theories: "Every time a philosopher of language turns a corner, he runs into a philosopher of mind who is pounding the same beat" (Fodor, 1987,

I suspect that Fodor has grasped the answer when he suggests that cognition is symbol manipulation. Yet part of the essence of a symbol is that it represents something – an object, a group of objects, an abstract entity - external to the subject. By adopting a position of methodological solipsism, Fodor can only explain the internal syntactic relations among symbols; he does not do justice to the relation between symbols and the external world.

Churchland, perhaps, offers the best account of this external relation to the world. By shifting from a description of mental functioning to a description of neural functioning, Churchland is able to describe the representation of sensory stimuli in the brain. He also offers some evidence that computation over these representations follows the same physiological rules as does representation itself. However, Churchland's relentlessly bottom-up approach is unlikely to allow us to distinguish between the "state space vectors" arising from the taste of an orange and those caused by, for example, a perception of injustice or a sudden overwhelming sense of existential dread. Churchland's eliminative materialism suggests that much of our mental life, or at least the language in which we account for it, is a mere illusion. Even if that is the case, there must be (on a materialist account) something about our brains that allows us to *have* these illusions. We may not really have beliefs, but we have the capacity to believe that we do. And, again, Churchland's bottom up approach is unlikely to be able to explain this fact.

---

p. xi).

To summarize, folk psychology and the propositional attitudes are an important part of the way that we experience our mental life, and any satisfactory philosophy of mind must address our capacity to understand our thoughts and actions in intentional terms. Fodor accepts this, but his theories are unable to provide a bridge between rational cognitive activities and the physiological mechanisms from which they emerge. Churchland approaches this distinction from the other side, and threatens to do away with rational cognition altogether, in favour of a stripped down mechanistic simulation of symbol manipulation. Baker emphasizes the need to consider rational behaviour in the context in which we actually behave; descriptions of beliefs and desires must take into account not merely the "aboutness" of the descriptions themselves, but also the world those descriptions are about.

In the next chapter, I will argue that the categories of folk psychology can only be defended when they are considered as the result of our ability to use symbols. Terrence Deacon argues that the faculty of symbol manipulation has made a qualitative difference in our cognition, one which has infected all of the other ways in which we represent. This capacity, however, is not ontologically distinct from any of the other capacities possessed by our species; it is an outcome of the evolution of our brains, and can be explained (though not necessarily detected) using Churchland's bottom-up approach. Deacon's theory also shows how rational cognition, in Fodor's sense of "hypothesis testing" emerges from the physiological mechanisms of the brain. And, finally, it ties our symbols to the world that those symbols represent. Deacon's description of human

beings as "the symbolic species" answers the questions raised by Fodor, Churchland and Baker, and allows us to see the whole problem, rather than a disparate collection of parts.

CHAPTER 2:

Ultimately, it seems that the debate over folk psychology is intimately connected with the role of language in cognition. Each of the three authors discussed in the first chapter has very definite views on how language should be seen, and these relate directly to their position on the status of folk psychology. Churchland views language as having a primarily social, communicative function (Churchland and Churchland, p. 219); it does not play a large role in constituting thought, and so cannot be guaranteed to reflect the cognitive processes that produce our behaviour. Even though, of course, language[1] is mediated by the brain, it is merely the product of a small, and relatively discrete, subsystem of the brain (ibid., p. 218).

For Fodor and for Baker, language does play an important role in our perceptions of the world, and in behaviour. Moreover, both claim that folk psychology provides a valid description of much of that behaviour. Like Churchland, Fodor is interested in developing a philosophy of mind that is continuous with scientific approaches to cognition; his concern is to justify the use of folk psychological terms in the context of a more rigorous psychology. By contrast, Baker is not interested in developing a philosophy of mind, so much as a socially based philosophy of language. Her approach to folk psychology involves its explanation in terms of the context(s) in which it is used.

[1] Or, in the case of Fodor, the underlying language of thought that natural language reflects.

Like Churchland, she considers language to be a social phenomenon, and she claims that it is best understood through an analysis of social practice, rather than of cognition.

Each of these authors, then, though disagreeing in their theories, share fundamental assumptions with each of the others. Fodor and Baker want to retain folk psychology, if not as a full-blown theory, at least as a useful, and largely correct account of human behaviour. Churchland insists that folk psychology *is* a theory, or at least a proto-theory, and rejects both its descriptive value and the beliefs, desires, and so on that it postulates. Baker and Churchland agree that language merely serves a social, communicative purpose, while Fodor insists that it is a reflection of the processes underlying cognition. Finally, Churchland and Fodor adopt a position of methodological solipsism, effectively divorcing cognition from the world in which it takes place. For Baker, language, and folk psychology, can only be understood in terms of social interaction.

It seems to me that each of these theorists is partially correct. Like Fodor, I believe that language is an important part of human cognition, and that it is involved in the constitution, rather than merely the description, of our thoughts. Our thoughts, though, are usually about some feature of the world, and I agree with Baker's claim that folk psychology is an inherently social practice. I do not agree, however, that it can only be analyzed in a social context. Instead, I share Churchland's interest in the neurobiological mechanisms that underlie human behaviour. However, I believe that the job of these theories is to explain how our common sense, folk psychological explanations work, not to explain them away. The question I want to ask, then, is how,

given what we know about how the brain works, can we explain our ability to use talk of beliefs and desires to make sense of the world?

In order to answer this question, I will be using Terrence Deacon's theory of brain-language co-evolution, and his description of human beings as "the symbolic species". Although Deacon does not himself address the issue of folk psychology, I will show that his theory addresses the positions advocated by Churchland and Fodor, and also illustrates how the methodological biases of each have led them to an overly narrow view of folk psychology. In addition, Deacon's approach overcomes the limitations imposed by methodological solipsism, and emphasizes that language is a reflection of the way we understand the world. This addresses Rudder Baker's concerns about the role of social practice in language and in folk psychology.

Deacon's research is in the area of biological anthropology. He studies the structure of the primate nervous system and of fossil skulls in order to explain the evolution of the human brain. In The Symbolic Species, he develops a theory of brain-language co-evolution that argues that the ability to use language is a consequence of massive changes in brain structure. However, the evolution of the brain is also due, in part, to the emerging linguistic capacity of early hominids. At the same time, languages evolved (and much more quickly than the brain!) to adapt to the structure of the brain.

In this thesis, I will be considering Deacon's discussion of language and of the brain, but I do not address the co-evolutionary theory that forms the core of his book.[2] Instead, I am concerned to draw out the implications of his views on language and

neuroanatomy for the problem of folk psychology. While Deacon does not address these issues, I believe that his work has profound implications for philosophy of mind.

Deacon uses C.S. Peirce's categories of icon, index and symbol.[3] I will begin by outlining these categories as described by Peirce. I will then describe Deacon's use of Peirce's categories. Deacon extends Peirce's categories to apply to all cognition – in human beings and in other animals. He claims that human beings are the only species whose members routinely cross "the symbolic threshold", and describes why this is the case in terms of the evolutionary development of the human brain. I will discuss Deacon's theories in detail, and then assess his claims regarding the relation between our capacity for symbolic thinking and language. While Peirce claims that "all thinking is in signs", Deacon suggests that human thought is strongly influenced by the activity of the prefrontal cortex, and is therefore largely symbolic.

A sign (or representamen), according to Peirce, is "something which stands to somebody for something in some respect or capacity" (Peirce, 1897, p. 11). Thus, a sign must have an object, which it represents, or stands for, as well as an interpretant, which links the sign and its object. Icons, indices and symbols are all signs, and are differentiated by the kind of relationship that they bear to their object.

An icon represents its object through its similarity, in some respect, to it. "That is, a quality it has *qua* thing renders it fit to be a representamen" (ibid., p. 16). Moreover,

---

[2] Reviews of Deacon's book have tended to emphasize the evolutionary implications of the theory. Both Ruse (1998) and Turner (1998) situate Deacon's theories with respect to the work of Chomsky and of Pinker, while Nixon (1998) briefly outlines Deacon's discussion of consciousness.

[3] These categories, however, are part of a much larger schema. In this work, I will not try to situate the "Second Trichotomy" of icon, index and symbol in the larger context of Peirce's semiotic.

it possesses that quality regardless of whether its object exists, "such as a lead-pencil streak as representing a geometrical line" (ibid., p. 15).

An index "is in dynamical (including spatial) connection both with the individual object, on the one hand, and with the senses or memory of the person for whom it serves as a sign" (ibid., p. 18). An index is "really affected" by its object; a symptom is an index of a disease, because it is caused by its object. Not all indices are causally related to their objects, however. A clock indicates the time of day, but is not caused by it. Peirce describes an index as existing in the imperative mood: "Anything which focuses the attention is an index. Anything which startles us is an index, in so far as it marks the junction between two portions of experience" (p. 19). As a result of the indexical relationship between these two portions of experience, we learn to associate them, or develop the habit of associating them.

The third class of sign is a symbol. The connection between a symbol and its object is the result of a convention. Because it is not a "natural" sign, in the sense of an icon or an index, a symbol "would lose the character which renders it a sign if there were no interpretant" (p. 15). While all three classes of sign depend upon being interpreted to function as a sign, only symbols *acquire* their character as sign from being interpreted. Thus, for example, "[a]ll words are symbols, because a word can only signify what it does if it is understood to have that signification" (ibid.).

Peirce describes the relationship between icons, indices and symbols as hierarchical. Because he describes a symbol as "a law, or regularity of the indefinite future" (p. 22), he claims that the symbol must be grounded in a concrete relationship

between sign and object. "Consequently, a constituent of a symbol may be an Index" (p. 22). In a sense, a symbol is a class of which indices are members. Furthermore, Peirce suggests that "[t]he only way of directly communicating an idea is by means of an icon, and every indirect method of communicating an idea must depend for its establishment upon the use of an icon" (ibid.). Both indexical and symbolic reference, then, depend on iconic references. Moreover, a symbol "cannot indicate a particular thing" (p. 22), so symbolic reference also depends on indexical relationships between particular objects. This hierarchy, as I will show, plays an important part in Deacon's use of Peirce's categories.

