The Anatomy of a Minimal Mind

Vincent Joseph Torley

Submitted in total fulfilment of the requirements
of the degree of Doctor of Philosophy
February 2007

Philosophy Department
The University of Melbourne

Abstract

This thesis is entitled “The Anatomy of a Minimal Mind”. By a “minimal mind” I mean the simplest kind of mind that could exist. As there is widespread philosophical disagreement about what a mind is, or what mental states are, I refrain from assuming at the outset that a minimal mind has to be phenomenally conscious, or subjectively aware of events in its surroundings. My objective is to identify the requirements that an organism would have to satisfy, before it could be credited with possessing a mind of any sort, however rudimentary. I then attempt to develop a detailed model of this minimal mind, using a conservative methodological approach: we should not interpret an organism's behaviour as a manifestation of underlying mental states unless doing so enables us to make better scientific predictions about its behaviour and/or explain its behaviour more fully.

In section A, chapter one, I discuss the philosophical background to the contemporary discussion of minds and mental states. I defend the controversial claim that only *living things* can be said to have minds or mental states, and I argue that there are no convincing grounds for rejecting the common view that mental states are real phenomena. This in no way implies the more controversial view that all mental phenomena share some distinguishing feature that characterises them as mental. If, however, there is some distinguishing property which is common to *all* mental states, and *only* those states, then the two most promising philosophical candidates for this property would surely be *consciousness* and *intentionality.* I argue that the different varieties of consciousness distinguished by philosophers fail to “carve nature at the joints”. I then analyse the strengths and weaknesses of three common definitions of intentionality. Lastly, I examine Dennett’s intentional stance and argue that the two ways in which it can be formulated are in fact quite distinct. I propose that *one* of these formulations, which I refer to as the agent-centred intentional stance, can be used to help us identify creatures with minimal minds.

In section B (chapters two to eight), I attempt to identify the necessary conditions for intentional agency in creatures, by examining several *broad categories* of behavioural and biological properties that have been proposed in the philosophical and scientific literature as relevant to having a mind, and sifting through them, all the while attempting to put together a *constructive definition* of a "minimal mind". In particular, I discuss sensory capacities (including discriminatory ability and perception); memory; flexible behaviour patterns; the ability to learn; self-directed movement and control; the ability to correct one's mistakes; and the ability to form concepts. Within each category of "mind-relevant" properties, I examine the different ways in which these properties are realised by different kinds of organisms, at various levels of complexity. The

biological case studies that I discuss range from the relatively simple (viruses) to the most complex (vertebrates, especially birds and mammals).

In section C, I list about a dozen detailed conditions that an animal has to meet before it can be said to possess this kind of "minimal mind", which, I argue, is the most basic kind of mind anything can have. Perhaps the most crucial condition is that the animal possess an internal “minimal map" by which it represents the means it has to adopt to achieve its ends, enabling it to steer itself around its environment. I argue that animals whose maps are of the right sort can be said to have beliefs, desires and intentions.

Finally, I claim that these "minimal minds" come in no less than four different varieties. Operant agency, navigation using visual landmarks, tool use and the social practice of following a guide are all behaviours that manifest mental states. Although these states are not phenomenally conscious states, I argue that the intentionality they possess is fundamentally the same as that found in conscious mental states. In the end, I conclude that many insects and spiders, as well as octopuses and squid, and of course fish, qualify as having minimal minds.

Declaration

This is to certify that:

(i) the thesis comprises only my original work toward the PhD except where indicated in the Preface;

(ii) due acknowledgement has been made in the text to all other material used;

(iii) the thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Vincent Torley

Acknowledgements

It is with a sense of humility that I offer my thanks to all those who have assisted me in the preparation of this thesis. First and foremost, I would like to acknowledge a deep debt of gratitude to my dearest wife, Futaba, for her encouragement, faith and selfless acts of kindness, which enabled me to see this project through to the end. Without her generosity of spirit, this thesis would not have been possible. Above all others, it was the presence of my wife and our baby son Ken that sustained me in the final stages of this thesis, and gave me the determination to finish it.

I would also like to thank my mother, Mrs Jill Torley, for taking an ongoing personal interest in the direction of my thesis, for assisting me with my research on animals and for regularly sending me clippings of topical articles through the mail. My mother's ideas and thoughtful comments helped shape the content of this work.

I wish to thank my mother-in-law, Mrs Ayako Soshi, for her moral support and patience with me during the last four years, as she generously provided me with a room and a phone line for Internet access.

My brother-in-law Tatsundo also deserves special thanks for letting me use his personal computer to do online research and to edit my thesis online.

My father Mr John Torley and my brothers Christopher, Philip and Paul all encouraged me to finish this thesis, but Philip deserves special thanks for arranging the printing of the draft version of this thesis with the University of Queensland printery, and for sending it to the School of Graduate studies at the University of Melbourne.

I would especially like to thank my thesis supervisor, Professor Brian Scarlett, whose insightful comments sharpened the presentation of my arguments, and whose generosity with his time went far beyond the call of duty. I would also like to thank Dr Francois Schroeter for taking over the supervision of this thesis during a brief period when my current supervisor had to go on leave. Finally, I would like to thank Professor Barry Taylor, the Head of the Philosophy Department, for encouraging me to finish this work.

I would also like to thank Mr Ryan Phinisee, my business partner, for his encouragement and moral support during the final stages of this thesis, when he was kind enough to lend me his computer after mine had given up the ghost. Mr James Bower also deserves special thanks for converting this thesis into a PDF document and for his technical assistance with images.

Many thanks are due to the academics who were kind enough to lend me their time and answer my scientific queries - especially those relating to chapters one through four of this thesis. I would like to thank Dr Charles Abramson, Dr Guenther Albrecht-Buehler, Dr Yehuda Ben-Shahar, Dr Bjorn Brembs, Dr Michel Cabanac, Dr Eric Chudler, Dr Rodney Cotterill, Dr Josh Dubnau, Dr David Edelman, Dr Randy Gallistel, Dr Bertram Gerber, Dr Martin Heisenberg, Dr Martin Giurfa, Dr Robert Kentridge, Dr Steven Laureys, Dr Joseph LeDoux, Dr Bernd S. Muller, Dr Lauren Myers, Dr Jennifer Mather, Dr Randolf Menzel, Dr Bill Nuttley, Dr Jaak Panksepp, Dr James David Rose, Dr Dan Ryder, Dr Catharine Rankin, Dr George Stefano, Dr Giulio Tononi, Dr Kirk Visscher and Dr David Wood.

I would also like to offer my deepest thanks to the staff at Atsugi Nishi high school, Ohara high school, Hadano Minamigaoka high school and Shonan Gakuen in Japan, for allowing me to use their computer facilities for my thesis research and writing. Without their generosity, this thesis could not have been completed.

This thesis is dedicated to all creatures, great and small.

Beyond the shadow of the ship,
I watched the water-snakes:
They moved in tracks of shining white,
And when they reared, the elfish light
Fell off in hoary flakes.

Within the shadow of the ship
I watched their rich attire:
Blue, glossy green, and velvet black,
They coiled and swam; and every track
Was a flash of golden fire.

From *The Rime of the Ancient Mariner* by Samuel Taylor Coleridge.

Table of Contents

Section C: The Conditions for Intentional Agency

List of Tables, Figures and Illustrations

Section A: Background to the Problem of Identifying Minds in Other Creatures

Introduction

0.0 What this thesis is about

This thesis is entitled “The Anatomy of a Minimal Mind”. By a “minimal mind” I mean the simplest kind of mind that could exist. I attempt to identify the requirements that an organism would have to satisfy, before it could be credited with possessing a mind of any sort, however rudimentary. I then attempt to develop a detailed model of a minimal mind, in which I describe what kinds of features the *most basic* kind of mind would have to possess – hence the title of this thesis.

In this chapter, I address several questions pertaining to methodology:

- What kinds of entities fall within the scope of this thesis?
- What assumptions do I make about “minds” and “mental states”?
- How should we go about identifying the occurrence of mental states in other organisms?
- What are the appropriate sources of evidence that would justify the attribution of mental states to other kinds of organisms?
- How generous should we be in assessing claims for mental states in other organisms? Putting it another way, where does the onus of proof lie?
- What terminological conventions will I follow when describing minds and mental states?

0.1 Scope of this thesis

I have chosen to confine my philosophical quest for a minimal mind to *living things.* I am aware that this leaves me open to a charge of methodological narrowness, as I shall be excluding the feats of man-made robots and supercomputers from my definition of mind. I reject this charge, for two reasons, which I elaborate in chapter one.

My first argument is a *reductio ad absurdum.* I contend that once we start attributing mental states to non-living systems, we end up with an infinite regress, leading to panpsychism – a view I consider philosophically defective, as it renders us incapable of distinguishing between messages and their senders.

The second argument which I put forward is that non-living systems, unlike living things, are incapable of qualifying as true individuals, since they lack ends of their own. Hence, they cannot be said to have minds of their own. Specifically, I defend the claim that the distinction between intrinsic and extrinsic finality is a clear-cut one, and that living things exemplify the former kind of finality, while contemporary artifacts, including robots and computers, instantiate the latter. I concede that future artifacts may well possess intrinsic finality, but I contend that if they did, there is no good reason why we should not regard such artifacts as alive.

0.2 Assumptions I make regarding minds and mental states

0.2.1 Eliminative materialism and the reality of the mental

Of course, the quest for a minimal mind would be a fool's errand if it turned out that there were no such things as "minds" or "mental states". According to eliminative materialists, the very idea that we have "mental states" of any sort (e.g. beliefs or desires) is an *illusion* created by our uncritical acceptance of what these philosophers derisively refer to as "folk psychology" – roughly, our everyday understanding of mental states. Proponents of *eliminative materialism* (henceforth **EM**) propose that we simply abandon our attempts to account for human and animal behaviour in terms of the alleged "mental states" of the individuals concerned. The only valid explanations of events are those that are physicalistic as well as causal. The implications for *animals* are obvious: none of them can possibly have *beliefs* and *desires* if these terms do not refer to anything real. Nor can there be a *minimal* mind, if there are no minds at all in the natural world.

In chapter one, I briefly examine the contemporary debate regarding EM, and conclude that there are no good reasons for saying that folk psychology is incompatible with physicalism, or with what we currently know about neuroscience or psychology. However, even if EM is wrong and folk psychology is a valid way of describing *human* minds, it does not follow that our folk theory of *animal* minds is correct. To complicate matters further, different human cultures ascribe different levels of mental sophistication to animals and other organisms; some even ascribe mental states to all natural objects (animism). Human predictions about animal behaviour are also notoriously fallible.

We therefore need to acknowledge at the outset of our enquiry that the most appropriate *scientific* terminology for talking about other animals' mental states – which is what we are trying to construct here, on the basis of empirical evidence – need not map onto our everyday discourse about animals. We may have to jettison most of what we believe about animal minds.

At the present time, scientists tend to eschew mentalistic terms when discussing other animals, partly because of a long-standing view that these terms denote private (and hence unknowable) inner states, and partly because there are no currently accepted scientific criteria for identifying the states they correspond to, in animals. The terminology that scientists employ when talking about animals reflects their goal of systematically predicting and explaining animal behaviour, within the methodologies of their disciplines. However, in this thesis, I shall argue for the ascription of mental states to non-human animals, precisely on the grounds that scientists cannot model and predict their behaviour satisfactorily without positing these states. In other words, EM is scientifically counter-productive, even when applied to other animals.

0.2.2 The unity of the mental

One could also challenge the existence of “minds” and “mental states” in a less radical way, by arguing that even if the states that philosophers collectively dub “mental states” *are* real, the category itself is an artificial one, as mental states do not possess any *common characteristic* which can serve to identify them as such. In recent times, Rorty (whose own views place him in the EM camp) has criticised the apparent arbitrariness of philosophical attempts to unify these phenomena under the “mental” umbrella, arguing that “the attempt to hitch pains and beliefs together seems ad hoc – they don’t seem to have anything in common except our refusal to call them ‘physical’” *(*1979, p. 22). The current fashion of designating pains, itches and other sensations as “mental states” would have seemed even more alien to Aristotle and Aquinas, both of whom, when they used the term “mind” (*nous* and *mens* in their respective languages), tended to restrict it to what we would call intellect or understanding (O’Callaghan, 2000; Sorabji, 1993).

Because the “unity of the mental” is a philosophically controversial issue, it will *not* be taken as a “given” in this thesis. Since this thesis is meant to be a *constructive* enquiry, I shall deliberately refrain from attempting to define “mind” or “mental states” at the outset, and thereby committing myself to tendentious philosophical positions that may prejudice my enquiry. My aim here is to allow the property which underlies or unifies the states we currently designate as “mental”, whatever it may be, to emerge in the course of my investigation.

To date, the only promising candidates for unifying the suite of phenomena that philosophers refer to as “mental states” are the features of *consciousness* and *intentionality*. The definition of the terms “consciousness” and “intentionality”, as well as the issue of the relation between them, remains an area fraught with controversy. Accordingly, I shall devote the first chapter of my thesis to: outlining various definitions of consciousness in the philosophical and scientific literature, and discussing their applicability to animals as well as any other organisms that may possess a minimal mind; surveying the history of philosophical attempts to define intentionality and the problems these attempts have generated, with reference to non-human organisms; and reviewing alternative views of the relation between consciousness and intentionality. My discussion will be principally focused on issues that are germane to the question of what a minimal mind might be, and what kinds of minds non-human animals, which lack language, could be said to possess.

0.2.2.1 Consciousness as a hallmark of mental states

The notion that *consciousness* is what underlies mental states is commonly traced back to Descartes, although as we shall see in the following chapter, his usage of the term differed from our own in several important ways. Contemporary philosophers distinguish several different senses of “consciousness”. When most people use the term “conscious”, they use it to denote a state of subjective awareness, with an indefinable feeling of “what it is like”. This variety of consciousness is what philosophers call phenomenal consciousness. Block (1995) defined *phenomenally conscious* states as states with a subjective feel or phenomenology, which we cannot define but we can immediately recognise in ourselves, distinguishing them from *access conscious* states, or mental representations which are (i) poised to be used as a premise in *reasoning*, (ii) poised for rational control of *action*, or (iii) poised for rational control of *speech.* (Block (2005) has since amended his definition: the key feature of access consciousness is now said to be the fact that the information it contains is made widely available - or “broadcast” - in a global workspace to the brain’s “consumer” systems.) Another, higher-level kind of consciousness is *reflexive consciousness,* or an individual’s capacity for second-order representations of its mental states. Phenomenal consciousness, access consciousness and reflexive consciousness are all varieties of *state consciousness*, which is defined as consciousness as applied to mental states and processes, as opposed to *creature consciousness*, or consciousness as applied to a living organism (Rosenthal, 1986). The latter may be subdivided into *intransitive creature consciousness* – i.e. being awake as opposed to asleep or comatose – and *transitive creature consciousness* – i.e. the ability to perceive and respond to objects, events, properties or facts, thereby making one conscious *of* them.

Questions of relevance to animals and other organisms here include the *ontological* question regarding the nature of phenomenal consciousness, as well as *epistemological* question of how we can know which creatures possess it (the Distribution Question) and how we can know what their experiences are like (the Phenomenological Question) (see Allen, 2005, for an overview). In the following chapter, however, I argue that we first need to address the more fundamental questions of whether the foregoing categories of consciousness are well-defined for other kinds of organisms, and whether they actually carve nature at the joints, so to speak. I contend that the different varieties of consciousness distinguished by contemporary philosophers generally fail on one or both counts, and I nominate some new categories of consciousness which may be more productive for future research.

Nevertheless, the question of whether mental processes are necessarily or even paradigmatically conscious states is a philosophically controversially one. Some philosophers argue that there is nothing particularly odd about *unconscious* mental processes, as opposed to conscious ones. For instance, Lakoff and Johnson (1999, p. 10) insist that "most of our thought is unconscious, not in the Freudian sense of being repressed, but in the sense that it operates beneath the level of cognitive awareness, inaccessible to consciousness and operating too quickly to be focussed on". Other philosophers (e.g. Searle, 1999, p. 88) take a contrary view, distinguishing between *non-conscious* and *subconscious* brain states and recognising only the latter as mental, because they are at least *potentially* conscious. Searle (1992) embraces what he calls the Connection Principle, according to which a state cannot qualify as mental unless it is available to consciousness.

0.2.2.2 Intentionality as the mark of mental states

Consciousness is not the only feature that has been invoked to unify mental states. Brentano (1874) proposed *intentionality* as the property which defines the domain of the mental. “Intentionality” is a philosophical term which derives from the Latin word *intentio*, which in turn derives from the verb *intendere*, which means being directed towards some goal or thing, such as a target. Intentionality can be defined as “the power of minds to be about, to represent, or to stand for, things, properties and states of affairs” (Jacob, 2003). In a much-quoted passage, Brentano claimed that *all* mental phenomena, and *only* mental phenomena, exhibited the characteristic of direction towards an *object*, which is included within the mental phenomenon itself – “[i]n presentation, something is presented, in judgement something is affirmed or denied, in love loved, in hate hated, in desire desired, and so on”.

Brentano explicitly equated what he called “direction towards an object” with what Scholastic philosophers of the Middle Ages referred to as “the intentional inexistence of an object” in the mind. Brentano’s innovation was to *define* mental phenomena as “those phenomena which contain an object intentionally within themselves" *(Psychology from an Empirical Standpoint,* 1874 / 1911 / 1973, pp. 88-89). It is certainly true that a large number of otherwise disparate states – e.g. perceptions, beliefs, desires, intentions and many other "propositional attitudes" – instantiate the property of intentionality, insofar as they are about or represent objects and states of affairs.

The word "intentionality" should be carefully distinguished from both the ordinary meaning of the word "intention" and the philosophical meaning of the word "intension" (Byrne, 2006). *Intentions* are just one of many types of intentional states; for instance, beliefs, desires and perceptions are also intentional states. However, these mental states are not *intensional* (with an 's'); only sentences are. A sentence S is intensional if and only if substitution of some co-referring expression A in S sometimes yields a sentence with a different truth value from that of S. A sentence can be intensional and yet have nothing to do with intentionality – for instance, the sentence "Necessarily, the number of elements found in nature is 92" has a different truth value from "Necessarily, 92 is 92", yet neither refers to intentional states – while some sentences that report intentional states are arguably not intensional: "Tom saw Felix" and "Tom saw Steven's cat" have the same truth value, even if Tom does not know Felix is Steven's cat.

The success of Brentano's enterprise of defining "mental states" in terms of intentionality therefore depends on our having criteria which can be used to identify states that exhibit the property of "direction towards an object", and which pick out *all* the phenomena which we would wish to call "mental", and *only* those phenomena. As we shall see in the following chapter, there is no universally accepted definition of "intentionality" (indeed, there are several different accounts in the contemporary philosophical literature of what it means). Additionally, the question of whether intentionality can be used to distinguish mental states from other phenomena is a controversial one: some artifacts exhibit a kind of "aboutness" – for instance, a compass carries information about the location of the North magnetic pole – which has some points in common with the "aboutness" of mental states, while some mental states – e.g. so-called "raw feels" and feelings of depression – don't seem to be "about" anything. There are various schools of philosophical thought regarding Brentano's thesis that intentionality is the mark of all mental states. Some philosophers hold to a *radical irrealist* picture of the mind, rejecting the view that the word "mental" represents a single property, let alone a "natural kind". In the words of Richard Rorty *(*1979, p. 22): "the attempt to hitch pains and beliefs together seems ad hoc – they don't seem to have anything in common except our refusal to call them 'physical'". Others subscribe to a view that can be called **"anti-intentionalism"**, arguing that intentionality fails to account for the *phenomenal* character of our conscious mental states. Philosophers in this camp all hold that phenomenal consciousness is not derived from intentionality, but differ among themselves as to whether it is separable or inseparable from intentionality, and if it is inseparable, whether it is essential to intentionality. Finally, so-called **"intentionalists"** attempt to explain phenomenal consciousness in terms of intentionality: phenomenal states are intentional states.

For animals and other organisms, relevant questions regarding intentionality include the following:

- Does intentionality presuppose a capacity for language?
- Do intentional states necessarily have a prepositional content?
- Does intentionality presuppose possession of concepts?
- What is the simplest form of intentionality that can exist in creatures?
- Can we provide a naturalistic account of at least some simple version of intentionality, and

does this account have to be a biological account?

- Is it possible for an animal to have intentional states, even in the absence of phenomenal consciousness?
- Conversely, do animals have phenomenal states which lack intentionality?
- Can animal consciousness and intentionality be explained separately, or is one the key to explaining the other?

0.2.2.3 Dennett’s intentional stance

Dennett’s *intentional stance* – which he discusses in his influential book, “Kinds of Minds” (1997) and elsewhere – is obviously relevant to my enterprise of identifying and describing minimal minds in nature. The chief advantage of the intentional stance, as Dennett sees it, is its predictive convenience. There are two other methods of predicting an entity's behaviour: what Dennett calls the *physical stance* (using scientific laws to predict the outcome - e.g. the trajectory of a bullet fired from a gun), and the *design stance* (assuming that the entity has been designed to function in a certain way, and that it is working properly - e.g. that a digital camera will take a picture when I press the button). The latter stance saves time and worry if the inner workings of the entity in question are too complex for behaviour to be rapidly predicted from a physical stance. Sometimes, however, even an entity's functions may be bafflingly complicated, and we may try to predict its behaviour by asking: what does it know (or at least, believe) and what does it want? The example Dennett employs is that of a chess-playing computer. I may not understand its program functions, but if I assume that it wants to win and knows where the pieces are on the board, how to move them and what the consequences of each possible move will be (up to a certain number of moves ahead), then I can make a good guess (perhaps a wrong one, given the limits of my memory and imagination) as to what it will do next in a game.

The chess-playing computer is what Dennett calls an intentional system: an entity whose behaviour can be more easily predicted from an *intentional stance*, where the entities are treated as if they were agents who choose to behave in a certain way, because of their underlying *beliefs* about their environment, and their *desires* – or (in what Dennett regards as an alternative description), because of their *information states* that enable them to achieve their *goals.* Insofar as intentional systems are said to have beliefs and desires, they exhibit the philosophical property of *aboutness:* beliefs and desires have to be about something. If I believe that the food in front of me is delicious, I have a belief about the food, and a desire relating to it (a desire to eat it). Dennett suggests that we can usefully regard living things and their components from an intentional stance, because their behaviour is "produced by *information*-modulated, *goal*-seeking systems" (1997, p. 34).

In the first chapter, I evaluate the significance of Dennett's *intentional stance* for my quest for a minimal mind*.* There are three issues which are pertinent here. First, how realistically should we construe the beliefs described in Dennett's intentional stance? The fact that Dennett is willing to ascribe beliefs even to thermostats (1997, pp. 34-35) suggests that he regards ascriptions of belief to entities as a useful device for predicting their behaviour. However, one could proceed by defining the search for "mental states" in organisms as a search for behaviour that can *only* be explained by attributing beliefs and desires to the entities engaging in that behaviour, and then employ Dennett's intentional stance as a way of narrowing down our search for mental states.

Second, is Dennett correct in equating "belief-talk" and "desire-talk" with "their less colorful but equally intentional alternatives; semantic information-talk and goal-registration-talk" (1995a), or are there some philosophically significant differences between the "belief-desire" description of the intentional stance and the "information-goal" description? After all, talk of information and goals has a decidedly less mentalistic flavour than talk of beliefs and desires. If the two descriptions are non-equivalent, we might ask whether systems whose behaviour can be adequately described using the latter stance can really be said to have minds at all, and whether beliefs and desires are a *sine qua non* for having minds.

Third, is there a fundamental difference between the intentionality of devices like thermostats and that of human agents? Searle (1999) thinks so: he makes a threefold distinction between the *intrinsic intentionality* possessed by conscious agents such as humans and other animals; the *derived intentionality* of words, sentences, pictures, diagrams and graphs, whose meaning depends on what other people (language users) think; and the *“as-if” intentionality* which we metaphorically attribute to systems whose behaviour is analogical to that of people and other animals. Dennett, on the other hand, regards the attribution of intentionality to thermostats as more than metaphorical: he argues that if we are to explain what all thermostats have in common, we “have to rise to … a level that invokes belief-talk and desire-talk (or … semantic information-talk and goal-registration-talk” (1995a). Additionally, for Dennett the distinction between intrinsic and derived intentionality is redundant because the brain is itself an artifact of natural selection, and the "aboutness" of our brain states (read: mental states) has already been determined by their "creator, Mother Nature", who "designed" them (1997, p. 70).

0.2.3 Biological assumptions about mental states

Although I refrain from making the controversial assumption that “mental states” have something in common, there are four broad biological assumptions which I shall make in this thesis regarding “mental states” in living organisms, whatever they turn out to be.

First, I assume that mental states (whatever they are) don't just "pop up" in an entity, for no reason. Any creature that possesses mental states must have some innate *capacity* for having these states. (The same requirement would apply to any artificial device that was found to possess these states.)

Second, a living creature's capacity for mental states is grounded in its biological characteristics. I am not here equating mental states with biological properties; rather, I simply assume that differences in organisms' mental *capacities* can be explained in terms of differences in their physical characteristics. This in no way commits me to the more speculative **supervenience thesis**, which states that *all* mental properties and facts supervene on physical properties and facts, so that any difference in two individuals' mental *states* reflects an underlying physical difference between them.

Third, I assume that the mental capacities of *animals* supervene upon (or are grounded in) states of their *brains and/or nervous systems.* I am not, however, assuming that every *organism* with a mind must have a brain, or even a nervous system; indeed, I discuss alleged instances of mental states in organisms lacking nervous systems, such as bacteria. In the following chapters, I shall attempt to identify the set of biological capacities that warrant the ascription of mental states - however rudimentary they may be - to an organism.