Deacon accepts Peirce's proposal that icons, indices and symbols are related hierarchically. In fact, he claims that Peirce's "most fundamental and original insight" with respect to interpretation was his realization that "the difference between different modes of reference can be understood in terms of *levels* of interpretation" (Deacon, p. 73). However, the distinction between each kind of sign lies not in the ways that it can be interpreted, but in the way in which a given sign generally *is* interpreted. Whereas Peirce tends to talk about icons, indices and symbols (i.e. the signs themselves), Deacon emphasizes the role of the interpreter in fixing the meaning of a sign, and talks in terms of iconic (etc.) *reference*. Moreover, the way(s) in which a given interpreter is *able* to understand a sign depends on that interpreter's cognitive abilities. Different kinds of reference are available to different kinds of interpreters. "It's not just the case that we are able to interpret the same sign in different ways, but more important, these different

interpretations can be arranged in a sort of ascending order that reflect a prior competence to identify higher-order associative relationships" (ibid.).

Deacon starts with a description of iconic reference. He describes this kind of reference in negative terms: while we *explain* an icon in terms of the resemblance between a sign and its object, "the interpretive step that establishes an iconic reference is essentially prior to this, and it is something ...that we *don't* do" (ibid., p. 74, my emphasis). For example, a moth with protective colouration has wings that resemble tree bark in pattern and colour: birds will often fail to notice the moth on the tree. Deacon uses this example of "no representation at all" to illustrate iconic reference because it clarifies his shift in emphasis "from the relationship [between sign and object] to the process behind it" (ibid., pp. 75-6). The moth's wings, to echo a point recognized by Peirce, are not inherently iconic – they are interpreted to be so – or, in the case above, are not so interpreted. The bird failed to distinguish between the moth and the tree because it responded to both of them with the same interpretive perceptual process. It was this interpretation that established the iconic relationship between moth and bark. (This example also reinforces Peirce's claim that an icon retains its iconic character even when it is not interpreted as a sign.) Deacon concludes that "iconic resemblance is not based on some prior ground of physical similarity, but in that aspect of the interpretation process that does not differ from other interpretive process" (p. 76). For a similarity to be grounded in iconic reference it must be recognized. Thus, for Deacon, iconic reference is a default position, holding between an icon and its object when the interpreter is inattentive or is prevented by a "minimal nervous system" (ibid.) from making higher-

level discriminations. Iconic reference, then, requires that we discriminate between a sign and its object, but emphasize the similarities between them. "That facet or stage of my interpretive recognition process that is the same for a sketch and the face it portrays is what makes it an icon" (p. 76).

Following Peirce, Deacon describes an index as something which functions as a sign in virtue of its spatial or temporal relationship to its object, or in virtue of "predictable co-occurrence" (p. 77). This co-occurrence on its own, however, is not enough to guarantee an indexical relationship. "What makes one [thing] an index of another is the interpretive response whereby one seems to 'point to' the other" (ibid.). Deacon describes this "pointing to" as the result of a relationship that is constructed by the interpreter, and examines indices in terms of the abilities necessary to create these relationships. Basically, he says, an "indexical interpretation is accomplished by bringing [an] assembly of iconic relationships to bear on the assessment of new stimuli" (Ibid.). When we, for example, smell smoke, and look for a fire, we expect to find a fire because past experience has taught us to associate the smell of smoke (an index) with fire (its object). The current experience of the smell of smoke calls to mind past, similar experiences – an iconic reference based in the perceptual similarity of the experiences. It also calls to mind past experiences in which the smell of smoke was associated with fire, and these similar experiences are also icons of each other. Finally, Deacon claims, a third iconic relationship is established by the repeated correlation of smoke and fire in these past experiences. This last iconic relationship is the key to the indexical relationship. "Because of this [relationship, we] recognize the more general similarity of the entire

present situation to these past ones, not just the smoke and not just the fire, but also their co-occurrence, and this is what brings to mind the missing element in the present case: the probability that something is burning" (ibid., p. 78).

We are able to make novel indexical references because of our memory of previous iconic references. Deacon emphasizes, however, that these new references are not explicable simply in terms of perception and learning. The description of an object as an index or an icon "refer[s] to the *inferential* or *predictive* powers that are implicit in these neural processes" (p. 78). These powers imply something that is not physically part of a situation, but that is "virtually re-presented by producing perceptual and learned responses like those that would be produced if they were present" (ibid.).

Indexical representation underlies what Deacon calls "the common sense idea" of symbolic association. For example, when we learn to pair a word (written or spoken) with an object in the world, we are merely extrapolating past correlations of the word and the object. This simple association is not truly symbolic. What really goes on in symbolic reference is, according to Deacon, something other than simple correlation.

Unlike indexical reference, symbolic reference holds even when the sign and its object are not consistently related in space or time. In the story of "The Boy who Cried Wolf", "the indexical function of...the word 'wolf' fails because of its lack of association with real wolves, *even though the symbolic reference remains*" (p. 82). Those who heard the boy never stopped understanding what he meant by "wolf". Deacon points out that a physical link between a word and its object can be rare, or even impossible, as in the case

of angels and of unicorns. Still, we understand what these words mean, and what the object is to which they refer.

Deacon suggests that this understanding is possible because words do not refer simply to objects, but also to other words. "In fact, they are incorporated into quite specific individual relationships to *all* other words of a language" (ibid.), in the way in which a dictionary or a thesaurus "maps" words onto other words. Moreover, if this shared mapping fails (Deacon gives the example of slang terms that "radically re-use" other words, such as "plastered" for intoxicated), so does the reference.

Deacon recognizes that this duality of reference is reflected in Frege's classic distinction between sense and reference (pp. 83, 61). In this schema, the sense of a word depends on its relationships with other words, while its reference "is something in the world which corresponds with this term and its sense" (p. 61). Icons and indices, then, have reference, but only symbols have sense.

Deacon notes that most theories claim that "we use the sense [of a word] to pick out its reference, not vice versa" (p. 83). It is sense, he claims, that creates the referential relationship between words that "forms a system of higher-order relationships that allows words to be *about* indexical relationships, and not just indices in themselves" (ibid.).

In most cases, we do not use words as indices that point directly to some object in the world. Rather, words occur in a context along with other words. Deacon describes the indexical power of words as *distributed* in the relationships between words. "Symbolic reference derives from *combinatorial* possibilities and impossibilities, and we therefore depend on combinations both to discover it (during learning) and to make use of

it (during communication) (ibid.). Nothing is a symbol in isolation: rather we should speak of a symbol system.

How do we learn to use symbols? Deacon suggests that the difficulties inherent in the acquisition of symbolic reference are "a consequence of the fact that what determines the pairing between a symbol (like a word) and some object or event is not their probability of co-occurrence, but rather some complex function of the relationship that the symbol has to other symbols" (ibid.) Symbol use, then, is not just the result of learning these relationships but also of *unlearning*. "Learning is, at its base, a function of the probability of correlations between things, from the synaptic level to the behavioural level. Past correlations tend to be predictive of future correlations. This is the basis for indexical reference" (ibid.) Symbolic reference, however, can only take place if the importance of these indexical relationships is minimized and subordinated to relationships among other symbols. These symbolic relationships are generally quite unlike the highly correlated relations between indices. "Words that carry similar referential function are more often used alternatively and not together, and words with very different (complementary) referential functions tend to be adjacent to one another in sentences" (ibid.).

Learning to use symbolic, rather than indexical, relationships "is initially a change in mnemonic strategy, a recoding. It is a way of offloading redundant details from working memory, by recognizing a higher-order regularity in the mess of associations" (p. 89). Rather than having to remember many concrete indexical relationships, symbol users can look for abstract relationships between groups of indices. Once learned,

however, the use of symbolic reference "takes no more time than the process of perceptual recognition" (p. 93).[4] While the underlying indexical associations may take time and effort to learn, the recoding of these associations into a system of symbols "is not *learned* in the same way; it must instead be discovered or perceived, in some sense, by reflecting on what is already known" (ibid.).

Deacon notes that the combinatorial possibilities arising from even a small set of symbols is staggeringly large, and that reflection upon various possible symbolic relationships often results in what psychologists have called "insight learning".[5] Deacon contrasts symbolic insight with learning by rote. The insight occurs when we are able, instead of blindly following an algorithm, to understand the logical relations behind it.

This insight is what Deacon calls the "symbolic threshold". Crossing this threshold involves "a radical change in cognitive strategy" (p. 67). Association, Deacon points out, requires that the stimulus and the object or event to which it refers be linked, if not all the time, at least frequently. Rote learning requires an indexical relationship if it is not to be extinguished. By contrast, the symbolic relationships that we understand become independent of the indexical references that underlie them.

Deacon further claims that it is only human beings that have crossed the symbolic threshold. While there is evidence that some of the great apes are capable of developing

[4] The reasons for this will be discussed in the second part of this chapter.

[5] Many insight problems can be solved using indexical relationships (" 'visualizing' the parts of a relationship in a new way", (p. 94)). Deacon cites Wolfgang Kohler's classic experiments in which a chimpanzee had to figure out that a banana suspended from the ceiling of a cage could be reached by stacking two wooden boxes. This insight was not symbolic, but resulted from physical, though apparently not goal-directed, manipulation of the boxes.

an impressive capacity for symbolic reasoning,[6] the fact is that this capacity is unexploited by any species other than human beings. In the next section of this chapter, I will describe the structural and functional features of the human brain that, according to Deacon, made our symbolic competence possible. In order to provide a complete account, I will use not only Deacon's own argument, but will supplement his analysis with that of other theorists. While Deacon's evidence appears to agree with that of other researchers, some of the important points supporting his theory are made more clearly by other writers. In addition, I am especially concerned to show that the neurological changes that make language use possible have, in fact, infiltrated the whole brain. Much of the supplementary information I have gathered reinforces this point.

The Evolution of the Symbolic Species

Language use, Deacon claims, is unique to humans. The communicative strategies employed by other species, no matter how complex they may be, are not true languages, in that they lack both "the hierarchic-recursive logic of sentence structure and the distributed multilevel associative relationships that support symbolic reference" (p. 148). To explain *why* language is unique, Deacon suggests that we look for features that are unique to the human brain. The most obviously distinct part of the brain is the prefrontal cortex, which is much larger in primates than in other animals, and larger in humans than in any other primate.[7] Deacon describes how the evolutionary expansion of this portion of the brain has resulted in a radical rewiring of the entire brain. This

[6] Deacon discusses at length the work of Sue Savage-Rumbaugh and Duane Rumbaugh (pp 84-92), who have taught Kanzi, a pygmy chimpanzee, to use a system of signs. The process by which Kanzi acquired competence with the symbols is thought to be similar to that of language acquisition in children.

rewiring, in turn, changed the cognitive strategies available to human beings, and resulted in the evolution of our linguistic capabilities.