Finally, I make the extremely modest assumption that at least *some non-human animals* possess the requisite capacities for mental states of some sort. In making this assumption, I do not wish to commit myself to the more controversial philosophical position that some non-human animals possess phenomenal consciousness. That remains to be seen. But even if our folk psychology of animals is liable to be mistaken in many ways, it would be presumptuous to deny that mental states occur in any non-human animals. Indeed, the assumption that some animals possess mental states – especially desires and other feelings – is woven into our own language to such a degree that animals often serve as primary *exemplars* of these states. To deny mentality to all non-human animals would thus render much of our everyday terminology about emotions meaningless.

0.2.4 Necessary and Sufficient Conditions for the Occurrence of Mental States

One of my *provisional* objectives in this thesis is to list the necessary and sufficient conditions for possessing “mental states”, whatever they turn out to be. Such an attempt may well fail. That in itself would be a philosophically significant result. We should not expect to find neat definitions for every concept, and the concept of "mind" may prove too elusive for such a definition.

Then again, it may not. My aim is not to define "mind" in all its possible varieties, but to define the conditions an individual would have to satisfy before it could be said to possess the most primitive kind of mind there could be - a "minimal mind", as I call it.

It might be argued that the concept of mind, like that of a game (discussed by Wittgenstein), is incapable of definition, because it is inherently *open-ended.* But even though the concept of "mind" appears to be open-ended, there is no reason why the concept of a *minimal* mind should be. A minimal mind may well turn out to be definable in terms of a small, finite set of properties. However, I refrain from assuming that there is a *unique* set of sufficient conditions for having a mind. On the contrary, there may well be several varieties of "minimal minds".

Above, I outlined my reasons – which I elaborate in the next chapter – for treating “aliveness” as a *necessary* condition for having mental states. Additionally, one of the issues I examine in this thesis is whether being alive constitutes a *sufficient* condition for having a mind of some sort.

Finally, I do not assume that subjectivity is a defining (and hence necessary) property of mind. It may turn out to be the case that for creatures with *minimal* minds, the element of phenomenal consciousness is wholly lacking from their mental states.

0.3 How should we go about identifying the occurrence of mental states in other organisms?

In this thesis, I propose to adopt an *a posteriori*, biological, "bottom-up" approach to the philosophical problem of animal minds. Instead of first attempting to define what a minimal mind is and then seeking to determine which animals fall within the scope of my definition, I shall begin by trying to define what an *animal* is. This is not merely a scientific matter: while a zoologist may be able to tell us how animals differ from closely related organisms such as plants and fungi, it is the task of philosophy to untangle questions such as what it means to be an *organism* (i.e. "alive"), or whether a robotic bee should be classified as an animal. In chapter one, I examine and reject proposals for attributing mental states to *non-living* entities, and argue that “minds” are only found in organisms, which can be defined as entities possessing the property of intrinsic finality.

One sensible way of identifying mental states in animals and other organisms might be to first examine the *biological* properties that define living things, and attempt to identify those properties that may be relevant to having a mind of some sort. One would start with a large group, such as the set of all living organisms - which I shall refer to as L for convenience - and carefully examine the definition of "organism", as well as the general properties of organisms, for anything that may be relevant to having a mind. A philosophical "winnowing process" could then be applied to these features, to ascertain whether singly or in combination, they sufficed to define the conditions for having mental states. *If* these features proved to be insufficient, one would then narrow one's focus to a *smaller* set of organisms - such as the set of all animals (call it A) - and once again critically examine the definition, as well as those universal traits of animals that might be relevant to having a mind. One could review successively smaller sets of organisms in the same way - the set of all animals with nervous systems, the set of animals with a brain, and so on - until one found a set of physical and/or behavioural characteristics that was sufficient to warrant the ascription of mental states to a creature. These characteristics can be said to define a set M of all creatures with mental states.

This is the strategy I propose to adopt in this thesis. In the process of converging from L to M, I hope to build up a set of conditions that may be relevant to the possession of a mind by an individual. As each new condition is added, the question of whether the set of conditions is necessary and/or sufficient for having a mind will be re-visited.

Henceforth, I shall generally focus on organisms which are developmentally mature and physically normal, as my primary concern is to identify the *species* whose members can be said to have minds, rather than ascertain which individuals have minds.

It might be argued that L (the set of all living things) is too large a set to begin with, in our quest for M (the set of creatures with mental states). After all, don’t we all know that microbes don’t think? In my opinion, it would be rash to make such an assumption at the outset. Di Primio, Muller and Lengeler (2000) have catalogued an impressive array of sensory and communicative capabilities, which are found even in simple bacteria:

> 1. Bacteria have internal and external *sensors* of different types.
> 2. Bacteria can synthesize sensors and effectors when required and eliminate them when no longer needed (a solution in response to changes of the environment less frequently found in higher organisms).
> 3. Bacteria have been *able to move* for about 3 billion years now by means of rotating effectors (flagella) that act like a ship's screw...
> 4. Bacteria *react to stimuli in indirect ways and the coupling between stimuli and responses is modifiable.*
> 5. Bacteria identify and compare stimuli at different times (a process based on sensory stimulation and a *simple memory*).
> 6. Bacteria are able to *integrate different (e.g. positive and negative) stimuli* when given simultaneously.
> 7. Bacteria show *purposeful* (goal-oriented) behavior...
> 8. They *communicate* by means of pheromones (signaling molecules) and by exchanging genetic information (quasi-sexual behavior).
> 9. They *co-operate* and *compete* in both an intra- and interspecific way (i.e. with bacteria of the same and of other species) (Di Primio, Muller and Lengeler, 2000, pp. 4 – 7, italics mine).

Taken together, these capacities could be regarded as *prima facie* evidence of mental states, so it seems prudent to begin our search for mental states by casting a wide net, and starting with the set of all organisms.

If M turns out to be a subset of L, then how should we construct a sufficient set of conditions for a species' being a member of M? What I propose to do, in chapters two to eight, is narrow down my search by examining several *broad categories* of behavioural and biological properties that have been proposed in the philosophical and scientific literature as relevant to having a mind, and sift through them, all the while attempting to put together a *constructive definition* of a "minimal mind". In particular, I discuss **sensory capacities (including discriminatory ability and perception); memory; flexible behaviour patterns; the ability to learn; self-directed movement and control; the ability to correct one's mistakes; and the ability to form concepts.** Within each category of "mind-relevant" properties, I examine the different ways in which these properties are realised by different kinds of organisms, at various levels of complexity. The biological case studies which I invoke range from the relatively simple (viruses) to the most complex (vertebrates, especially birds and mammals). In other words, I propose to converge from L towards M *within each category* of "mind-relevant" properties. Unlike most previous philosophical authors, I do not use these studies to *illustrate* some pre-defined philosophical distinction, but rather to *elucidate* the various ways in which the above-mentioned categories of properties are instantiated by living things. This procedure allows me to put forward a list of criteria for having mental states and formulate a constructive definition of a minimal mind.

It should be borne in mind that the step-by-step accumulation of necessary and/or sufficient conditions for having a mind may not simply converge towards a single set of animals. M may turn out to be defined by more than one set. There may turn out to be separate "islands of mentality" in the animal kingdom. Nor should it be assumed that animals which are phylogenetically closer to members of M are necessarily smarter. For instance, echinoderms (such as starfish) are more closely related to mammals than octopuses are, yet the evidence for rudimentary mental states in octopuses appears to be more convincing than the evidence for minds in starfish.

0.4 Appropriate Sources of Evidence

0.4.1 Thought experiments versus empirical observations

There are many different kinds of evidence for mental states that merit serious philosophical consideration, but there is one kind of "evidence" that should, I believe, never be appealed to. I reject arguments or thought experiments pertaining to mental states which are based on mere *logical possibility* as philosophically illegitimate. To show that a state of affairs is logically possible (or not obviously logically impossible) does not establish that it is physically possible. We can imagine organisms that look and even act like us, but have no subjective experiences, as in Chalmers' "zombie world" (1996, pp. 94 - 99); we can also imagine entities such as gas clouds, force fields or ghosts having mental states. All this proves is that mental states are not *logically* supervenient on their underlying physical states. However, as Chalmers himself points out (1996, p. 161), they may still be *physically* supervenient on these states.

0.4.2 Singular versus replicable observations

The use of animal *anecdotes* has been discredited since the days of Darwin and Romanes, who were prepared to rely on second-hand accounts of observations from naturalists and pet-owners who wrote to them. However, I would suggest that the insistence by Thorndike and Morgan on controlled, replicable laboratory experiments, while commendable for its scientific rigour, misses the point. From a scientific perspective, the key question to be asked when assessing an observation is not: “Is it replicable?” but: “Is it *reliable?”* Laboratory experiments which have been replicated will score highly on an index of reliability, as the risk of error is low. But the risk of error is also low when a singular observation is made by an acknowledged expert in the field. I conclude that there is no good scientific reason for excluding such a singular observation. What scientists should then proceed to do is further investigate this observation and endeavour to explain it within some general framework.

As regards *controlled experiments*, I have decided to err on the side of caution and not give credence to experimental observations that other researchers have tried but failed to replicate. Recent research, which has not yet been replicated, will be admitted, if published in a reputable scientific journal, but any new claims made will be treated with caution. I also reject studies whose follow-up has produced *conflicting results*.

0.4.3 Laboratory versus natural observations

There is something to be said for observing animals in their *natural state*, as cognitive ethologists do, simply because such observations maintain the network of relationships between an organism and its environment. An organism in a laboratory is an organism uprooted: the nexus of connections is severed, leaving us with less information about the interactions which characterise its lifestyle. Rigour is secured, but at a high price.

On the other hand, if the research is designed to measure the relation between a small number of variables, laboratory controls eliminate contamination by external factors.

In other words, the methodologies of behavioural science and ethology should be seen as complementary, rather than contradictory. Observations of animals in the wild will therefore be admitted if they are reliably attested by an acknowledged expert in the field.

0.5 How generous should we be in assessing claims for mental states?

Simplicity is generally regarded as an explanatory virtue, and many philosophers (beginning with the Spanish philosopher Gomez Pereira, whose claim that animals are true machines predated that of Descartes by eighty years) have invoked **Occam's razor**, which tells us "never to multiply entities beyond necessity", to dispense with the attribution of minds to animals, on the basis that it was the simplest reading of the available evidence.

Other philosophers have used Occam's razor in a *contrary* sense, arguing that the most parsimonious explanation of the pervasive neurophysiological and behavioural resemblances between human beings (who can certainly feel) and animals is that animals also have feelings (e.g. Griffin, 1976, p. 20). However, one problem with this argument is that similarity comes in degrees. How similar does an animal's brain have to be to ours before we can be sure it has mental states? Alternatively, if having a mind depends on possessing a "critical mass" of neural organisation, even animals with brains like ours may miss out, if they fall below the cut-off point.

Morgan's Canon is also used to dispense with mentalistic explanations:

> In no case may we interpret an action as the outcome of the exercise of a higher faculty, if it can be interpreted as the outcome of one which stands lower in the psychological scale (cited in Bavidge & Ground, 1994).

Even leaving aside worries about its terminology of "higher" and "lower" psychological faculties, the key insight, that *nature* must be parsimonious in the way it "designs" (i.e. selects for) organisms that can adapt to their environment (Bavidge and Ground, 1994, p. 26) contains a hidden assumption, disputed by Griffin (1994, p. 115) that it is more complicated for nature to generate adaptive behaviour by means of mental states than by other means.

The demand for "simplicity" has proven to be a double-edged sword, leaving us unsure how to wield it.

The methodology which I would like to propose here, for evaluating a claim that a certain kind of behaviour in an organism is indicative of a mental state, is to proceed by asking: "What is the most *appropriate* way of describing this behaviour?", rather than "What is the *simplest* way of describing it?" We should use mental states to explain the behaviour of an organism if and only if there is some scientific advantage in doing so: i.e. a more comprehensive and/or more empirically accurate description, model and prediction of the organism’s behaviour, than that afforded by other modes of explanation.

The methodology I am proposing here has implications with respect to the ongoing controversy as to whether mental states occupy a *continuum* from human beings to the smallest cell, as some scientists and philosophers (Godfrey-Smith, 2001; Birch, 2001; Chalmers, 1996, p. 292) have argued, or whether there is a *clear-cut divide* between organisms that have minds and those that do not, as others (Humphrey, 1993, pp. 195-196) have maintained. I have proposed that we should ascribe mental states to an organism *if and only if* doing so allows us to describe, model and predict its behaviour more comprehensively, and/or with a greater degree of empirical accuracy than alternative, non-mentalistic accounts. Either mental states do or do not further our scientific understanding of the behaviour in question. The decision to impute these states is not one that admits of degree, although the epistemic grounds for making the decision might be much stronger for some animals than for others. On methodological grounds, then, I am committed to looking for *"on-off" criteria* for ascribing these states to organisms. Failure to find them would lend support to the continuum hypothesis.

Even if there is a clear-cut divide between organisms that have minds and those that do not, we should still expect to find a vast range in the mental capacities of organisms, given their biological and behavioural diversity.

0.6 What terminological conventions will I follow when describing minds and mental states?

Philosophers and scientists guard against using inappropriately mentalistic language when describing the behaviour of organisms which may or may not possess mental states. Some verbs in the English language are peculiarly reserved for mental states. The choice of these verbs may change over time: at one time, the suggestion the attribution of sense or memory to a mindless entity would have seemed odd, but today, people commonly talk about the sensor in a thermostat, or the memory of a computer (or even a piece of deformed metal). The table below, which I believe reflects contemporary usage, sorts some everyday terms into mentalistic and non-mentalistic categories.

The following terms will be treated as *mentalistic* unless clearly indicated otherwise:

1. The phrase "act intentionally". In common usage, intentional agency presupposes the occurrence of mental states.
2. The verbs "feel", "believe", "desire", "try" and "intend", and the associated nouns "feeling", "belief", "desire", "attempt" and "intention". In ordinary parlance, these intentional terms are currently used to characterise either states of a subject ("feel", "feeling"), proposed or attempted actions by an agent ("intend", "intention", "try", "attempt"), or explanations for an agent's actions ("believe", "belief", "desire").
3. The words "perceive" and "perception", as opposed to "sensation". Modern usage draws a distinction between "sensation" and "perception" in an organism: the former is usually said to arise from immediate bodily stimulation, while the latter usually refers to the organism's *awareness* of a sensation (Merriam-Webster on-line dictionary, 2006, definition (1a) of "sensation"). *Philosophers, however, do not always adhere to this pattern of usage.* It would be prejudicial to endorse these distinctions at this stage, but we should allow for the possibility that there may be organisms that can be appropriately described as *having sensations while lacking perceptions.*
4. The verbs "remember", "recall" and "recollect". The verb "remember" retains a distinctly mentalistic connotation in ordinary usage: it refers not only to stored information being retrieved but also to the subjective feeling of its coming into one's mind. In popular parlance, machines are never said to "remember" anything. The verbs "recollect" and "recall" are even more strongly mentalistic, as they signify the intentional act of bringing something back to mind.
5. The words "learn" and "learning" will generally be treated as mentalistic, unless indicated otherwise (e.g. in the chapter on learning, where I address the possibility of *non-mentalistic*

learning in worms). This mentalistic usage is challenged by Wolfram (2002, p. 823), but I believe there is currently no verb in common use that can replace the peculiarly mentalistic flavour of "learn" in English. The word "learn" usually means "to gain knowledge or understanding of or skill in by study, instruction, or experience" (Merriam-Webster on-line dictionary, 2006). *However, we should keep an open mind.* According to the above definition, gaining a "skill" by "experience" is learning. In our examination of organisms' abilities, we may find that some living things, despite lacking minds, are capable of feats that can be described as the acquisition of skills through experience. In that case, we would have to call this "learning", simply because it would be a violation of our existing linguistic conventions not to do so.

6. The words "know" and "cognition". Some philosophers and AI researchers use the terms "know" and "cognitive" in a more general, mind-neutral sense. At present, however, popular usage treats these terms as mentalistic.

The following terms will be treated as *mind-neutral* or *non-mentalistic* in this thesis:

1. The general-purpose verbs "act" (unless followed by "intentionally") and "react". These verbs should not be regarded as mentalistic, as they are routinely used by chemists and biologists without any mentalistic connotations. In popular parlance, too, these verbs may be used in a neutral sense.
2. The verbs "seek", "search", "pursue", "attack", "avoid" and "compete". These verbs simply describe goal-oriented behaviour by entities, without any mentalistic connotations.
3. The verbs "attract" and "repel". These verbs simply describe the state of being or not being a goal.

4. The verbs "communicate", "signal" and "respond", as well as the noun "message". Scientific usage has appropriated these words, and popular usage has followed the trend. For instance, bacteria are commonly said to communicate with each other, without carrying any mentalistic overtones.
5. The verb "detect" and the nouns “detector” and "sensor". These verbs are often applied to inanimate artifacts (e.g. motion detectors, thermal sensors), although no-one speaks of these artifacts as having "sensations". The verb “sense” occupies a linguistic grey area, as it may be used with or without mentalistic connotations.
6. The noun "memory" (but not the verb "remember"). Contemporary usage allows us to speak of artifacts as having a "memory", from which they retrieve stored information.

Mental states are sometimes divided into two categories: cognitive and affective. In this chapter, when I use the term **"cognitive mental states"**, I mean *beliefs* in particular, as well as any higher-order judgements that are founded upon those beliefs. However, I do *not* wish to prejudice my enquiry at the outset by assuming that animals with beliefs and/or desires necessarily have subjective, "phenomenally conscious" mental states.

0.7 Preview of my conclusions

In the course of my enquiry, I conclude that *consciousness* is the less promising of the two hallmarks proposed for mental states, and that it fails to shed light on the question of what a minimal mind might be. My main reasons for arriving at this conclusion are: first, that many of the various senses of “consciousness” distinguished in the philosophical literature are poorly defined, being either too vague or, worse, empirically inadequate; second, that a mountain of neurological research over the past seventy years has established that consciousness, in the phenomenal sense of the word, is confined to “higher” animals; and third, that there is currently no satisfactory theory of how phenomenal consciousness (or any other variety of consciousness) can unify the domain of the mental.

This leaves us with *intentionality* as a more promising avenue of enquiry for explaining what mental states have in common. In chapter one, I examine Dennett’s *intentional stance*, and argue, contrary to Dennett, that an explanation of an entity’s behaviour in terms of its beliefs and desires is indeed psychologically richer than an explanation that invokes “information” and “goals”. I thus distinguish between two kinds of intentional stance: an *agent-centred* stance and a *goal-centred* stance, respectively. I argue that the terminology invoked by the goal-centred stance is teleonomic but mind-neutral, so we have no reason to regard an item of behaviour that can be accounted for using this stance as a manifestation of an underlying mental state. If, on the other hand, it turns out that some kinds of animal behaviour are better explained in terms of agent-centred stance (which invokes “beliefs” and “desires”) than in terms of a goal-centred stance, then I would argue that these kinds of behaviour should be regarded as manifestations of mental states.

My survey of the behavioural capacities of animals and other organisms forces me to confront the philosophical issue of whether intentionality can be *naturalised*. Briefly, I reject attempts by Dennett and other authors to account for intentionality in purely biological terms as philosophically unsatisfactory. Instead, I follow up on a proposal by David Beisecker, who puts forward a naturalistic proposal that accounts for the *normativity* of intentional ascriptions, while avoiding biological reductionism.

Some philosophers have proposed that intentionality comes in many different flavours. If so, then the kind that is most relevant to my enquiry is the “aboutness” that characterises *beliefs*. The “aboutness” that characterises sensory *perceptions* may indeed prove to be explicable in terms of a goal-centred (non-mentalistic) intentional stance, but beliefs are non-controversially mental states. If I can establish some criteria for attributing beliefs to other creatures, then I will have shown that they possess *minds*. However, demonstrating that these criteria must be satisfied by even a *minimal* mind is quite another matter: in order to accomplish this, I shall argue that the behaviour of a creature which is naturally incapable of having beliefs possesses neither of the “hallmarks” proposed earlier for mental states – consciousness and the *right kind* of intentionality. First, the implied presupposition that creatures with sensory perceptions have “raw feels” or some other primitive form of *consciousness* (and hence, a mind of sorts) is at odds with scientific findings that only higher animals, such as mammals and birds, can be said to possess even the most rudimentary form of what neurologists refer to as *primary consciousness*. Second, the behaviour of organisms that lack beliefs can be accounted for in terms of what I refer to as a *goal-centred* intentional stance, which, as I argue in chapter one, does not require us to posit mental states.

The foregoing argument implies that having *beliefs* is a necessary condition for possessing even a minimal mind. Nevertheless, the attribution of beliefs to *beings without language* is notoriously problematic. Any philosophically adequate discussion of these problems would have to include the following: Davidson’s “holism of the mental” (which requires belief-holders to possess the *concept of a belief* as one of their background concepts); our inability to specify the *content* of alleged animal beliefs; and the modest argument that in order to have beliefs, one must possess *concepts* of some sort (for which there is no firm evidence in animals, despite their impressive discriminative abilities). In the last part of my thesis, I therefore attempt to develop a robust *model* of a minimal mind, based on Ramsey's metaphor of belief as a *map*, which warrants the ascription of *truth-tracking representations* and *propositional attitudes* to creatures instantiating the model, even if they lack phenomenal consciousness. I also describe the primitive *concepts* that such a minimal mind could be said to have. In particular, I focus on "instrumental" concepts relating to means and ends, which do not necessarily imply the concept of a permanent "object".

The most novel aspect of my thesis lies in my claim that there are *at least four* fundamentally kinds of maps that an animal can construct which enable it to steer itself around its world (Ramsey). An animal may construct: *kinesthetic maps* based on its associations between its own bodily movements and their consequences; *spatial maps* based on its association between landmarks in its environment and the animal's goals; *tool maps* based on its associations between different ways of moving an external object and their consequences; or *social maps* based on its associations between its past experiences of attending to different individuals in its group, and their usefulness in helping or hindering the animal attain its goals. I argue that since there is no compelling reason why any one of these kinds of minds should be regarded as more fundamental than the rest, we are indeed entitled to speak of *four different kinds of minimal mind*. As it turns out, all four kinds of minds are distributed fairly widely in nature.

I argue that none of the four different kinds of minimal mind requires phenomenal consciousness, which appears to be confined to relatively few kinds of animals. If the account of belief which I defend in this thesis is correct, then beliefs are inherently intentional states but not necessarily conscious ones. This does not mean that intentionality is more "basic" than consciousness, but it does imply that it is independent of consciousness. This finding has clear implications for the philosophical debate as to whether unconscious mental processes are *bona fide* mental states, as some philosophers insist, or whether they can be called mental states only insofar as they are potentially conscious, as Searle (1999) and others assert. If my account is correct, then the domain of the mental is considerably larger than the domain of phenomenal consciousness. On the other hand, the domain of creatures possessing minimal minds is only a small subset of the set of creatures that have what philosophers refer to as transitive creature consciousness.

Chapter 1 – Two Approaches to Identifying Mental States in Other Creatures: Consciousness and Intentionality

In this chapter, I defend the controversial claim that only *living things* can be said to have minds or mental states. After a brief discussion of “folk psychology” and “theory theory”, I conclude that there are no convincing grounds for rejecting the common view that mental states are real phenomena. This in no way implies the more controversial view that all mental phenomena share some distinguishing feature that characterises them as mental. If, however, there is some distinguishing property which is common to *all* mental states, and *only* those states, then the two most promising philosophical candidates for this property would surely be *consciousness* and *intentionality.* I argue that the different varieties of consciousness distinguished by philosophers are in many cases poorly defined and biologically ill-informed, and that they fail to “carve nature at the joints”. I then analyse the strengths and weaknesses of three common definitions of intentionality. Lastly, I examine Dennett’s intentional stance and argue that the two ways in which it can be formulated are in fact quite distinct, rather than being equivalent to one another as Dennett supposes. I propose that one of these formulations can be used to help us identify creatures with minimal minds.

1.1 Why only living things can be said to have minds

1.1.1 Attributing minds to non-living systems leads to panpsychism

My first argument for restricting minds to living things is a *reductio ad absurdum.* I contend that once we start attributing mental states to non-living systems, we end up with an infinite regress, leading to panpsychism. I argue that panpsychism is deficient because it fails to explain why we ordinarily impute intelligence to the senders and receivers of a message, but *not* to the message itself, let alone the medium for conveying it.

The following thought experiment illustrates why many people believe that certain *non-living* systems could – in principle at least – have mental states. Imagine the most "primitive" kind of living creature to whom you would be prepared to attribute mental states. Now suppose that there exists a non-living system – say, a robot – whose behaviour is as sophisticated and complex as that of the living creature. It would surely be a form of "biological chauvinism" to deny that this system has mental states too. You might object that at the present time, no non-living system exists whose behaviour is sophisticated enough to match that of the creature you had in mind. However, it is a fairly safe bet that in the future, such a system will exist, thanks to advances in technology. Two recent inventions that highlight this point are the InsBot robot, which is capable of infiltrating a group of cockroaches, influencing them and altering their behaviour (Pest Control Times Online, 16 November 2004), and the world's first fully autonomous robotic fish, which can avoid objects and swim around a tank in a very lifelike way, using embedded sensors (BBC News, 6 October 2005).