The embryonic development of an organism depends not merely on the genetic code, but also on the organizational influences provided by local interactions between populations of cells. The differentiation of a group of cells into distinct tissues and organs is a complex process dependent largely on the chemical environment in which the cells develop. In the case of the brain, the scope of these "local" interactions is increased because of neuronal morphology: the cell's axon (output fibre) and dendrites (input fibres) can be quite long. "Because neurons are specialized for cell-cell communication over long distances, they can utilize an additional level of structural information over and above the regional segregation of tissues and cell lineages to help organize their functions" (Deacon, p. 194). As a neuron grows, it sends out a long axon, which is chemically attracted to a particular region of the brain. Upon reaching its destination, the axon branches and makes multiple synaptic connections with its "target neurons". The differentiation of the particular neuron can depend not only on the cells in its immediate vicinity, but also on the synaptic connections it makes with cells in other regions of the brain.

The development of these synaptic connections is, however, only part of the story. The fetal brain makes many more of these connections than will be found in the mature brain. Following the phase of neuronal growth and synaptic proliferation comes a period in which many of these connections die off. Which connections remain can depend

[7] Brodmann calculated that the prefrontal cortex represents 29% of the total cortex in humans, and 17% in

either on the expression of a preprogrammed "suicide gene," or on competition between cells for local resources. "[This competition] turns out to be a very important mechanism for the developmental sculpting of parts of the nervous system, serving to match the proportions of one to another" (ibid., p. 195).[8]

Deacon terms the role of regional brain size in the development of the brain as "displacement".[9] "In general terms, relative increases in certain neuronal populations will tend to translate into the more effective recruitment of afferent and efferent connections in the competition for axons and synapses" (p. 207). In human beings, the increased size of the prefrontal cortex (relative to other brain structures) has resulted in a greater connectivity of this region with the rest of the brain.

It is a truism in biology that "structure dictates function". However, Deacon needs to show *how* the changes in prefrontal size and connectivity are linked to the changes in neural function that underlie language use. He must specify exactly *which* areas of the brain are closely interconnected with the prefrontal cortex, and what the functional outcome of these connections appears to be.

There are several ways in which these interconnections can be studied. First, clinical studies provide clues as to the function of a particular brain region by showing which behaviours are affected by damage to that area. In the case of the prefrontal cortex, this often involves developing specific and sophisticated tasks to reveal deficits that are otherwise "silent" (Pribram, 1985).

---

the chimpanzee; by contrast, it accounts for 7% in the dog, and 3.5% in the cat (cited in Fuster, p. 3)

[8] Gerald Edelman refers to this process as a form of "neural Darwinism". His theories, however, are concerned primarily with smaller-scale connections between groups of functionally similar cells.

Second, detailed anatomical studies can be performed using other primates. Aboitiz and Garcia, for example, use the connections that can be demonstrated in the macaque brain to extrapolate to the connections between homologous areas in humans.

Finally, new technologies have been developed that allow researchers to study – noninvasively – the level of brain activity in different areas of the brain as human volunteers perform specific tasks. These procedures are particularly helpful, because they allow scientists to investigate normal, as well as damaged, brains. Theories of brain function, however, continue to rely on all three methods.

The prefrontal cortex is defined as the part of the cerebral cortex that receives projections from the mediodorsal nucleus of the thalamus (Fuster, 1989, p. 2). This distinguishes it from the precentral and limbic portions of the frontal cortex, which are concerned primarily with the control of movement and emotion, respectively. The prefrontal cortex can be further subdivided based on both cortical cytoarchitecture (e.g. Brodmann's areas), and apparent functional similarities.

The frontal lobes are reciprocally connected with areas of the thalamus, the limbic system, the basal ganglia and the cerebellum. However, it is primarily its synaptic connections with other areas of the cerebral cortex that are thought to have made the development of language possible (Aboitiz and Garcia, 1997). In addition to receiving input from various sensory systems, (visual, auditory, olfactory and somatic), the prefrontal cortex is also reciprocally connected with areas in the parietal and temporal lobes that are involved in higher-level associative processes. In particular, Wernicke's

[9] Deacon describes displacement as "an evolutionary mechanism", however, the process is the same in the

area appears to be reciprocally connected to several portions of the prefrontal cortex. Damasio (1992) suggests that Wernicke's area acts as a multimodal convergence zone, in which speech sounds are mapped to objects, allowing them "to be used subsequently to invoke conceptual meanings" (p. 534).

While damage to Wernicke's area appears to affect auditory comprehension, it is no longer seen as the area in which language comprehension takes place. Rather, comprehension is viewed as a "complex process [involving] numerous cortices of different modalities and hierarchies distributed over the entire brain" (Damasio, p. 534).

The multimodal associations made in Wernicke's area appear to be "shallow". While they involve relations between modalities, but there is no evidence that these relations are symbolic. Rather, they reflect Deacon's description of indexical reference. (By contrast, symbolic reference is used in tasks requiring "deep" or semantic processing, in which the meaning of words is analyzed. These activities require the involvement of the left prefrontal cortex (e.g., Gabrieli *et al.*, 1998).) Seen in this light, the resulting deficits in auditory comprehension following damage to Wernicke's area might arise from failure of the indexical relationships upon which symbolic reference is built.

This hypothesis is supported both by clinical pictures of frontal lobe damage and by the theories of prefrontal function to which this clinical evidence has contributed. A global theory of prefrontal effects on behaviour has been difficult to develop because of the multiple anatomical subdivisions of the cortex and the subtle variations in function of each. "Because different prefrontal areas are connected to different cortical and

---

development of an individual brain.

subcortical structures, when they are damaged, they produce slightly different types of impairment" (Deacon, p. 259). Still, Deacon suggests that there appear to be common traits among the many different specific deficits associated with different areas of the prefrontal cortex. Two specific deficits described by Deacon have particularly interesting implications for symbolic reference. The first involves the transfer of "learning sets". The second is the performance of tasks that require that the subject take another person's perspective.

The prefrontal area appears to be necessary for the transfer of information from one learning task to another (p. 262). This transfer requires "using information from previous trials, but divorcing it from specific stimuli" (ibid.). Deacon had noted, earlier in his argument, that this transfer is, in its simplest form, actually not symbolic. Rather it is a "complex form of indexical association...often confused with symbolic association" (p. 80). Some of the more complex forms of these tasks (at which only "some of the larger brained primates succeed" (p. 262)), actually *are* symbolic, in that they "involve transfer of an inverse pattern of associations from one task to another" (ibid.).

Related to this first deficit is the inability of some patients with prefrontal damage to take an allocentric perspective on a problem.[10] Deacon describes this task as "analogous to using a mirror. In order to think in allocentric terms, one is required to perform a systematic mental reversal of response tendencies, and so egocentric information must be continually used as the frame of reference, but responses based on it must be specifically inhibited and inverted" (pp. 263-4).

Both of these tasks, then, require the abstraction of information from the environment, and the ability to manipulate this information independent of its original context. Deacon further tells us that "[p]refrontally damaged patients exhibit a tendency to be controlled by immediately correlative relationships between stimuli and reinforcers, and this disturbs their ability to entertain higher-order associative relationships" (p. 264). Clinical evidence suggests that prefrontal involvement is necessary if we are to use symbolic, rather than indexical, reference.

Fuster (1989) claims that the prefrontal cortex as a whole appears to be involved in the temporal organization of behavior. He suggests that this involves three main tasks: provisional (short-term) memory, preparatory set, and interference control. Fuster uses the phrase "short-term memory" as an equivalent to "working memory".[11] It refers to the ability to retrieve past events, stored in long-term memory, that are relevant to the current situation, and to use them to analyze the new situation. Preparatory set involves both the anticipation of future events, and the preparation for them. "Experience allows the organism to form, out of old elements, the cognitive scheme of the plan that is to serve as the template for new action" (p. 164).

Most interesting, for Deacon's purposes, is the function that Fuster calls "interference control". This involves the recognition and suppression of stimuli that are irrelevant to a particular goal-directed behaviour. This often requires that a response to

---

[10] Frith (1999) notes that these tasks involve an ability to take what Dennett calls "the intentional stance", or to have, in Premack and Woodruff's terms a "theory of mind".

[11] Often, the term "working memory" is used to mean "holding information in mind while not acting on it" (Deacon, p. 262). This use of the term incorporates Fuster's short-term memory, and aspects of his "interference control".

the interfering stimulus be suppressed so that the organism can make another, more appropriate, response.

This is similar to the kind of cognition that Deacon claims is required in learning a language. Crossing the symbolic threshold requires the suppression of indexical reference in favour of the symbolic relationships among words. "The contributions of prefrontal areas to learning all involve, in one way or another, the analysis of higher-order associative relationships. More specifically, judging from the effects of damage to prefrontal regions, they are necessary for learning associative relationships where one associative relationship must be subordinated to another. These are the critical learning problems during symbol acquisition" (Deacon, p. 264).

Deacon suggests that the failure of other species to acquire language arises from the "less effective regulation of competing learning tendencies" by the prefrontal cortex. Conversely, the predominance of the prefrontal cortex in human brains suggests that "human information processing should be biased by an excessive reliance on guidance by the kinds of manipulations that prefrontal circuits impose upon the information they process" (p. 237). However, there are limitations to this influence. The prefrontal cortex appears to be necessary for the creation of new symbolic references, and is less important in the use of relationships or concepts that have already been mastered. Once a particular task is learned, "the level and extent of metabolically activated ventral frontal and cingulate cortex is significantly reduced" (p. 267). Once the novelty wears off, it appears, so does the prefrontal activity. In addition, damage to the prefrontal cortex does not affect either the production or the comprehension of speech. While the prefrontal

cortex disproportionately influences human cognition, its influence affects the functioning of the entire brain.

This fact allows us to reconcile two apparently contradictory aspects of the role of language in thought. The first is that language is not simply a module added to an otherwise language-free brain. Rather, it has infiltrated all of human thought. The second point is that linguistic/symbolic processing has not *replaced* pre-linguistic reference, but functions in parallel along with older, indexically based, cognitive strategies. The attempt to make an absolute distinction between linguistic and nonlinguistic cognition arises from a false dichotomy.

Anatomically, the first point is best expressed using Deacon's concept of displacement. The massive expansion of the prefrontal cortex in human beings has changed the structural relationships existing throughout the brain. The influence of the neural processes that occur in the prefrontal cortex constrains "the way the parietal cortex handles tactile information, the way the auditory cortex handles sound information, the way the visual cortex handles visual information" (Deacon, p. 265).[12] This means, for instance, that there is no way to separate our perceptual experiences from our conceptual ones. The higher-order discriminations that we make mean that there is no such thing as a "pure" perception free of conceptual distinctions.

To return to Peirce's vocabulary, symbolic, indexical, and iconic reference are inextricably mixed in human thought. This point is also made, in a different context, by

[12] At the same time, Deacon cautions against concluding that the prefrontal cortex is simply "the place where symbols are processed" In fact, "widely distributed neural systems must contribute in a coordinated fashion to create and interpret symbolic relationships. The prefrontal cortex is only one of these" (p. 266).

a commentator on Peirce's work, Douglas Greenlee. I will outline Greenlee's arguments in some detail, because they make clear the way that symbolic reference has infiltrated human cognition. Greenlee critiques Peirce's claim that the relationship between icons, indices and symbols is hierarchical, and argues that icons and indices have symbolical components as well. Moreover, iconicity also has an indexical component. Like Peirce, Greenlee focuses on human cognition, which has been infected by the evolution of symbolic reference.

The defining mark of an icon, according to Peirce, is its similarity to its Object. This, however, implies that "similarity is a sufficient ground of iconic representation" (Greenlee, p. 75), which, combined with Peirce's further assertion that everything is like something else in at least some respect,[13] "leads to absurd consequences" (ibid.). Any object that is perceived or conceived resembles another object in at least one respect, and so iconically represents it. Moreover, this object must iconically represent still others, "so that the innocent attempt to think of anything at all must loosen a flood of irrelevant associations into the mind" (ibid., p. 76).

In fact, however, representation does not work this way, and the similarities that we *do* perceive, are often the result of learned discrimination. Greenlee suggests that we should distinguish between resemblance and similarity, where the former term is used to designate only those similarities which are recognized. It is resemblance, not similarity that is "the true ground of iconic reference" (p. 77).

[13] Greenlee draws this assertion from an 1892 paper by Peirce, which demonstrates that, for any trait used to identify an object, there is at least one other object which shares that trait.

Greenlee further claims that resemblance relies on the establishment of a convention or rule "to the effect that the sign should be interpreted to signify in a certain way" (ibid.). The icon, then, is "symbolical" in that it depends on a conventional interpretation of the sign that emphasizes only a particular *kind* of likeness between the icon and its object.

According to Greenlee, the index, too, has a symbolic component. While Peirce defines an index as " 'a sign which refers to the Object that it denotes by virtue of being really affected by that Object'" (quoted in Greenlee, p. 84), he also claims that indicators of *any* sort are indices. Peirce lists relative pronouns, possessive pronouns, adverbs of place and time, and prepositions as words that serve an indexical function (Peirce, pp. 21-22). Yet he also claims, less than a page later, that all words are conventional signs, and therefore symbols. Greenlee further points out that *all* indices are symbolic. Even "Peirce's favorite example, that of the weathercock...signifies, in its typical use, as much by virtue of a rule of interpretation, delimiting the range and possibilities of its reference and confining these to a representation of the direction of the wind, as it does by virtue of a causal connection to its object, the wind. The two factors are inseparable" (p. 89).

Both icons and indices, then, have a symbolic component. Peirce further suggests that *all* signs have an indexical component: "it would be difficult, if not impossible, to instance an absolutely pure index, or to find any sign, absolutely devoid of the indexical quality" (p. 18). Greenlee notes that, in his later writings, Peirce describes the index as exerting a "physiological compulsion" which draws the interpreter's attention to its object. Furthermore, this physiological compulsion is rooted in habit. According to

Greenlee, Peirce's use of the notion of habit to explain the reference of signs is "plausible". It explains both the conventional (symbolic) and the indexical aspect of signs. In the latter case, Greenlee equates the idea of a conventional response with a habitual one, though he notes that because "habits can vary in strength, it is not difficult to see that some sign conventions are readily and arbitrarily alterable" (p. 91).

Peirce suggests that the interpretation of a sign as icon, index, or symbol is the result of a habit. This claim is compatible with Deacon's description of Peirce's categories. Although Greenlee rejects Peirce's early insistence on the hierarchical progression of icon, index and symbol, Deacon accepts it. According to Deacon, the hierarchy must be seen in the context of the evolution of animals with more complex nervous systems. While all of the signs discussed by Greenlee are those recognized and interpreted by human beings, Deacon demonstrates that "pure" iconicity and " pure" indexicality play an important role in the behavior of other animals. In his discussion of iconic reference, Deacon described a bird's failure to recognize a moth on a tree trunk as resulting from a lack of attention. However, the failure to make discriminations can also result from an inability to perceive them due to the structure of an organism's nervous system. The ability to make at least some discriminations can also be the result of learning, whether in the context of indexical or of symbolic reference. Regardless of the level of reference, the neurophysiological mechanisms underlying the interpretive process are likely to be the same, or highly similar.

It is here that Deacon's discussion of the brain overlaps with Churchland's theories. According to Churchland, there is no difference between organisms whose

nervous systems are " 'calibrated' by natural selection" (Churchland and Churchland, p. 217) to produce certain behaviors, and those whose calibrations are the result of experience rather than (or, more accurately, as well as) evolution. In both cases, behaviour hinges on the fact that "specific neural responses are regularly caused by types of state in the organism's normal environment" (ibid.), regardless of how these neural responses came into being. The level of analysis with which Churchland is concerned is primarily that of physiology, as it is reflected in connectionist theories of representation and cognition. At this level, Peirce's "habit" cashes out in terms of cellular response patterns and strengthened synaptic connections within small groups of neurons.

Habits, however, vary in strength, and there is a vast difference (at a behavioural level) between "simple fixed action pattern stuff" (ibid.), such as the pecking of herring gulls, and the flexible responses to the environment that human beings and other animals are capable of producing. In the latter case, a habit is best considered as a disposition to behave in a certain way, a disposition that can be broken or modified under a variety of circumstances. To return again to the synaptic level, no matter how strong the links between neurons, they are always open to modification.

The plasticity in these synaptic relationships suggests that, even at a physiological level, the PDP programs discussed by Churchland are too simplistic. They cannot fully account for the behaviour of neurons, much less of organisms. Edelman points out that "at the level of their finest connectivity, rich nervous systems like those of vertebrates cannot have precise, prespecified, point-to-point wiring, and that, in general, uniquely specifiable connections do not exist" (1989, p. 41). Perhaps we should think of neuronal

circuits as having dispositions, or tendencies, as well. Even the most "hard-wired" firing patterns, such as those that underlie the pecking of herring gulls, are modified by the environment in which they occur (Hailman, 1969).

In more complex actions, the response to a given stimulus can vary, and often involves a choice among several alternatives. It is at this level that behaviour reflects Fodor's characterization of cognition as a kind of hypothesis testing. This process involves the memory of past experiences, and the anticipation of the outcomes of various possible choices of response. However, as Deacon has described it, this kind of cognition can be the result of purely indexical reference. It does not require that the animal make the abstract associations that define symbolic reference.

For Fodor, there is no qualitative distinction between human cognition and that engaged in by other species. The language of thought is fundamentally the same for all animals; feline "mentalese" differs from that of human beings only in complexity. Our linguistic accounts of our actions merely reflect, fairly accurately, the underlying cognitive processes that take place in mentalese. Like Churchland, Fodor sees language as simply a newly evolved module that enables us to articulate what is going on in our heads.

Deacon, however, argues that language *changes* what goes on when we think. The capacity for symbolic reference does not, however, replace iconic or indexical reference; as Peirce claims, it is built upon them, and, in turn, enriches them. For instance, the ability to use symbolic reference changes the way in which we perceive the environment. Even simple relationships, such as those arising from iconic reference,

depend upon the way in which we categorize a situation. We think of Bach's music as having more in common with Beethoven's than with the Beatles', but whether we think of it as more similar to a song by the Beatles or to a painting by Caravaggio depends on the way in which we choose to classify them. In this case, it is not clear where iconicity stops and symbolism begins. It is clear, however, that the systems of symbols we are able to use do affect our perception of the world. Without some knowledge of the defining features of the Baroque era, we would not link Bach with Caravaggio. Similarly, a knowledge of musical theory and history can change the way we appreciate both Bach and the Beatles.

On Deacon's account, then, language is a fundamental part of human cognition. In the next chapter, I will discuss how his theory applies to folk psychology. I will argue that the description of human cognition as largely linguistic addresses problems raised by Churchland, Fodor, and Baker. It also relates the three approaches, and addresses questions left unanswered (or unasked) by each of these authors. Peirce's distinction between the three kinds of reference explains why we can describe the behaviour of other animals in intentional terms. It also accounts for the uniquely human characteristic of abstract thought, which includes the use of folk psychology to explain our own actions, and those of others.

CHAPTER 3:

One of the purposes of a philosophy of mind should be to explain everyday, folk psychological talk of the mental in more precise terms. It is generally acknowledged, though, that this explanation must be compatible with what we know about the workings of the brain. In the first chapter, I outlined the theories of three authors, Paul Churchland, Jerry Fodor, and Lynne Rudder Baker, each of whom has a definite explanation of folk psychology and definite views as to the methodology by which it should be studied. These three points of view seem incommensurable.

In the second chapter, I described Terrence Deacon's account of the evolutionary changes in the brain that enabled the human species to acquire language. The development of the capacity for symbolic representation that underlies language has radically changed the organization of the human brain (and, to a lesser extent, that of the other higher primates) compared to the rest of the animal kingdom.

Deacon's characterization of humans as "the symbolic species" provides an excellent starting point for a theory of folk psychology. It enables us to address the primary concerns of Churchland, Fodor and Baker, and it also provides a framework within which these disparate positions can be, if not reconciled, at least compared. Deacon's interdisciplinary approach to human cognition encompasses the more limited methodologies of Churchland and Fodor, and is also compatible with Baker's view of folk psychology as a social practice.

In the first section of this chapter, I will describe two approaches to the definition of belief. Baker's theory follows a tradition that describes beliefs as dispositions to behave. According to Fodor, beliefs are actually propositions, stored in the head, and written in the language of thought. However, neither theory gives natural language its proper emphasis in the explanation of folk psychology. In contrast to these theories, I will outline Peirce's description of the "fixation" of belief, and show how it is related to his semiotic theory. In the second part of this chapter, I turn to Deacon's work and show its relevance to the problem of folk psychology. According to Deacon, the capacity to use symbols, and therefore language, has radically changed human cognition. In the final section, I will argue that the practice of folk psychology is a result of the ability to use language[1] and that, more importantly, it is a necessary tool for coping with the world that language opens up to us. Although Deacon's theory uses scientific data to explain language and the mind, it is not incompatible with theories that approach folk psychology in purely philosophical terms. I will argue that Deacon's description of symbolic reasoning shows that it is this ability that gives rise to such concepts as normativity, agency and morality, and, further, that these concepts are at the centre of our use of folk psychology to explain human behavior.