The flaw in the above thought experiment is that it implicitly assumes that *behaviour* of the right kind is sufficient to warrant the attribution of mental states. Let us follow that line of thinking, and see where it takes us. If a *non-living* system – whether man-made or existing in nature is irrelevant here – which behaves in a sufficiently sophisticated and complex way qualifies as having mental states, then why should we not view the *ensemble* of (human and/or natural) entities *causing* the system to behave in that way as an "intelligent" system too, since this system (taken as a whole) is also capable of generating "intelligent" behaviour? Admittedly, the entities in the ensemble may be (spatially and/or temporally) far removed from the "intelligent" behaviour they give rise to – but why should that matter? The important point is that these entities can reliably generate the behaviour in question, when they are all working in tandem. We can now push the argument back one step further: the complete set of entities causing the *ensemble's* behaviour can also be viewed as an "intelligent" system, since it, too, can be regarded as a cause of "intelligent" behaviour … and so on. As we go further back, the remote causes of a system's "intelligent" behaviour will multiply and encompass more and more processes, and eventually incorporate all natural processes. In other words, the logical consequence of regarding "intelligent" behaviour as a sufficient warrant for the attribution of intelligence is a kind of *panpsychism*. (If, on the other hand, we can identify good philosophical reasons for restricting mental states to living individuals, then we can avoid the above regress, as the question of whether the *ensemble* of entities causing an organism's "intelligent" behaviour is *also* intelligent will only arise if the ensemble is itself a living thing. The only conceivable situation in which this situation would occur is where the organism in question is a parasite, living inside a host.)

The most sophisticated modern defence of panpsychism can be found in the writings of Steve Wolfram (2002), who endorses (2002, pp. 845, 1177, 1195) a scientific form of *animism*:

> [T]here are certainly many systems in nature whose behavior is complex enough that we often describe it in human terms. And indeed in early human thinking it is very common to encounter the idea of animism: that systems with complex behavior must be driven by the same essential spirit as humans.
>
> But for thousands of years this has been seen as naïve and counter to progress in science. Yet now essentially this idea – viewed in computational terms through the discoveries of this book – emerges as crucial (2002, pp. 844-845).

Wolfram’s views follow straightforwardly from his project of defining mind (or intelligence) as the behaviour of any systems which is capable of performing *computations*, coupled with his demonstration that this capacity can be found in virtually all *natural* systems, as well as artificial systems. Wolfram’s case can be summarized in the following six steps.

1. If anything can be said to be the distinguishing hallmark of intelligence, it has to be *complex behaviour.* Wolfram equates intelligence with complexity (2002, p. 844).

2. Complex behaviour can be defined as the ability to perform sophisticated calculations or *computations.* The latter term is defined broadly: "it is possible to think of any process that follows *definite rules* as being a computation - regardless of the elements it involves" (2002, p. 716, italics mine). Indeed, "all processes, whether they are produced by human effort or occur spontaneously in nature, can be viewed as computations" (2002, p. 715). In natural processes, the *rules* are defined by the *laws of nature*, instead of programs written by human beings (2002, p. 716). [One might object that the ability of mathematicians and other theorists to make *intuitive generalisations* which defy reduction to concrete calculations also qualifies as intelligent. Wolfram’s response is that intelligence has to manifest itself in a concrete, physical process (i.e. a *computation)* in order to generate results (2002, p. 721). In other words, a purely "general" intelligence would be utterly unrecognisable.]

3. According to Wolfram's **Principle of Computational Equivalence** (or P.C.I.), there is in fact an *upper limit* to complex behaviour in our universe. Thus, anything that achieves this upper limit can be considered intelligent. This upper limit is achieved by a universal system – i.e. a system which is computationally equivalent to a universal Turing machine. A *universal system* can be

used to perform any calculation - that is, one that can be programmed to follow any rule - so long as the function described by the rule only applies to a *finite* number of states (2002, pp. 642 - 644, 721). (There could conceivably be systems that can exist in any one of an infinite number of different states, but using Occam’s razor, Wolfram argues that there is no reason to suppose such systems actually occur in nature. Gray (2003) points out that there is a machine called a **universal CA,** which can do more than a universal Turing machine, because it has infinitely many parallel processors, but Wolfram could reply that there is no reason to suppose that such a system exists in reality, as our universe appears to be finite.)

4. A corollary of the P.C.I. is that any universal system is as smart as any other: "such a system can emulate any of the kinds of systems that we considered - even ones whose construction is more complicated than its own" (2002, p. 720). The only difference between universal systems is that some may require more time and resources to complete their calculations than others. However, any of them can eventually solve any problem that the others can.

5. Universal systems are surprisingly commonplace: "a vast range of systems - even ones with very simple underlying rules - should be equivalent [to universal systems] in sophistication of the calculations they perform" (2002, p. 822). Indeed, "unless it is obviously simple essentially *any behavior that one sees* should correspond to a computation of equivalent sophistication" (2002, p. 726, italics mine). The weather, the flow of sand in a sand pile, and the motion of a turbulent fluid are just a few examples (2002, p. 822).

6. Hence, intelligence can be found wherever there are systems with the ability to perform complex calculations. Since such systems are commonplace in nature, it follows that intelligence

is ubiquitous in the cosmos. Wolfram approves of the primitive, animist notion that the weather has a mind of its own: when we say this, "we are in effect attributing intelligence to the motion of a fluid" (2002, p. 822).

I would suggest the Achilles' heel in the foregoing argument is the first premise – the equation of intelligence with complexity. Accepting this equation leaves us unable to distinguish between entities that *send* or *generate* messages, systems that *transmit* messages, devices that *decode* messages, and receivers that *comprehend* them: since all of these systems are capable of highly complex behaviour, all of them would qualify as "intelligent". Nevertheless, the distinction seems to be a fundamental one. Ordinarily, we impute intelligence to the senders and receivers of a message, but *not* to the message itself, let alone the medium for conveying it. The decoder of a message is considered intelligent only if it shows signs of comprehending the message, in its response to it. However, if we accept Wolfram's computational theory of mind, and the panpsychism which it entails, then we are no longer able to draw these distinctions. Wolfram's argument implies the absurd conclusion that messages embody intelligence in the same way as their senders, as the sending of a message and the transmission of a message are both processes that can be described in computational terms. The conclusion I draw is that Wolfram's theory is an incomplete account of what minds actually do. Entities with minds, like everything else in nature, certainly undergo rule-governed transformations, but that is not all they do.

At the outset, Wolfram deliberately excludes *purpose* from his definition of intelligence, on the grounds that it is too hard to discern, especially in other species such as birds (2002, pp. 825-827). This, I would suggest, is where he goes wrong. Discerning the intentions of a message-sender is a difficult but by no means intractable problem. The first step towards solving such a problem is to identify the message-sender's *built-in ends*, which I will now discuss.

1.1.2 Non-living systems lack intrinsic finality

The foregoing argument against denying mental states to non-living systems is a negative one: doing so compels us to overlook certain distinctions which are philosophically significant. My second and more substantial argument is that non-living systems are not true individuals, since they lack ends of their own. Rather, they are mere *aggregates* of parts. Hence, they cannot be said to have minds of their own.

Now, it might be objected that the parts of some non-living systems – i.e. artifacts – are closely integrated with one another, in a very complex fashion, enabling them to work together harmoniously to achieve some end or goal. However, my point is that this goal is not one which benefits the system, but its *designers*. In other words, the finality exhibited by artifacts is *extrinsic* rather than intrinsic. Because an artifact is incapable of "selfish" ends, it fails to qualify as an individual in its own right. It therefore makes no sense to construe the selfless behaviour of such a system as the pursuits of an individual agent, with a mind of its own.

Cameron (2000, 2004) is an articulate contemporary exponent of the view that the property of having *intrinsic ends* is what distinguishes living from non-living things. Whereas the end of an artifact depends on the *use* to which it is put (2000, p. 331), organisms have their own ends:

> [O]rganisms have and are ends in a sense (as beneficiaries of actions) that the tools and instruments of organisms do not have them; the ends of the tools are derivative in their being upon the prior ends of the wholes they serve (Cameron, 2000, p. 334).

As this thesis is about mental states and not organisms, I shall content myself with summarising his case here. Cameron contends that *non*-teleological definitions of life, such as the *cluster definitions* which are currently in vogue, are both scientifically deficient and philosophically unworkable. Cameron's own position is a neo-Aristotelian one. First, he maintains that teleological discourse is perfectly compatible with modern-day science. Additionally, Cameron distinguishes between the *derived* ends of complex artifacts and the *intrinsic* ends of biological organisms. Finally, Cameron proposes a simple definition of life (2000, p. 335): to be alive is to possess *intrinsic ends.*

Contrary to a view that is widely held by historians of science, Cameron does not consider Aristotle to have been a *vitalist* (someone who believes that the "aliveness" of an organism is completely *independent* of the interactions between its parts), but rather, a *weak mechanist* of sorts*,* since Aristotle's philosophical works indicate that he accepted the thesis of *upward determination:* the higher level properties and parts of organisms depend for their existence on micro-level properties and parts (Cameron, 2000, pp. 36-41). However, Cameron also demonstrates that Aristotle viewed living things as having higher-level *teleological* properties whose causal powers are not exhausted by the causal powers of their underlying parts. These properties of an organism exert their own causal influence over the course of their micro-parts' careers *without breaking any physical laws*. The causal powers of these higher-level properties belong to the domain of *final* causality, which for Aristotle is *not* reducible to the other categories of causality. Because Aristotle believed in *downward causation*, Cameron classifies him as an *emergentist*.

Cameron also rebuts on textual grounds (2000, pp. 86-135) what he regards as a false, animistic interpretation of Aristotle’s views by some commentators, who construe Aristotle as holding that not only biological processes, but also necessary and regular occurrences in the inanimate world required final causes, or that all things possessing a nature had a final cause. Cameron argues convincingly that these interpretations of Aristotle are based on a mis-reading of what he actually taught; Aristotle’s teleology was in no way animistic.

Teleology continues to play a vital role in modern science. Statements cited by Cameron (2000, pp. 171, 219-220) from contemporary biologists attest to scientists’ recognition of *teleology* as a pervasive feature of the natural world, wherever *life* is found. As Karen Neander, a philosopher of biology, points out, "the apparent explanatory power of teleological explanations which appeal to biological functions is quite robust" (1991b, p. 127; cited in Cameron, p. 171). Neander calls teleology the "conceptual glue" of biology and declares that it would be "hard to exaggerate" the concept's importance to biology (1991b, p. 137; 1995, p. 127; cited in Cameron, 2000, p. 171). However, because final causation (teleology) is incompatible with a "scientific world-view that countenances only efficient causation" (Buller, 1999, p. 6), modern scientists have attempted to construct accounts that *reduce* teleological properties to something more scientifically acceptable – usually either *structural* properties (as in the systems approach to teleology) or *historical* properties (as in the etiological account of functions). Cameron (2000, pp. 168-213; 2004) argues at length that such attempts fail for three main reasons. First, they undermine the authority of biologists whose specialty is not evolution (e.g. physiologists) to attribute biological functions to organs (such as the heart), without reference to their evolutionary past. (This failing is peculiar to the etiological account of biological functions.) Second, they are vulnerable to counter-examples: the conditions they specify are neither necessary nor sufficient for possessing a biological function. Third, they fail to account for the goal-directedness and biological normativity of teleological properties.

As Cameron is well aware that teleological explanations are currently scientifically unfashionable, he carefully rebuts some common reasons for excluding *final causality* from scientific explanations, refuting oft-repeated assertions that teleology implies a pre-existing commitment to animism, or panpsychism, or the existence of God, or the benevolence of nature, or a belief in "progressive" evolution, or vitalism (2000, pp. 215-219). **Emergence** may seem a strange, almost magical notion to some, but Cameron contends that if we accept Hume's argument that "we are in a natural state of ignorance with regard to the powers and influence of all objects" when we consider them *a priori*, it follows that "there are no *a priori* bars on what types of events, properties, or entities might emerge from the causal interactions of complex groupings of micro-entities" (2000, p. 228). Nor can there be any bars on *downward causation:* for all we know, higher-level states such as final causes may be able to affect and even direct micro-level states.

Finally, Cameron (2000, pp. 331-335) argues that there is a distinction between the *intrinsic ends* of living things and the ends of artifacts, which are *derivative* upon the purposes of their designers or users. Artifacts are in no sense "alive". Cameron (2000, p. 335) puts forward a simple definition of life: to be alive is to possess *intrinsic ends.* Cameron also argues (2000, pp. 327-335) that Aristotle was in fact well aware of the distinction between *intrinsic* and *derived* ends, insofar as he taught that the body *parts* of an organism possess ends, but only in a secondary sense, which derives its intelligibility from the nature and finality of the organism as a whole, which possesses intrinsic ends.

Cameron carefully distinguishes between the various senses in which a thing may possess intrinsic ends. Plants, animals and people are all alive, but not in the same way: the life of plants, animals and humans can be identified with the activity of the reproductive soul, the sensitive soul and the rational soul, respectively (Cameron, 2000, pp. 327, 336).

I would like to elaborate Cameron’s point on intrinsic finality below, as his thesis deals principally with living organisms, rather than artifacts. Some recent philosophers have argued that the distinction between intrinsic and derived finality is blurred, while others have queried the notion that artifacts lack intrinsic finality.

1.1.2.1 Intrinsic versus extrinsic finality: Is the distinction a clear-cut one?

Leahy (1994) has argued that the distinction between intrinsic and extrinsic finality is blurred by human domestication of animals:

> Furthermore it is not even clear that the distinction in *telos* is that marked. What is in the interests of animals is increasingly decided by human beings. A good guide-dog, sheep-dog, circus, zoo or farm animal, thoroughbred or pet Siamese is treated on the basis of criteria proper to the different roles imposed upon them by human beings... (1994, p. 46).

The recent introduction of *genetically modified organisms*, which are expressly created for human ends, lends an even greater plausibility to Leahy's case.

However, the foregoing argument merely establishes that some organisms possess *both* intrinsic and extrinsic ends. The existence of the latter in domesticated and genetically modified organisms constitutes no reason to deny the reality of the former, and even in those cases when intrinsic and extrinsic ends happen to coincide, they remain conceptually distinct. The point made by Cameron (2000, pp. 331-335) is that it is only in virtue of its *intrinsic* ends that an organism can be said to be alive.

1.1.2.2 Do artifacts possess intrinsic finality?

Varner (1998) is another critic of the enterprise of defining life in terms of having intrinsic ends. Although he recognises the distinction between artifacts and living things as one which matters philosophically and ethically, Varner is leery of resorting to an extrinsic vs. intrinsic dichotomy to ground this distinction. Citing Nagel (1979), he argues (1998, p. 66) that some artifacts can be regarded as "goal-directed systems" with ends that can be specified *independently of the goals of their human producers.* To illustrate his case, he offers the example of a Patriot missile, uncovered by an alien scientist long after a nuclear holocaust has wiped out all intelligent life on earth. The alien may be able to deduce that Patriot missiles are meant to intercept projectiles, without knowing a thing about late 20th century aerial warfare.

(We can strengthen Varner's case by imagining a 21st century new generation missile, with unforeseen military capabilities, designed and refined not by human beings but by a factory of robots, following the instructions of a computer, after all human life on earth has been wiped out in a nuclear conflict. Let us say that the computer was originally programmed by a long-dead mathematician, not to build a particular missile, but to follow a search algorithm for identifying and comparing possible designs for missiles. In this case, the design is actually "found" by the computer, and its capabilities were not foreseen by the program designer, so its built-in ends cannot be adequately specified in terms of the intentions of the human designer of the original search algorithm.)

Now, while it may be true that certain goal-directed features of artifacts can be appreciated without any knowledge of what they were designed for, the point remains that these artifacts are not *self*-directed. In Varner's example, the parts are not designed to maintain the missile, but to enable it to shoot down projectiles. The missile does not "benefit" in any sense from doing so - its mission is a "suicidal" one. Its finality is thus *extrinsic*, rather than intrinsic. (Likewise, a computer designed by human beings has a *telos*, but it is extrinsic: it is designed to perform computations.)

Lest the foregoing argument be considered too sweeping and hasty, I would like to flesh it out by proposing four physical properties that one might reasonably expect to find in a physical system possessing intrinsic finality, none of which are found in a Patriot missile or any other contemporary artifact:

(i) an internal *master program* that regulates the formation of the system, as well as the structure of its internal parts. All living things possess this kind of internal program (Koshland, 2002); artifacts (such as a Patriot missile) do not. At most, what they have is a built-in program that co-ordinates their internal parts, *after* they have been already assembled by some *externally* directed process;

(ii) *novel chemical properties* which appear when the different parts of the system are put together. When we examine the biomolecules within a living cell, we find that these novel properties abound. Birch (2001) refers to the relations between the parts of a system whose properties cannot be described in isolation from one another as *internal relations*.

Unlike the biomolecules in a cell, the parts of a Patriot missile do not acquire any new chemical properties by virtue of being put together; rather, the missile is an assemblage of pre-existing parts whose properties can be described in isolation from one another, like the bricks in an office block, and the workings of the ensemble can be subsequently deduced from an understanding of these external properties alone (Birch, 2001);

(iii) a *nested hierarchy of organisation*, in which the "components are ... organised into a hierarchy of compound individuals" (Birch, 2001, online). In a living thing, the parts are organised in a *nested hierarchy*, from macromolecules to organelles to cells to tissues to organs to organisms. A nested hierarchy is a necessary (but not sufficient) condition for the *telos* of the parts to be completely subsumed within that of the whole. Because human-built computers (and Patriot missiles) have no such hierarchy of internal organisation, they are, as Birch remarks, mere *aggregates*, not individuals;

(iv) *dedicated functionality*, where the range of *activities* of the parts can only be understood *by reference to the whole.* For each layer in the nested hierarchy of a living thing, the entire repertoire of functions of the parts supports the next highest level of organisation. No such dedicated functionality exists in a Patriot missile, or in today's computers. Living things are built from the bottom up, by "dedicated", *intrinsically* adapted parts; today's human-built computers are designed from the top down, out of parts which have to be modified in some way, to suit the designers' ends (Geer, 2002).

I wish to emphasise here that I am *not* proposing the above four properties as a set of sufficient conditions for possessing intrinsic finality, but simply to illustrate my point that there are several striking physical differences between living organisms and contemporary artifacts, which strongly suggest that artifacts do *not* possess intrinsic finality. It is, however, quite possible that *future* artifacts may instantiate these properties. Should they ever do so, then the question of whether artifacts can indeed possess *intrinsic* (in addition to extrinsic) finality will need to be re-examined, and we might then conclude that some artifacts are indeed alive.

Finally, I would like to point out that restricting the attribution of mental states to those entities that can actually *benefit* from having them – in other words, entities that possess *intrinsic finality* – has an additional advantage: it prevents the commencement of the infinite regress I referred to in section 1.1.1, thereby enabling us to avoid the panpsychism implied by the regress.

1.2 A defence of mental states against eliminative materialism

Many philosophers and scientists claim that our everyday understanding of mental states constitutes a primitive "theory of mind", which is referred to in the philosophical literature as "folk psychology". If this claim is true, then we have to ask whether folk psychology is: (i) a *universal* psychological theory, which is substantially correct, as well as adequate for explaining most or all of the phenomena investigated by modern psychologists; (ii) a *partial* theory, which is able to account for a limited range of psychological phenomena; or (iii) a *radically mistaken* theory, as eliminative materialists maintain. Universal theories which cover all events within their scope are relatively rare in science – the modern atomic theory of chemistry is perhaps the best-known example – and advocates of folk psychology have never claimed that their theory is anything like a complete one. Nor do they claim that their theory is free from errors. The question that concerns us is whether the errors are radical enough to warrant jettisoning folk psychology. If so, then my search for a "minimal mind" in this thesis would be a pointless quest.

The term "folk psychology" is used in contradistinction to the view that the content of our mental lives is immediately available to us, through the incorrigible process of introspection. Sellars (1956) attacked this introspectionist "myth of the given", as he called it, by presenting an alternative myth of how our ancestors learned to attribute internal thoughts to others, as a way of accounting for their behaviour; later on, they learned to ascribe mental states to themselves.

Stich and Ravenscroft (1994) distinguish two senses of "folk psychology": an *externalist* sense, which refers to the theory of mind implicit in our everyday *discourse* about mental states, which is said to implicitly define terms such as "believe", "want" and "desire"; and an *internalist* sense, which refers to an internal representation which each of us possesses, that allows us to *predict* and *explain* the *behaviour* of ourselves and others. Folk psychology may be true (or false) in one sense without necessarily being true (or false) in the other. Likewise, we can distinguish two senses (externalist and internalist) of the higher-level claim – known as "theory theory" – that our everyday understanding of people's mental states is equivalent to a kind of theory – namely, a folk theory of mind. Theory theory (external) refers to the claim that our everyday *talk* of mental states implicitly constitutes a folk theory of mind, while theory theory (internal) refers to the claim that our capacity to *predict* and *explain behaviour* is grounded in an internally represented theory of mind – namely, folk psychology in the internal sense – which may be either learned or innate.

1.2.1 Relevance for other animals

Theory theory can be applied to other animals too. First, human beings' everyday discourse about *other animals'* mental states may be said to presuppose a folk *theory* of animal minds, which implicitly defines terms such as "believe", "want" and "desire", insofar as they apply to animals. Second, every human being may also be said to possess an internal representation – either innate or learned – that allows him/her to *predict* and *explain* the *behaviour* of other *animals.* This internal representation can also be regarded as a theory.

Neither the externalist nor the internalist versions of “theory theory” have gone unchallenged in the philosophical literature. Theory theory (external) has been challenged along two grounds. First, it has been argued that our linguistic ascriptions of mental states to agents are not theoretical, but can be cashed out as *dispositions* to behave in a certain way. The main difficulty with the dispositionalist account (Chisholm, 1957) is that it is impossible to define belief straightforwardly along dispositional lines, as someone with a belief (e.g. that it will rain) will only be disposed to behave in a certain way (e.g. take an umbrella) if she has certain desires (e.g. not to get wet) and other attendant beliefs (e.g. that umbrellas ward off rain). (For a further discussion, see Schwitzgebel, 2006). Likewise, the behaviour of someone with a particular desire will depend on her surrounding desires and beliefs. Granted that with other species of animals, the attribution of surrounding desires and beliefs is far less problematic than with human beings, it remains the case that any attempt to define even the simplest kinds of beliefs and desires in animals will require us to invoke a host of accompanying beliefs and desires, which also need to be defined. Thus the enterprise of accounting for even other animals’ mental states in atheoretical terms is fraught with problems.

The second principal challenge to theory theory (external) is the argument that our everyday talk about mental states is simply too *vague* and under-defined to constitute a proper theory that is capable of defining terms such as “believe” and “desire”. Ravenscroft (2004) discusses the problems associated with one widely used approach, pioneered by Lewis (1972), of defining mental states in terms of popular *platitudes* that everyone accepts. A similar argument can be mounted for other animals, that our everyday discourse about their mental states is far too vague and inconsistent to constitute a proper theory: as I pointed out in the Introduction, there are considerable differences, both within and across human cultures, as to what level of mental sophistication should be ascribed to animals and other organisms, with some cultures even regarding all natural objects as possessing mental states (animism). Now, it might be said that in the case of human beings, there is no need to define mentalistic terms, as *introspection* can give us a non-theoretical understanding of what it means for a human being to have a belief, desire or some other mental state. Be that as it may, the point I wish to make here is that when we are talking about other species of animals, this avenue is not available to us. We cannot “look inside their minds”. If we wish to place our discourse about other animals’ mental states on a firm philosophical footing, we have no choice but to define their mental states in *theoretical* terms. The philosophical task that will be attempted in this thesis is therefore to construct a robust theory of animal minds – in particular, the simplest kind of mind that an animal could have.

Theory theory (internal) can also be challenged when applied to other animals. Simulation theorists maintain that our capacity to predict and explain people’s behaviour is not grounded in an ability to create and *internally represent* theories about other human agents, but in a capacity to *simulate* (i.e. pretend to have) their mental processes: deciding what one would do in such circumstances then enables one to predict their decisions. Likewise, one could plausibly argue that for other animals, the best way to predict their behaviour is to put ourselves in their shoes, and ask ourselves what we would do in those circumstances. Simulation theory is certainly a promising approach to the quest for a minimal mind – one could even attempt to define the set of creatures possessing minds as those whose behaviour can only be accurately predicted by performing a simulation. On the other hand, some authors (e.g. Ravenscroft, 2003) have argued that the ability to perform simulations itself presupposes possession of an internally *hardwired* theory of human behaviour; if they are right, then simulation theory collapses into theory theory (internal).

Whatever the merits of simulation theory, we need to bear in mind that the internal representations – be they simulations or otherwise – that enable us to predict animal behaviour are likely to be quite erroneous. Human predictions about animal behaviour are notoriously fallible, even under "ideal" observational conditions. Our internal representations therefore need to be supplemented by a theory that defines animal mental states as rigorously as possible, so that scientists who study animal cognition and emotions can do their work properly. In the Introduction, I suggested the attribution of mental states to animals was justified only if it helps scientists achieve their goal of *systematically* describing, explaining and predicting animal behaviour. If this proposal is correct, then the ideal way of defining mental states in animals is simply the one that allows researchers to realise this scientific goal.

I conclude that if we are to justify the ascription of mental states to other species of animals and address the task of systematically explaining their behaviour, then we require the external and internal versions of theory theory for these philosophical endeavours, respectively.

1.2.2 Eliminative materialism: arguments in favour

Eliminative materialism (EM) amounts to the assertion that: (i) theory theory (external) is true, but (ii) the externalist account of folk psychology is radically false: terms such as "belief" and "desire" do not refer to anything real. As we saw above, our ascription of mental states to animals cannot be placed on a firm philosophical footing unless claim (i) is true. The discussion below focuses on claim (ii).