From Disposition to Proposition

Among those who wish to defend folk psychology from the eliminativists, there is a difference of opinion as to the nature of belief. For Baker, calling something a belief allows us to explain an agent's disposition to behave in a particular way under certain

[1] As opposed to the possibility that language merely allows us to express the predictions that doing folk

circumstances. According to Fodor, a belief is a proposition; a statement of belief merely puts into words the underlying mental structure that *is* the belief.

The suggestion that a belief is a disposition was prominently defended by Gilbert Ryle. In his book The Concept of Mind, Ryle distinguishes between a disposition and an occurrence or episode. "To say that a person knows something, or aspires to be something, is not to say that he is at a particular moment in process [sic] of doing or undergoing anything, but that he is able to do certain things, when the need arises, or that he is prone to do and feel certain things in situations of certain sorts" (p. 116). Moreover, Ryle emphasizes that dispositions can be categorized differently, and "there are lots of different kinds of dispositional words" (ibid.). Some relate to a specific kind of activity and signify that a person has episodes of a particular behavior. For example, a smoker need not be currently smoking a cigarette to be described as a smoker (ibid., p. 117), however, that description would not be true if he never smoked at all. Other dispositional terms are examples of what Ryle calls "determinable dispositional words. They signify abilities, tendencies or pronenesses to do, not things of one unique kind, but things of lots of different kinds" (p. 118). "Belief" is, according to Ryle, an example of this second kind of dispositional word.

Baker's theory of folk psychology has been influenced by Ryle.[2] According to Baker, to attribute a belief to someone is to describe a disposition. "Since the term 'belief' is just a nominalization of 'believes that', *S* has a belief if and only if there is some proposition *p* such that *S* believes that *p*. Whether *S* believes that *p* depends solely

---

psychology allows us to make.

on what *S* would do, say, and think in various circumstances. Although *S* may not always manifest beliefs in behavior, there must be some circumstances in which *S's* belief makes a difference to what *S* would do, say, or think" (1995, p. 21).

Fodor, on the other hand, claims that beliefs are literally propositions in the head. "LOT [Language of Thought] wants to construe propositional attitude tokens as relations to symbol tokens. According to standard formulations, to believe that *P* is to bear a certain relation to a token of a symbol which means that *P*" (1987, p. 135). Even on this account, however, there is a sense in which beliefs can be viewed as dispositions. While they are stored as propositions ("written" in Mentalese), beliefs need not ever be translated into their natural language equivalent. "For example, it has probably never occurred to you before that no grass grows on kangaroos. But, once your attention is drawn to the point, it's an idea that you are quite capable of entertaining, one which, in fact, you are probably inclined to endorse" (1985, pp. 36-7).

In a sense, Fodor seems to want to claim, you have held the belief that "no grass grows on kangaroos" all along. In another sense, however, you most certainly did not. Intuitively, we want to say that, in order to *count* as a belief, a proposition should claim something that really matters. While we might assent to the proposition "no grass grows on kangaroos", it is difficult to imagine a situation in which we would really care, one way or another, whether it does.

The analysis of belief in terms of dispositions suggests that a belief is simply a passive state, which becomes active only under certain relevant circumstances. By

---

[2] Though Baker, unlike Ryle, sees belief explanations as *causal* explanations.

contrast, Peirce's view emphasizes the active *construction* of a belief when he suggests that a belief is "a rule for action." This characterization must be distinguished from the claim that a belief is "a disposition to act". The difference arises from Peirce's suggestion that a belief is something that we reach from a starting point of doubt. Thus "it is something that we are aware of [and] it appeases the irritation of doubt" (1877, p. 144). For Peirce, the formation (or "fixation") of a belief is an active reformulation of our perceptions of the world, and the process of developing a belief is an attempt to determine the best way to act in a perplexing situation.

Peirce admits that his characterization of belief and doubt is a bit hyperbolic. "To speak of such a [trivial] doubt[3] as causing an irritation which needs to be appeased, suggest a temper which is uncomfortable to the verge of insanity." Yet even mild forms of doubt cause us to think. "However the doubt may originate, it stimulates the mind...to activity" (1878, p. 164).

However, Peirce also characterizes a belief as "the establishment of a habit" (ibid., p. 166). Once the habit is in place, it becomes a disposition. "Belief does not make us act at once, but puts us into such a condition that we shall behave in a certain way, when the occasion arises" (1877, p. 150). Again, a belief, for Peirce, *results* in a disposition, but the disposition can arise from several methods of "fixing" a belief.

First, a belief can be "constitutional" as opposed to acquired. A constitutional belief seems to be instinctive (or, in Churchland's words, "fixed action pattern stuff").

[3] In Peirce's example, the doubt is caused by needing to decide whether to use a nickel or five pennies as streetcar fare.

Acquired beliefs can be fixed by the methods of tenacity or authority, or by the *a priori* or scientific methods.

The method of tenacity consists in "taking any answer to a question which we may fancy, and constantly reiterating it to ourselves, dwelling on all that may conduce to that belief, and learning to turn with contempt and hatred from anything which might disturb it" (1877, p. 151). This method, Peirce notes, does not work well. In addition to the inconveniences resulting from, for example, resolutely continuing to believe that fire does not burn, the method of tenacity falls down before the "social impulse" (ibid., p. 152). We are bound to notice that others have different beliefs, and that their beliefs may be just as good as our own. "Unless we make ourselves hermits, we shall necessarily influence each other's opinions; so that the problem becomes how to fix belief, not in the individual merely, but in the community" (ibid., p. 152).

Tenacity is also reflected in the second method of fixing belief, that of authority. This method allows individuals to share beliefs with their community, because all members of the community adopt their beliefs from the same authority. Even here, though, the beliefs of one community may be different from those of another, simply because the authorities in each community differ. In order to develop beliefs that are true for all communities, we must turn to the *a priori* method, which "promise[s] to deliver our opinions from their accidental and capricious element" (p. 155). This method may run into the same problems as the method of authority, because "though governments do not interfere, sentiments in their development will be very greatly determined by

accidental causes" (p. 155). Without examining these causes themselves, we cannot hope to develop beliefs that can be held universally.

Peirce suggests that the best method of fixing belief arises from "the scientific method". This method introduces the notion of a reality with which all belief must conform in order to be universal. It is the conviction that there is such a reality that makes doubt "a source of dissatisfaction".

The fundamental hypothesis of this method is: "There are real things, whose characters are entirely independent of our opinions about them; those realities affect our senses according to regular laws, and, though our sensations are as different as our relations to the objects, yet, by taking advantage of the laws of perception, we can ascertain by reasoning how things really are and any man, if he have sufficient experience and reason enough about it, will be led to the one true conclusion" (p. 156).

Under this method, a belief is not allowed to stand unexamined, as with the methods of tenacity and of authority. Nor is it justified in isolation from its causes, as is the case with the *a priori* method. Yet a belief fixed by the scientific method is still a belief – a rule of action or a habit.

Peirce's analysis, unlike the kind of dispositional analysis offered by Ryle, allows the division of beliefs according to how they were arrived at by the believer. It also emphasizes that beliefs *are* "arrived at"; in other words, that we believe something because we have had reason to think about it. Even our most trivial beliefs are the result of our attention having been drawn to a subject, inspiring some doubt which we subsequently appease in the formation of beliefs.

In human beings, language plays an important role in the formation (or, in Peirce's terms, the fixation) of our beliefs. Even the method of tenacity requires that we are capable of expressing the belief we adopt, even though we may not express a justification for holding it. Expressing a belief requires that we are able to deal with the belief *as* a belief, or habit of response, rather than simply treat each instance in which we act in a particular way as an isolated occurrence.

Peirce's writings on belief are strongly tied to his writings on semiotic. As Greenlee notes, the notion of habit is crucial to the interpretation of a sign. Deacon's appropriation of the distinctions between icon, index and symbol also has implications for the way in which we describe "belief." In the second section of this chapter, I will describe "beliefs" in terms of Deacon's theory, and show that human cognition is unique in that our beliefs are, at least sometimes, the outcome of symbolic thought, and the expression of this thought in language. The practice of folk psychology is not merely rooted in human biology, it is also a reflection of the symbolic activity that constitutes our humanity.

The Evolution of Opinions

Deacon's thesis, in essence, is that massive changes in the organization of the human brain, which have co-evolved with an expansion in the size of the prefrontal cortex, have enabled us to categorize the world in a way available to no other species. Language, also unique to humans, is merely "the outward expression of an unusual mode of thought – symbolic representation" (Deacon, p. 22). Most theories view folk

psychology as a purely linguistic affair; this is one of the few points (perhaps the only point) upon which the three authors I discussed in the first chapter seem to agree.

If we regard folk psychology as the result, not merely of (public) language use, but of the (cognitive) symbolic capacity underlying language, the whole picture changes. Unlike the theories of Churchland, Fodor and Baker, Deacon's theory of human language use accounts for the radical difference in the cognitive strategies used by humans, compared to those available to other animals. At the same time, Deacon's work addresses the major concerns of all three authors. The use of symbols can be explained in terms of the evolution of the brain. More specifically, symbolic processing relies on the same physiological mechanisms of neuronal firing and synaptic modulation on which Churchland bases his philosophy of mind. While Fodor's notion of a symbol is different from that of Peirce and Deacon,[4] Deacon's description of symbolic reference addresses Fodor's concerns about the syntactic relations between symbols, and about the role of language in cognition.[5] And because a symbol system, in Deacon's sense, relies on lower-level indexical relationships, which link a sign directly with its object, Deacon's theory avoids the methodological solipsism of Churchland and Fodor. Moreover, because a symbol is a sign in virtue of a conventionally agreed-upon interpretation, Deacon's theory can address Baker's characterization of folk psychology as a social practice.

[4] Because of Fodor's methodological solipsism, and his commitment to a computer-processing model of cognition, he believes that the meaning of a symbol reduces to its syntactical relations with other symbols. As Searle phrases it: "they aren't even *symbol* manipulations, since the 'symbols' don't symbolize anything. In the linguistic jargon, they have only a syntax but not semantics" (1980, p. 199).

[5] Though, of course, the conclusions to be drawn from Deacon's theory are opposed to Fodor's own ideas.

Deacon's theory relies, in part, upon the comparison of human cognition with that of other animals, as do those of Churchland and Fodor. Churchland emphasizes the continuity of human thought with that of "lower" animals, and concludes that language has not made much difference to human cognition. We do not need language (or folk psychology) to explain behavior. Fodor agrees that much of human behavior is continuous with that of other animals, but draws the opposite conclusion. All animals (or at least those whose nervous systems are complex enough to permit the representation of the world) have beliefs, because they have a grammatically similar language of thought.