Proponents of EM have advanced several arguments advanced against the externalist account of folk psychology (for a discussion, see Ramsey, 2003). A common flaw in these arguments is that they fail to establish that folk psychology is a *radically false* theory, which is what they need to do. For instance, it has been urged (P. M. Churchland, 1981; P.S. Churchland, 1986) that the externalist account of folk psychology does a very poor job of explaining many features of our mental lives (such as dreams, altered states, memory, learning, mental illness and some peculiar features of consciousness). However, this merely shows that folk psychology is an *incomplete* theory, not that the theory itself needs to be rejected *in toto* (Ramsey, 2003). If mental states arise from lower-level events in the brain, to which we have no introspective access, then it is hardly surprising that folk psychology fails to explain events that transpire on these lower levels. The ongoing controversy over folk psychology highlights a more fundamental philosophical problem, relating to the appropriate grounds for rejecting *any* kind of theory: at the present time, there is no consensus as to how much of a mismatch between theory and observation warrants our abandoning a theory altogether (Stich, 1996). There is also considerable controversy over whether folk psychology is a fruitful theory, as it defenders maintain (Greenwood, 1991; Horgan and Woodward, 1985) or a stagnant one, as its critics allege (P. M. Churchland, 1981; P. S. Churchland, 1986).

The most substantial argument mounted against the externalist account of folk psychology by its critics is that it is incompatible with modern scientific findings, either because the way we talk about mental states is considered to be at odds with a mature understanding of scientific psychology (Stich, 1983), or because it is held to be at odds with what we now know about neuroscience (P. M. Churchland, 1981). Beliefs and other propositional attitudes, which figure prominently in folk psychology, have two properties that are said to clash with the modern scientific attempt to explain events within the framework of a purely *physicalistic* ontology. First, their *syntactical* properties, as shown by their sentence-like structure, do not appear to map onto anything similar in the brain. Second, their *semantic* properties of “aboutness” (intentionality) do not correspond to anything in our current scientific models of how information is stored in the brain. A defender of folk psychology might respond that its incompatibility with neuroscience is only apparent: neurology and folk psychology look at *different levels* of the brain. However, the connectionist model of memory and inference recently developed by Ramsey, Stich and Garon (1990) allegedly demonstrates the impossibility of reconciling folk psychology with modern scientific psychology at *any* level of the brain.

> Since some connectionist models store information in a highly distributed manner, there are no causally discrete, semantically evaluable data structures that represent specific propositions. It is not just that these models lack the sort of sentential, compositional representations assumed in more traditional (or language of thought) models. Rather, it is that in these networks there are no causally distinct structures that stand for anything specific. Consequently, there do not appear to be any structures in these networks – even at a syntactical level of analysis – that might serve as candidates for identifying beliefs and other propositional attitudes (Ramsey, 2003, online).

Three points can be made in response to the above line of reasoning. First, all the above model shows is that folk psychology is incompatible with particular versions of physicalism – namely, connectionist models of memory and inference. However, the question of whether distributed networks accurately model the workings of the brain remains an empirically open one. Other versions of physicalism might be able to accommodate propositional attitudes.

Second, the assertion that in highly distributed networks, we cannot specify the semantic content of network elements that give rise to cognitive episodes has been questioned by some authors (Forster and Saidel, 1994), who have proposed that it is actually possible to identify causally relevant pieces of stored information, using sophisticated forms of analysis.

Finally, supporters of connectionist models also need to establish that their models provide a *comprehensive* account of the brain's decision-making processes, and that we have no need to invoke *additional* occurrences in the brain, in order to rule out the possibility of propositional attitudes. In the absence of such a demonstration, a defender of folk psychology could suggest that some kind of *top-down causation* takes place in the brain, thereby allowing a causal role for beliefs and other propositional attitudes. (Cameron (2000, pp. 228 ff.) has vigorously argued that this kind of causation is compatible with modern science, even if it is at odds with contemporary metaphysical prejudices.)

Some philosophers (e.g. Reppert, 1992) have argued that EM is self-refuting: while the bare proposition that there are no beliefs is not logically contradictory, the *assertion* that there are no beliefs is self-refuting, since one cannot make an assertion unless one *believes* it to be true. In reply, proponents of EM reject as question-begging the assumption that an individual's making an assertion – or exercising any other psychological capacity, for that matter – requires her to hold some belief (Ramsey, 1991, 2003).

A better approach might be to show that proponents of EM have certain *commitments* which are at odds with their theory. The commitments I have in mind here are ones which are *critical* to our everyday lives: without them we could not function as rational animals. In particular, the activities of arguing and critical thinking are activities which all of us are committed to engaging in, simply by virtue of trying to live a human life. The fact that proponents of EM *argue* for the truth of EM and against folk psychology, and *critically evaluate* arguments against EM, demonstrates their commitment to these human activities in a striking fashion.

The difficulty for EM is that since it denies the reality of beliefs, it appears unable to explain (a) what *purpose* these activities serve, or (b) why engaging in these activities is *worthwhile* for its own sake. Advocates of "folk psychology", on the other hand, have a ready answer to both these questions. First, the purpose of arguing with someone is to persuade that person to change their *beliefs.* Likewise, the purpose of critical thinking is to re-evaluate one's *beliefs*, and if necessary revise them. Second, these rational activities are worthwhile, because the public acknowledgement of beliefs creates new opportunities for human decision-making, enlarging the range of choices available to us.

This "rebuttal" of EM suffers from one major limitation, however: even if it is valid, it fails to establish that other kinds of animals (who cannot argue) have mental states, too.

1.2.3 Implications for animals

As there appear to be no compelling arguments in favour of EM, the quest for a *minimal mind* remains a defensible philosophical undertaking. Nevertheless, the absence of good arguments for EM does not establish that non-human animals have beliefs or desires; all we can say is that they *may* do so.

Arguments have been mounted against EM, based on the remarkable predictive successes of folk psychology *(internal)* (Ramsey, 2003). These arguments typically rely on an inference to the best explanation to support their conclusion that folk psychology *(external)* is roughly true and that our ascription of beliefs to other agents is warranted. However, these arguments cannot be treated as conclusive, since they tend to discount anomalous findings, and overlook the fact that folk psychology is unable to account for a broad range of mental phenomena (P. M. Churchland, 1981; P. S. Churchland, 1986).

Finally, even if these inferential arguments were powerful enough to discredit EM as a *global* scientific hypothesis that could be applied to *all* animals, it would remain an *empirical* question whether EM can predict and explain the behaviour of some or all *non-human* animals more successfully than folk psychology, which attributes beliefs, desires and other mental states to these creatures.

I shall therefore attempt a comprehensive *philosophical* overview, in the following chapters, of the general biological and behavioural capacities of animals that might be said to constitute evidence for their having minds of some sort, in an endeavour to generate a *constructive* definition of a minimal mind.

1.3 Do we need to unify the domain of the mental?

From a contemporary perspective, it may seem natural to view pains, sensations, feelings, desires, wishes, imaginings, memories, beliefs, ratiocinations and decisions as different kinds of mental states. My aim here in this short section is to illustrate the oddity of this way of thinking from an Aristotelian perspective, and to suggest an alternative way of looking at the phenomena that we are disposed to lump together as “mental”. For all we know, the idea that these phenomena share something in common may turn out to be nothing more than a modern prejudice.

There was no term in Aristotle’s lexicon for what we would call "mental states". The term *psuche* (soul, or life principle) will not do, as plants, which are said to lack perception, have a *psuche* because they are capable of being nourished (*De Anima* 2.4, 415a24-25, 415b27-28). Animals are characterised by virtue of their faculty of **perception** (*aisthesis*) (*De Sensu* 1, 436b10-12), without which it would be impossible for them to move around, but non-human animals are said to lack **reason** (*logos*) (*De Anima* 3.3, 428a4; *Eudemian Ethics* 2.8, 1224a27; *Politics* 7.13, 1332b5; *Nicomachaean Ethics* 1.7, 1098a3-4), **reasoning** (*logismos*) (*De Anima* 3.10, 433a12), **thought** (*dianoia*) (*Parts of Animals* 1.1, 641b7), **belief** (*doxa*) (*De Anima* 3.3, 428a19-24; *On Memory* 450a16) and **intellect** (*nous* - also translated as "mind") (*De Anima* 1.2, 404b4-6; all references cited in Sorabji, 1993, p. 14). Aristotle described *nous* (translated as "mind", but also rendered as "intellect" or "reason") as "the part of the soul by which it knows and understands" (*De Anima* 3.4, 429a9-10; cf. 3.3, 428a5; 3.9, 432b26; 3.12, 434b3). "[J]ust as the having of sensory faculties is essential to being an animal, so the having of a mind is essential to being a human" (Shields, 2003; see also *Metaphysics* 1.1, 980a21; *De Anima* 2.3, 414b18; 3.3, 429a6-8). Aristotle does not seem to have regarded perception and thought as even belonging to a common category (e.g. "knowledge", "cognition", "awareness" or "consciousness"). On the contrary, he sharply distinguished knowledge or cognition (*gnosis*) from perception (*De Anima* 3.8, 431b24), arguing that “thinking admits of being false and is enjoyed by no animal that does not also have rationality" (*De Anima* 3.3, 427b). The only term that Aristotle does apply to both perception and thought is *krinein* (*De Anima* 3.9, 432a16), which according to Ebert (1983) is best translated as **discrimination**, or a discerning activity. The related adjective *kritikos* means capable of judging.

Aristotle argues that since all animals have at least one form of perception (the sense of touch), they are subject to pleasure and pain, and can be said to experience a **desire** for the pleasant *(De Anima*, 2.3, 414b1 ff.); lacking reason, however, they are incapable of deliberation and thus incapable of free **choice** *(prohairesis) (Nicomachean Ethics*, 1111b7-9, 1113a11). Aristotle uses the word *epithumia* to denote the irrational craving of animals, as distinct from *orexis* (desire), which can be both rational and irrational. There is however, no suggestion in Aristotle's work that he regarded processes diverse as perceptions, thoughts, mental images and desires as members of a special domain of "mental states". All that these processes can be said to have in common is that they are acts of the animal soul – but then, so are nutrition and locomotion.

Aristotle's concept of the soul was later adopted by Thomas Aquinas. In his online essay "From Augustine's Mind to Aquinas' Soul", O'Callaghan (2000) discusses the Augustinian concept of "mind" as a single indivisible capacity or power underlying the faculties of intellect and will. Augustine, whose philosophy exerted a strong influence on the early Aquinas (as well as Descartes), identified the *mind* (which he equated to will, memory and intellect) with the essence of the soul, effectively sidelining its role as the body's biological *principle of life*. However, O'Callaghan argues convincingly that Aquinas, in his later years, wholeheartedly rejected this concept, returning to an Aristotelian view of the soul as the form or organising principle of the body. In Aquinas' mature work, the S*umma Theologiae*, the term "mind" *(mens)* is rarely used, and when it is employed, it is simply identified with *intellect.* Although we can speak of a unity of intellect and will, this unity "is not preserved in a special power that separates man *from* animals, … [but] in the unity of the soul that unites man *to* animals, insofar as it specifies the form that animal life takes in being human" (O'Callaghan, 2000). Thus we cannot separate the life of the mind from the *biological life* of the human animal; rather, for human beings, the life of the mind literally *is* the life of the animal.

In contemporary English usage, the term "mental" has a far broader application than it had in Aquinas' day: even sensory capacities are now commonly regarded as "mental". However, Aquinas' point can easily be generalised to cover the entire gamut of capacities that we would now describe as "mental": the only common characteristic shared by these capacities may simply be that they are *capacities of animals* – but then, the same could be said for capacities such as the ability to eat, drink, walk or swim.

1.4 Consciousness as the hallmark of the mental: philosophical background

The notion that consciousness could serve as the defining feature of mental states would have made no sense to Aristotle; indeed, Hennig (2004) asserts that "Aristotle did not use any such concept" (p. 18). The only place where Aristotle seems to have said anything about what we would call "consciousness" is in his brief and somewhat cryptic discussion (*De Anima* 3.2) of how it is that we can perceive that we are seeing or hearing.

The following account of the history of the term "consciousness" is a brief summary of the work of Boris Hennig (2004) who has carefully chronicled the manner in which the usage of term has evolved from Roman times down to the time of Descartes. The English word "consciousness" derives from the Latin *conscientia*, which is etymologically composed of *con-* and *scientia*. Literally, it means knowledge-with, or shared knowledge, and this seems to have been the original meaning of the term: in early Latin texts, a person is called *conscius* if she shares knowledge with someone else (Hennig, 2004, p. 21). In some pagan and early Christian texts, *conscientia* simply means common or public knowledge, but in other Latin texts, it has a more restricted meaning of knowledge shared with others on the basis of direct testimony. Seneca at one point (fragment 14, preserved by Lactantius) personifies *conscientia* and speaks of it as the witness of our thoughts and acts. Indeed, the Latin word *conscius*, taken as a noun, literally means “witness”.