Deacon's account, however, suggests that human cognition is qualitatively different from the kind of thought that other animals engage in. While other animals behave in ways that can often be predicted by positing beliefs and desires, humans are (again, with the possible exception of some of the other higher primates) the only ones who appear to have the capacity to engage in full-blown folk psychology. In order to make this point clear, I will distinguish between adopting the intentional stance and engaging in folk psychology. We take the intentional stance when we attribute beliefs and desires to others in order to explain their behavior. I will reserve the term "folk psychology" for those instances in which we attribute those beliefs and desires to other human beings and ourselves. I will argue that adopting the intentional stance requires nothing more than assuming that one's object of study is capable of iconic and indexical reference. By contrast, practicing folk psychology generally involves the recategorization of the world in symbolic terms. The intentional stance simply requires

that the agent in question have a goal; folk psychology requires that the agent have a worldview.

The concept of the intentional stance is the core of Daniel Dennett's theory of folk psychology. We adopt the intentional stance, according to Dennett, when we explain the behavior of a "system" (the term is that preferred by Dennett) in terms of the beliefs, desires, and so on, that led it to perform a particular action, or when we predict a future action based on the beliefs that we believe the system to have.

The intentional stance, however, applies equally well to any object or organism that responds to the environment. Even a thermostat, to use Dennett's example, can be said to believe that the room is too cold, to want the room to be warmer, to believe that to turn on the boiler will make the room warmer, and so to want to turn on the boiler. While we can accept that this explanation does account for what the thermostat does, it still seems counterintuitive to suggest that the thermostat actually *has* beliefs. Like Fodor, we may want to reserve that claim for animals with the capacity to represent the world. According to Dennett, however, "there is no magic moment in the transition from a simple thermostat to a system that *really* has an internal representation of the world" (1981, p. 74). The difference between the beliefs of thermostats and those of human beings is explained simply by the greater complexity of the latter.

Critics of Dennett's position counter that we (and possibly dogs and cats and apes) *are* different from thermostats, and that this difference is not merely the result of an increase in complexity. Searle (e.g. 1980), for example, describes this difference in terms of our having "original" intentionality, while thermostats and other artifacts have only

derived intentionality. Any form of intentionality that the latter do possess is merely a reflection of the (original) intentionality of their designer. We recognize this distinction in our everyday dealings, as well. When we take the intentional stance toward the behavior of inanimate objects, we consider it to be a metaphorical description.

In the case of animals, we can certainly explain and predict behavior in intentional terms, and often we are able to accurately predict what an animal will do in a given situation by appealing to its beliefs and desires. Intuitively, we think that animals do, in fact, have beliefs. However, their behavior can also be explained in terms of iconic and indexical reference. I will illustrate my point with reference to Fodor's description of his "strikingly intelligent" cat. According to Fodor, "Greycat...has, and acts out of, beliefs and desires. The reason, for example, that Greycat patrols his food bowl in the morning is that he wants food and believes – has come to believe on the basis of earlier feedings – that his food bowl is the place to find it" (1987, p. *x*). This example is exactly analogous to Deacon's description of indexical reference. "Physical contiguity (nearness or connectedness) or just predictable co-occurrence are the basis for interpreting one thing as an index for another" (Deacon, p. 77). Like Deacon's example (discussed in Chapter 2) of expecting to see fire upon smelling smoke, Greycat's expectation of food "is constructed from a set of relationships between icons" (ibid.). The expectation is an indexical interpretation "accomplished by bringing this assembly of iconic relationships to bear in the assessment of new stimuli" (ibid.). Greycat's feeling of hunger (his desire for food) is certainly something he has experienced before, and the current hunger is therefore iconic of past feelings. Similarly, Greycat has found food in his bowl before;

these past occurrences are also icons of each other. A third iconic relationship also plays a part in the development of Greycat's indexical association: past experiences in which the feeling of hunger was correlated with finding food in the bowl also resemble each other. This co-occurrence is a "more general similarity between the present situation and past ones", and is "what brings to mind the missing element in the present case" (Deacon, p. 78) – the expectation that food will be forthcoming.

So far, so good. But what is it that makes Greycat's expectation of finding food in his bowl any less an expectation than my own would be, were I to develop a craving for kibble? The answer is that I, unlike Greycat, would be able to *recognize* my expectation as an expectation. This recognition involves an abstraction from the present situation, and appreciating its similarity at this abstract level, to my feelings in other situations. Greycat's expectation of food is probably a similar experience to his expectation (to borrow another example from Fodor) that scratching at the door will result in his being let out. To abstract even further, his expectation is also similar when, due to fear, "he maintains an appreciable distance between himself and the nearest aggressive dog" (Fodor, 1987, p. *ix*), though in this case, Greycat expects that his actions will result in *preventing* a particular outcome. We can grant that Greycat feels something like what we would call expectation, and something like what we would call fear. It seems much less likely, however, that Greycat is capable of feeling apprehension.[6]

The feeling of apprehension is unlike an expectation in that it is the result of a symbolic association. It requires that we have not merely feelings of expectation and

[6] Not, of course, to be confused with the claim that Greycat is not capable of apprehension

fear, but *concepts* of expectation and fear. These concepts arise from indexical reference in much the same way as indexical relationships arise from iconic reference. Greycat, sitting at his food bowl, would need to step back, mentally, from his expectation, and reflect upon the fact that he is sitting and expecting.

This is only a part of what is involved in symbolic reference; the feeling of "expectation", in order to function as a symbol, must also be able to refer to other symbols. Apprehension is a feeling that we recognize as being part expectation and part fear. According to Deacon's explanation, symbolic reference is built upon indexical reference. However, in order to use symbolic reference, an animal must be capable of suppressing the indexical references of a sign and concentrating instead on the conventional, or symbolic, features of the sign.

Fodor could, perhaps, reply that, while a feeling such as apprehension might be recognized by humans but not by Greycat, this difference can be explained by noting that feline Mentalese is less complex than the human language of thought. According to Fodor, the symbols of Mentalese are innate, and we are certainly more cognitively complex than cats. This objection, however, does not make Greycat capable of symbolic thought; instead it suggests that human reasoning, too, can be explained indexically, which is a doubtful, and somewhat bleak, suggestion. Fodor's claim is problematic, too, because he characterizes cognition (whether human or animal) as a process of hypothesis testing. Yet, indexical reference simply does not support the flexibility required to juggle various hypothetical interpretations of a particular situation. Odds are that Greycat, as he sits by his food bowl, has only one hypothesis in mind.

Human beings, by contrast, have the ability to develop multiple hypotheses. Moreover, these hypotheses are of a different order than the expectations and beliefs that other species are capable of entertaining. "Because of our symbolic abilities, we humans have access to a novel higher-order representation system that not only recodes experiences and guides the formation of skills and habits, but also provides a means of representing features of a world that no other creature experiences, the world of the abstract" (Deacon, p. 423). Whereas animal beliefs can be explained in terms of dispositions and indexical reference, human beliefs are inherently symbolic and can be expressed propositionally. We are aware of having beliefs.

This is the root of my distinction between taking the intentional stance and doing folk psychology. When we describe the beliefs of a thermostat or a cat, we do not assume that the object of our descriptions *knows* that it has these beliefs. When we describe the behavior of other people in the terms of folk psychology, we assume that they are, or can become aware, that they have the beliefs that we attribute to them. Should they not agree, we assume that they can tell us about the beliefs that they *do* have.[7]

According to Dennett (1991), these exchanges are not about beliefs. He claims that we must draw a distinction between beliefs, which are the same in humans and animals (and thermostats), and opinions, which are unique to language users. Since we communicate our beliefs to others, "we have to give an edited version of what we are actually thinking, so what comes out is a few sentences" (p. 89). From there, it is easy to

mistake our sentences for "copies or translations of the very states in the minds of the beings we're talking about" (ibid.), rather than recognizing them for the "edited abstractions or distillations" that they actually are.

Because we must communicate in propositions, Dennett suggests, it is easy to fall prey to "Fodor's industrial-strength Realism" (R.P., p. 102 in Lycan), which is "the illusion of discrete, separate, individuable beliefs" (1991a, p. 88). Despite this characterization, Dennett does not go so far as to say that opinions are unreal. Rather, they are "real abstractions", and reflect our awareness of what Dennett describes as "the brute existence of pattern" (1991b, p. 120). Moreover, this position seems to be in accord with Fodor's insistence on the autonomy of psychological explanations.[8] Dennett suggests that "the psychology of opinions is really going to be rather different than the psychology of beliefs and that the sorts of architecture that will do very well by, say, nonlinguistic perceptual beliefs (you might say 'animal beliefs') is going to have to be supplemented rather substantially in order to handle opinions" (1991a, p. 90).

The question that Dennett does not address, however, is whether our "opinions" make a difference to our beliefs. Rather, he sees the intentional stance as a useful strategy from which to predict and explain the behavior of both animals and humans. In fact, the adoption of the intentional stance toward "oneself and one's fellow intelligent beings" is unavoidable (1987, p. 27), primarily because no other level of explanation works so well.

---

[7] Of course, people can be mistaken about their motives, or deceive themselves as to their beliefs. Even this capacity, though, is the result of symbol use.

[8] Though Dennett claims that Fodor's Realism is itself a "gratuitously strong" form of materialism (1991b, p. 120).

According to Dennett, the reason that the intentional stance works as well as it does is simply that "we are products of a long and demanding evolutionary process [that] guarantees that using the intentional stance on us is a safe bet" (ibid., p. 33). Evolution has ensured that we are generally rational creatures, who believe what we ought to believe, and want what we ought to want. If this were not the case, we would not have survived.

There are two reasons why this answer is not satisfactory. First, it does not tell us how those beliefs and wants are "regulated by the internal machinery" (ibid., p. 34). Deacon's theory of brain-language co-evolution does give us such an account, but without recourse to Fodor's problematic "Realism". More importantly, Deacon's theory also addresses the second problem with Dennett's suggestion that we evolved, by and large, to be rational creatures. Quite simply, we are often irrational. Moreover, Dennett's "rationality" refers to our possessing the beliefs needed to fulfill such Darwinian goals as finding food, avoiding becoming food for someone else, and producing and feeding offspring. The human world is much more complex than Dennett's account would have it. As Deacon points out, "Biologically, we are just another ape. Mentally, we are a new phylum of organisms. In these two seemingly incommensurate facts lies a conundrum that must be resolved before we have an adequate explanation of what it means to be human" (p. 23).

Deacon describes the human world as "a virtual world, full of abstractions, impossibilities and paradoxes" (p. 22). This virtual world is the one in which we experience those characteristics and preoccupations which we associate with being

human. The sense of self, of morality, and the search for meaning and purpose in the world all require symbolic representation. "In a real sense, we live our lives in this shared virtual world" (ibid.).

Symbols, Selves and Personhood

It is in this shared virtual world, the product of our symbolic representations of reality, that folk psychology is practiced. When we attribute beliefs to others, or talk about our own, we are referring to the kinds of abstract distinctions that differentiate my behavior from Greycat's. The ability to do that, however, is not merely an example of the human capacity for symbolic reference; it is also an outcome of it.

According to Baker, it is the ability to describe one's actions from the first person perspective that makes humans unique among animals. She distinguishes between being a human being, which refers to our biological continuity with the rest of the animal kingdom, and being a human person, which is the essential characteristic of a human individual. On her Constitution View, "*person* is an ontological kind" (Baker, 2000, p. 11). While she does not directly link her theory of personhood with her earlier work on folk psychology, she does acknowledge that "attitudes – like believing, desiring and intending – should be understood not primarily as brain states but as states of whole persons" (2000, p. xi).

In Baker's "Constitution View", a person is distinct from her body, but is constituted by it, in much the same way that a statue is not identical with the block of marble of which it is made. The characteristic that distinguishes between a human body

and a human person is the ability to take a first-person perspective. "Whether or not *x* is a human being depends on biological facts about *x*; whether or not *x* is a human person depends additionally on psychological facts about *x* – namely, on the Constitution View, on whether or not *x* has a capacity for a first person perspective" (ibid., p. 8).

Baker also distinguishes between weak and strong first-person phenomena. Animals are capable of a first-person perspective in the weak sense of the term. This sense requires merely that an experience be had from a particular physical point of view. "We attribute beliefs and desires (perhaps in the vocabulary of aversions, appetites, and learning states) to nonhuman animals, which seem to be reasoning from a certain perspective" (ibid., p. 61). For example, Greycat positions himself near his food bowl when he is hungry, because he is hungry and believes that food will be arriving shortly. Yet "although such an animal has beliefs and desires, he has no conception of belief or desire, or of himself as the subject or bearer of beliefs and desires. He acts from his own perspective without any conception of having a perspective that differs from other perspectives" (p. 62).

A strong first-person perspective requires not only that one can "*distinguish* between first person and third person; one must also be able to *conceptualize* the distinction, to conceive of oneself as oneself" (ibid., p. 64). The conditions for a strong first-person perspective are met when the word "I" cannot be eliminated from descriptions of a situation (ibid., p. 76). Greycat's "I believe it's dinner time" can be rephrased as "Dinner soon." By contrast, one cannot eliminate the second "I" in the thought that "I think that I will take a quick nap after dinner."

"From a first-person point of view," Baker claims, "one can think about oneself as oneself and think about one's thoughts as one's own" (p. 91). She also notes that this claim is different from the claim that an organism has intentional states, because "there are many intentional states that humans and other animals can have without being able to conceive of themselves in the first person" (ibid.). While having intentional states is necessary for being a person, it is not sufficient. The difference between the two cases is illustrated in the difference between having the belief that *x*, and knowing that "I believe *x*".

Yet Baker also notes that having the ability to take a first-person perspective is not the same as having a sense of self. She describes the latter as a "much thicker concept" (p. 87). Dennett, for example, describes the self as "a center of narrative gravity" (ibid., see also Dennett 1991c, p. 410). Baker agrees that "[a] self is that locus of personal integrity and coherence", but adds that "such a self is not required for a first person perspective" (Baker, 2000, pp. 87-8).

In fact, the construction of a self in this robust sense depends on the ability to take a first-person perspective. The question is, however, which "level" of selfhood we should take to be sufficient for the ascription of personhood. Dennett suggests that there are several characteristics that we associate with personhood, including rationality, moral agency, and "being *conscious* in some special way" (1978, p. 270), which is generally interpreted as being self-conscious. All of these concepts, though, are problematic if we wish to claim that they are sufficient conditions for personhood.[9] According to Dennett,

[9] Dennett also claims that all of these are necessary conditions for personhood as well.

"the concept of a person is...inescapably normative. Human beings or other entities can only aspire to being approximations of the ideal, and there can be no way to set a 'passing grade' that is not arbitrary" (ibid, p. 205).

It seems to me that Baker's approach to defining personhood is the better one. Rather than making personhood inherently normative, her definition suggests that personhood is the basis for the normative concepts described by Dennett. In this, Baker's Constitution View is compatible with Deacon's theories. The capacity for a first-person perspective was probably one of the earliest outcomes of our evolving symbolic capabilities. Deacon notes that the ability to learn symbol use has become universal in the human species. "Though there may be variations in this ability among people, essentially all of this variability is above the threshold for acquiring symbols" (p. 412). Both the first-person perspective and symbolic reference represent qualitative differences in cognition; on the human side of these thresholds many new possibilities open up. Because there are qualitative differences among individuals in the ability to manipulate symbols, the behaviors discussed by Dennett occur along a spectrum.

One of the new possibilities that symbolic reference makes possible is the practice of folk psychology. We often find ourselves describing people's actions in terms of their beliefs and desires. Fodor emphasizes the extent to which we are able to make accurate predictions about what others will do, based on this type of analysis. Deacon acknowledges this, albeit in a somewhat more cynical fashion. "In our everyday dealings with one another, we constantly try to anticipate others' behavioral plans or decision-making processes, either just to get along with them, or else the better to manipulate

them" (p. 425). The reason, according to Deacon, that folk psychology works so well is that our behavior is generally governed by convention. "Learned social habits provide a sort of unanalyzed folk psychology that gets us by" (ibid.).

Folk psychology, it seems, is a sort of self-fulfilling prophecy. As we learn to use folk psychology to describe the behavior of others, we are also learning to behave in accordance with the standards it imposes. In Peirce's terms, many of the beliefs governing social behavior are fixed by the method of authority. Even in the (relatively rare) instances in which folk psychology goes badly wrong, it is the social norms and practices expressed in folk psychological terms that allow us to judge someone's behavior as inappropriate or simply unusual. "Applications of common sense psychology mediate our relations with one another, and when its predictions fail these relations break down. The resulting disarray is likely to happen in public and to be highly noticeable" (Fodor, 1987, p. 2).

At the same time, not all of our behavior can be described as social habit, and, because of this, folk psychology is prone to revision. It is possible both to create new intentional categories, and to form beliefs about novel propositions. Not long ago, no one had beliefs about the desirability of globalization or of environmental protection legislation. Now, it is becoming increasingly important that we decide what we believe about these issues, and quickly. Similarly, it was not possible, not too long ago, to "be in denial about *x*", or to "feel conflicted over *y*."

Despite the effect of pop psychology on our understanding of ourselves, the proliferation of intentional attitudes, and concepts and events to which we apply them,

generally reflects the increasing sophistication of the individuals and societies that use them. It is at this level that Dennett's normative criteria come into play. We can never be perfectly rational or moral because these concepts themselves evolve as we use them. The same goes for folk psychology. "The ability to generate models of others' emotional states, and to exercise restraint or determination with respect to them, must be discovered through the manipulation of symbols. This makes such interpretations conditional on the maturation, experience, and even cognitive capacities of people" (Deacon, pp. 431-2).

It is folk psychology's unlimited flexibility and capacity for growth that make it such a powerful tool. One of the hotly contested questions in the philosophy of mind is whether folk psychology is a theory. As I explained in Chapter 1, Churchland and Fodor say it is; Baker claims that it is not). The best answer, though, is probably: "yes and no." It is a theory in that it is used to predict and explain human behavior. It is not a theory in that it is partially constitutive of the behavior it describes. In Deacon's terms, this aspect of folk psychology involves the creation of novel symbolic relations. In Peirce's terms folk psychology fixes belief using something like the scientific method. It is the active attempt to understand the causes of human action by describing its regularities. Peirce, however, did not distinguish between the purely physical world that human beings share with other animals and the "virtual world" available to humans alone. For Peirce, thought was as real and objective as any other force in the universe, whether that thought was iconic, indexical or symbolic in character.[10]

[10] Hoopes (1991) describes this position as "semiotic realism, [the] notion that thought is in visceral signs requiring inferential interpretation" (p. 160).

Deacon, however, suggests that symbolic reference is unique to human beings, and this claim alters somewhat the scientific method of fixing belief. According to Peirce, "*True* is an adjective applied solely to representations and things considered as representations. It implies the agreement of the representation with its object" (1861, p. 20). In the virtual world of symbolic reference, it is possible for more than one representation to be true; the same object or event can often be described in multiple, equally true, ways.

The virtual world "is governed by principles different from any that have selected for neural circuit design in the past eons of evolution" (p. 423). In order to live in this world, we have developed principles to help us understand it and to understand the behavior of the other people in it. At the same time, though, this abstract, psychological world is not completely divorced from biological reality. "The symbolic constructions of others' plausible emotional states, and their likely emotional responses to our future actions, are analogous to a whole new sensory modality feeding into our ancient social-emotional response system" (Deacon, p. 431). As a result, we often do not act in the evolutionarily rational way that Dennett describes. We are motivated by all manner of diverse goals, many of which are not recognizably Darwinian.[11] Most of what we do is likely a result of both abstract and concrete representations of the situation to which we are responding.

[11] Sociobiologists, however, claim that these novel goals are simply disguised forms of older imperatives, and that such diverse human concerns as rape, ambition, altruism and fashion exist solely to increase an individual's chances of reproductive success. For the case against sociobiology, see, for example: Rose, Lewontin and Kamin (1984), Chapter 9; Fausto-Sterling (1986), Chapter 6.

It is not always easy, however, to link the abstractions we make with more concrete indexical representations. Deacon suggests that "[the] ability to let our emotions be activated by the virtual experiences constructed with the aid of symbols probably makes us the only species in which there can be a genuine conflict of simultaneous emotional states" (ibid.). While symbolic reference allows us to develop a wide range of novel behaviors, these behaviors are often in conflict with older evolutionary imperatives. At a purely symbolic level, conflict also arises when alternative representations of a situation suggest different courses of action.

Folk psychology is, in its most important form, an effort to make sense of these conflicting mental states, to, as Peirce would say, move from doubt to belief. It is a continuing social project that attempts to make sense of the multiplicity of possible interpretations of the world and to determine which interpretation we should choose to act upon. Deacon notes that the prefrontal cortex is most active during "the *construction* of the distributed mnemonic architecture that supports symbolic reference" (p. 266), or, in Peirce's terms, during the development of a habit of action.

This process of symbol-making is the deepest and most important sense of having beliefs. The only way to cope with the confusion created by the symbolic understanding of the world is to use more symbols. Using symbols, in fact, is just what Deacon claims that human brains were designed to do. "[T]he neuroanatomical evidence of massively altered brain proportions and the anthropological and clinical evidence for universality of symbol learning across a wide spectrum of circumstances indicate that the human brain has been significantly overbuilt for learning symbolic associations" (p. 413). As

described in Chapter 2, the prefrontal cortex is involved in a large number of the synaptic connections in the brain. Because of this, "[s]ymbolic abilities were recruited to serve other social and pragmatic functions than those which selected for their initial appearance" (ibid.). Not only can we form beliefs about appropriate social behavior and which action will lead us to our goals, we can form beliefs about just about anything. Folk psychology began as a way of symbolically ordering the world; it is only derivatively a way of making truth claims about grass and kangaroos. We are able to entertain beliefs about all manner of trivial things because we have learned to form beliefs about the things that are important, primarily the behavior of ourselves and other people.

As Churchland, Fodor and Baker all recognize, folk psychology is intimately linked to the use of language. Yet, as Deacon suggests, language is itself merely the outward sign of symbolic thought and it is in this deeper context that folk psychology must be understood. When we describe behavior in terms of the beliefs and desires that give rise to it, we are looking at abstract patterns in behavior. These abstract patterns, however, are just as "real" as anything else; they are a higher-order description of the world, not merely a convenient fiction.

Deacon suggests that human beings are "the symbolic species" but it is important to realize that the use of symbols is intimately linked with the more concrete forms of reference that characterize cognition in other animals. While Deacon talks about humans' unique access to a virtual world, this does not mean that the virtual world is imaginary. It is merely abstract.

A philosophy of mind, as I have said before, should begin with our phenomenological experiences. These experiences, however, are influenced by the concepts we acquire from the social environment. Because of this, it is perhaps most important that a philosophy of mind can account for our ability to experience the world in so many different ways. Deacon's account of symbolic reference suggests the biological basis for this ability.

As for folk psychology, the practice is more complex than it is generally regarded to be by philosophers. Churchland criticizes it for failing to match what we know of the neurological mechanisms underlying cognition, while Fodor tries to force all varieties of cognition to match folk psychological descriptions. Neither author makes a proper distinction between human cognition, with its strong linguistic bias, and the kinds of cognitive processes governing behavior in other animals. Nor do they acknowledge that even in human beings there are many different kinds and levels of belief.

Because Churchland is interested solely in neurobiological explanations of mental phenomena, he seems to be willing to sacrifice the richness and variety of our mental lives to fit biological theories. Fodor, on the other hand, argues for the autonomy of psychological explanations. While this approach makes it possible for a theory of mind to be rooted in our commonsense conception of mind, Fodor's explanation is no more adequate than Churchland's. Fodor (Fodor and Pylyshyn, 1988) suggests that the connectionist models Churchland discusses may represent the "hardware" of the brain,

but that LOT is required to account for the properties possessed by the "software" of the mind. Neither Churchland nor Fodor can offer any way of bridging the gap between these two levels of explanation.

The problem of this "gap" arises because both Fodor and Churchland are too adamant about the primacy of their respective levels of explanation. Deacon, on the other hand, constructs his theory using evidence from a variety of disciplines, including anatomy, anthropology and psychology. Because of his multidisciplinary approach, Deacon is able to show that the gap between brain and mind is quite possibly a mirage. The mental capacities that allow human beings to, among other things, practice folk psychology arise from our biological makeup. However, they are not best explained, as Churchland insists, solely in terms of our physiology. A scientific theory of mind must show how physiology and anatomy and biochemistry and psychology work together to explain different aspects of the same phenomena.

Similarly, science (taken as a whole) and philosophy also ask similar questions about the mind, though answering them in very different ways. Both kinds of explanation are important. Baker emphasizes that folk psychology is a social practice, and suggests that it is best explained by analyzing the meaning of the concepts and terms it uses. What she does not address, at least in her earlier work, is what it is that makes us able to *use* these concepts. In her most recent book, however, she discusses the importance of the ability to take a first-person perspective and suggests that this ability is linked to the ability to entertain propositional attitudes.

In fact, a first-person perspective is the foundation of this ability, and both are the result of symbolic thinking. Because we have crossed "the symbolic threshold" we are able to conceive of ourselves *as* ourselves and also create complex intentional descriptions about ourselves. Once we pass the initial threshold, the systems of symbols we can construct become more and more elaborate and abstract.

As Dennett points out, there is a gap between recognizing ourselves as the subjects of intentional states, and statements and being ideally rational or moral. Doing folk psychology is an important part of becoming a rational and moral person, of creating a self. Any philosophy of mind that wishes to account for the biological, psychological and social aspects of this process must examine the role of folk psychology in the development of the human mind.

BIBLIOGRAPHY

Abiotiz, Francisco and Ricardo Garcia. 1997. "The Evolutionary Origin of the Language Areas in the Human Brain. A Neuroanatomical Perspective." Brain Research Reviews 25, 381-96.

Baker, Lynne Rudder. 1995. Explaining Attitudes: A Practical Approach to the Mind. Cambridge: Cambridge University Press.

-------. 2000. Persons and Bodies: A Constitution View. Cambridge: Cambridge University Press.

Churchland, Paul. 1988. Matter and Consciousness: A Contemporary Introduction to the Philosophy of Mind. 2nd ed. Cambridge, Mass,: MIT Press.

-------. 1989a. "On the Nature of Theories: a Neurocomputational Perspective." In Mind Design II: Philosophy, Psychology, Artificial Intelligence. (1997) Ed. John Haugeland. Cambridge, Mass.: MIT Press.

-------. 1989b. A Neurocomputational Perspective: The Nature of Mind and the Structure of Science. Cambridge, Mass.: MIT Press.

Churchland, Paul and Patricia Churchland (1984). "Stalking the Wild Epistemic Engine." In Mind and Cognition: an Anthology 2nd Ed. (1999) Ed. William G. Lycan. Oxford: Blackwell Publishers.

Damasio, Antonio R. "Aphasia." The New England Journal of Medicine. February 20, 1992, pp. 531-539.

Deacon, Terrence. 1997. The Symbolic Species: The Co-evolution of Language and the Brain. New York: W.W. Norton.

Dennett, Daniel C. 1978. "Conditions of Personhood." In Brainstorms: Philosophical Essays on Mind and Psychology. Cambridge, Mass.: MIT Press.

-------. 1978. "A Cure for the Common Code." In Brainstorms.

-------. 1981. "True Believers: The Intentional Strategy and Why it Works." In Mind Design II.

-------. 1991a. "Two Contrasts: Folk Craft versus Folk Science, and Belief versus Opinion." In Brainchildren: Essays in Designing Minds.

-------. 1991b. "Real Patterns" In Brainchildren.

-------. 1991c. Consciousness Explained. Boston: Little, Brown and Co.

Edelman, Gerald M. 1989. The Remembered Present: A Biological Theory of Consciousness. New York: Basic Books Inc.

Fausto-Sterling, Anne. 1992. Myths of Gender: Biological Theories about Women and Men. Revised Edition. New York: Basic Books Inc.

Fodor, Jerry. 1968. Psychological Explanations. New York: Random House.

-------. 1975. The Language of Thought. Hassocks, Sussex: The Harvester Press.

-------. 1985. "Fodor's Guide to Mental Representation: the Intelligent Auntie's Vade-Mecum". In The Future of Folk Psychology: Intentionality and Cognitive Science. Ed. John D. Greenwood. Cambridge: Cambridge University Press. 1991.

-------. 1987. Psychosemantics: the Problem of Meaning in the Philosophy of Mind. Cambridge, Mass.: MIT Press.

Fodor, Jerry and Zenon Pylyshyn. 1988. "Connectionism and Cognitive Architecture: A Critical Analysis." In Mind Design II.

Frith, Chris D. and Uta Frith. 1999. "Interacting Minds – A Biological Basis." Science, Vol. 286, November 26, 1999, pp. 1692-1695.

Fuster, Joaquin M. 1989. The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobes. 2nd Ed. New York: Raven Press.

Gabrieli, John D. E., Russell A. Poldrack and John E. Desmond. 1989. "The Role of Left Prefrontal Cortex in Language and Memory." Proceedings of the National Academy of Science, U.S.A. Vol. 95, February, pp. 906-913.

Greenlee, Douglas. 1973. Peirce's Concept of Sign. The Hague: Mouton & Co.

Hailman, J.P. 1969. "How an Instinct is Learned." Scientific American 221. No. 6, pp. 98-108.

Hoopes, James. 1991. Peirce on Signs: Writings on Semiotic by Charles Saunders Peirce. Chapel Hill: University of North Carolina Press.

Nixon, Greg. 1998. Review of Terrence Deacon's The Symbolic Species: The Co-Evolution of Language and the Brain. Journal of Consciousness Studies 5 (5/6), pp. 746-8.

Peirce, Charles S. 1861. "A Treatise on Metaphysics." In Peirce on Signs: Writings on Semiotic by Charles Saunders Peirce. Ed. James Hoopes. Chapel Hill: University of North Carolina Press. 1991.

-------. 1877. "The Fixation of Belief." In Peirce on Signs.

-------. 1878. "How to Make our Ideas Clear." In Peirce on Signs.

-------. 1897. "Logic as Semiotic: the Theory of Signs." In Introducing Semiotic: An Anthology of Readings. Ed. Marcel Danesi and Donato Santeramo. Toronto: Canadian Scholars' Press. 1992.

Pribram, Karl. 1985. "The Subdivisions of the Frontal Cortex Revisited." In The Frontal Lobes Revisited. Ed. Ellen Perecman. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Rose, Steven, R.C. Lewontin and Leon J. Kamin. 1984. Not in Our Genes: Biology, Ideology and Human Nature. London: Penguin Books.

Ruse, Michael. 1998. "The Evolution of Symbols." The Semiotic Review of Books, Volume 9(3).

Ryle, Gilbert. 1949. The Concept of Mind. Chicago: University of Chicago Press.

Searle, John R. 1980. "Minds, Brains and Programs." In Mind Design II.

Sprague, Elmer. 1999. Persons and their Minds: A Philosophical Investigation. Boulder: Westview Press.

Turner, Mark. 1998. "Poetry for the Newborn Brain" Bostonia, Spring, Number 1, pp. 72-3.