Augustine altered the meaning of *conscientia* by making it something private rather than public, and by limiting it to moral matters. Although *conscientia*, in Augustine’s writings, pertains to the interior realm, it does not refer to a person’s subjective knowledge of herself, but to the knowledge that an ideal observer (God) has about each of us. It is for this reason that Augustine claims that no human being can ever know whether another’s *conscientia* is pure or not (Hennig, 2004, p. 22). There are, however, other passages in Augustine where he uses the term in a sense which is closer to its etymological one.

In medieval Christian theology, the word *conscientia* came to stand for the moral conscience, or one’s knowledge about one’s own wrongdoings, as opposed to God’s knowledge about us. The term could also refer to our knowledge of how we should act.

Aquinas defines *conscientia* as the *act* whereby we apply knowledge – i.e. anything which can be expressed in the premises of a practical syllogism – to actions *(In Sent.,* 24, 2, 4, c.a.). There are two ways in which this knowledge may be applied to an action – first, in order to determine if the action has been successfully carried out, and second, in order to decide if the action was good or bad. Both cases are called *conscientia* (*Summa Theologiae*, Ia, q. 79, a. 13). Non-human animals, or “brutes”, are clearly precluded from any share in *conscientia* by virtue of their incapacity for reasoning or free choice *(De Veritate* 24.2). Although brutes are capable of forming natural judgements about certain determinate objects, they are ignorant of the basis of their judgements. Brutes do not deliberate about their actions; each of their judgements is determined to a single course of action *(De Veritate* 24.2, reply).

Later medieval philosophers treated *conscientia* as a habitual application of knowledge to actions (Hennig, 2004, pp. 33-36). By defining *conscientia* as a habit, these authors hoped to better explain how one’s moral conscience could be corrected and trained.

By the late medieval period, *conscientia* was thus regarded as an actual or habitual application of practical knowledge to a particular action. However, both Aquinas (*De Veritate*, 17, 1, ad 4) and Bonaventure (*In II Sent*., 39, 1, 1, c.a.) explicitly denied that *conscientia* causes the action it is about. The resolution of this paradox was most clearly articulated by Bresser, a Jesuit and a contemporary of Descartes, who distinguished between an action as a physical event and as a moral event – where the term “moral” simply signifies “subject to evaluation”. *Conscientia* is the cause of an action *qua* moral event, rather than *qua* physical event. In other words, *conscientia* means *whatever it is about an action that makes it subject to evaluation*, where the relevant norms are understood as public criteria or conditions of satisfaction, which can be shared by the agent (Hennig, 2004, pp. 36-39).

According to a widely held view, Descartes was the first writer to use the term *conscientia* in a way that did not fit the traditional meaning found in Aquinas and other authors; consequently, French, English and German translators of his writings – in particular, Pierre Coste, Locke and Christian Wolff – felt compelled to coin new words to denote the state of being aware – “conscience psychologique”, “consciousness” and “Bewusstsein”. Descartes is supposed to have posited a fundamental distinction between processes which can be performed absent-mindedly or while asleep, and activities or states that require our attention. Processes of the former kind were excluded from the sphere of mental states; these processes were deemed to be merely "automatic", while processes requiring our attention were characterised as "cogitative", or relating to thought. However, Descartes' conception of "thought" was meant to encompass *all* mental states. At first blush, the textual evidence, as illustrated by the following quote from Descartes’ *Principles of Philosophy* (1644), seems to support such an interpretation:

> By thought I understand all that of which we are conscious as operating in us. And that is why not alone understanding, willing, imagining, but also feeling, are here the same thing as thought (Haldane and Ross, 1970, I.222).

Elsewhere, Descartes wrote:

> [T]here are ... characteristics which we call mental [*literally cogitative, or relating to thought*] such as understanding, willing, imagining, sensing, and so on. All these are united by having in common the essential principle of thought, or perception, or consciousness [*conscientia, a state of being aware*] (Descartes' Reply to Hobbes' Second Objection, translation and footnotes by Ross, 1975-1979).

The above reading of Descartes could be described as the "standard" view. The key features of this account are firstly, that Descartes equated mind with consciousness, thereby *expanding* the definition of "mind" to include anything with a *subjective* aspect – aches and pains, sensations and emotions as well as thoughts – and secondly, that the body is nothing but an unconscious, insentient piece of clockwork. Recently, this view has been vigorously challenged by Baker and Morris (1996), who argue that the foregoing account overlooks an ambiguity in Descartes' usage of the term "sense" or "feel" *(sentire):* the term may refer to a *bodily process* involving sense organs, or it may signify our cognitive apprehension or *judgement* that we are undergoing such a process. The authors argue that for Descartes, "sensing", defined according to the latter, more restricted meaning is inherently *propositional.* Having a sense-perception (defined as a mental event) does *not* involve introspecting a *qualium*, as many commentators suppose, but having a *thought* with a particular *content*, which relates to a process occurring in one's body. Mental events are inherently cognitive; the real dichotomy in Descartes is not between conscious and unconscious, but between rationality and mere bodily sentience, or the moral/intellectual and the animal. Only cognitive events qualify as mental events, and the Cartesian mind, Baker and Morris argue, is simply *intellectus,* or the rational soul.

Baker and Morris make a powerful case that the popular notion that Descartes regarded animals as inert, insensate automata is a total misrepresentation of his views. Such a view implicitly assumes that Descartes regarded the mind alone as active, and the body as purely passive – a legend which the authors challenge on scholarly grounds. Moreover, it overlooks the extensive textual evidence showing that Descartes attributed not only external senses but also *internal* senses to animals:

> The first internal sense apprehends pain and pleasure, hunger, thirst, and other bodily appetites or needs; the second the emotions such as fear, anger, joy and wonder. Both are essentially concerned with the preservation of animal welfare (Baker and Morris, reply to Nadler, 1997).

Thus according to Descartes, animals can *feel* pain with an internal sense; what they cannot do is *think* that they are in pain. Because animals are sentient, they have the capacity to anticipate things and pursue goals (imagination), and to initiate and sustain movement (locomotion).

More recently, Hennig (2004) has mounted a further challenge to the received interpretation of Descartes, arguing carefully that Descartes did *not* use the term *conscientia* in a radical new sense, as is commonly supposed, but continued to use it in a moral sense. *Conscientia*, on Descartes' account, refers to the formal cause of our thoughts being evaluable as true or false. The only innovation here is that whereas late medieval philosophers viewed *conscientia* as applying to actions (which are evaluable as good or bad), Descartes applied the term to *thoughts.* However, *conscientia* is *not* a second-order thought about a lower-level thought, but simply what it is about a thought that makes it subject to evaluation, from the perspective of an ideal evaluator who (unlike ourselves) can never be wrong – God (Hennig, 2004, pp. 40-42).

In contemporary philosophy, the concept of “mental events” is somewhat broader than that which is commonly imputed to Descartes: it is now acceptable to speak of *unconscious* as well as conscious mental processes. Some philosophers hold to a very strong version of this view. For instance, Lakoff and Johnson (1999, p. 10) insist that "most of our thought is unconscious, not in the Freudian sense of being repressed, but in the sense that it operates beneath the level of cognitive awareness, inaccessible to consciousness and operating too quickly to be focussed on". Other philosophers (e.g. Searle, 1999, p. 88) take a contrary view, distinguishing between *non-conscious* and *subconscious* brain states and recognising only the latter as mental, because they are at least potentially conscious. Searle (1992) embraces what he calls the Connection Principle, according to which a state cannot qualify as mental unless it is available to consciousness.

The question of whether *consciousness* can serve as the *hallmark of mind* remains highly controversial among contemporary philosophers: some (such as Searle) consider it to be the true criterion of the mental, others regard consciousness as a secondary phenomenon that can be explained by intentionality, and still others view the term “mental states” as an artificial category (see Jacob, 2003; Siewert, 2003, for a discussion of these views).

1.5 Problems associated with philosophical and scientific definitions of consciousness

1.5.1 Varieties of consciousness

The definition of “consciousness” has been refined a great deal since the time of Descartes. Contemporary philosophers distinguish several different senses of “consciousness”. The “standard” reading of Descartes interprets him as being concerned with what we would now “phenomenal consciousness” – or roughly, subjective experience. Block (1995) defined *phenomenally conscious* states as states with a subjective feeling or phenomenology, which we cannot define but we can immediately recognise in ourselves, distinguishing them from *access conscious* states, or mental representations which are poised for free use as a premise in *reasoning*, and for the direct rational control of *action* and *speech.* (Block (2001, 2005) has since amended his definition: the key feature of access consciousness is now said to be the fact that the information it contains is made widely available (or “broadcast”) in a global workspace to the brain’s “consumer” systems.) Another, higher-level kind of consciousness is *reflexive consciousness,* or an individual’s capacity for second-order representations of its mental states. It has yet to be shown that any non-human animals are capable of reflexive consciousness, and even for chimpanzees, the evidence for such an ability is highly questionable (see Hauser, Chomsky and Fitch, 2002; Nissani, 2004; see also Horowitz, 2002; Emery and Clayton, 2004; Block, 2005). As Lurz (2003) remarks, "it is rather implausible that my cat... upon espying movement in the bushes... is conscious *that* she sees movement in the bushes, since it is rather implausible to suppose ... that my cat has thoughts about her own mental states". As this thesis is about *minimal* minds, I shall say no more about this variety of consciousness.

Phenomenal consciousness, access consciousness and reflexive consciousness are all varieties of *state consciousness*, which is defined as consciousness as applied to mental states and processes, as opposed to *creature consciousness*, or consciousness as applied to a living organism (Rosenthal, 1986). The latter may be subdivided into *intransitive creature consciousness* – i.e. being awake, as opposed to asleep or comatose, and having at least some sensory systems which are receptive in the way normal for a waking state (Rosenthal, 1993, p. 355) – and *transitive creature consciousness* – i.e. the ability to perceive and respond to objects, events, properties or facts, thereby making one conscious *of* them. What distinguishes the latter is that it is inherently relational: “[w]hen a creature senses something or thinks about some object, we say that the creature is conscious *of* that thing” (Rosenthal, 1993, p. 355). Whereas a *creature* can be both intransitively conscious and transitively conscious *of* something, *mental states*, as such, are not conscious *of* anything; thus a mental state can only be intransitively conscious.

Questions of relevance to animals and other organisms here include the *ontological* question regarding the nature of phenomenal consciousness, as well as *epistemological* question of how we can know which creatures possess it (the Distribution Question) and how we can know what their experiences are like (the Phenomenological Question) (see Allen, 2005, for an overview). More fundamentally, we can ask whether the foregoing categories are well-defined for other kinds of organisms, and whether they actually carve nature at the joints, so to speak. I shall argue below that the different varieties of consciousness distinguished by contemporary philosophers generally fail on one or both counts, and I shall nominate some new categories of consciousness that may be more productive for future research.

The chief philosophical merits of Rosenthal's (1986) distinction between *creature consciousness* (a property of creatures) and *state consciousness* (a property of creatures' mental states) are firstly, that it highlights two different ways in which we apply the term "conscious" in everyday speech, and secondly, that it enables us to make sense of the claim that some of a creature's mental states – e.g. its background beliefs and desires – are not phenomenally conscious states. However, this preliminary distinction sheds no light on how we should go about answering the Distribution Question: which animals are phenomenally conscious?

1.5.2 Intransitive creature consciousness

My research led me to conclude that Rosenthal's definition of *intransitive* creature consciousness is scientifically defective, especially when applied to animals. First, it fails to distinguish between two very different criteria scientists use for sleep and wakefulness in animals - *behavioural* *criteria*, which are satisfied to some degree by nearly all animals, and *brain-based* (electrophysiological) *criteria*, which are only satisfied by mammals and birds (Shaw *et al.,* 2000, p. 1834). Nearly all animals (vertebrates and invertebrates) show “consolidated periods of activity and inactivity” (White, 2000). These periods of rest and activity are organized in a circadian rhythm of about 24 hours, in both vertebrates and invertebrates (Shaw et al., 2000, p. 1834). Most cold-blooded animals spend at least part of their day in a state of inertia or torpor, which scientists refer to as **behavioural sleep.** This state is not to be confused with what scientists refer to as *restful waking* (also sometimes referred to as quiet wakefulness, rest or drowsiness). In invertebrates, behavioural sleep is typically identified by behavioural quiescence, elevated arousal thresholds, species-specific stereotypic postures, and increased rest after a prolonged period of wakefulness. The occurrence of these traits has already been confirmed in a variety of insects, and it is believed that other invertebrates with complex eyes, including many arthropods – not just insects – and cephalopods, also probably sleep (Kavanau, 1997, p. 258; Shaw *et al.,* 2000, p. 1834). In cold-blooded vertebrates, behavioural sleep is characterised by similar criteria: behavioral quiescence, elevated thresholds for sensory stimuli, occlusion of the pupillary apertures (i.e. closed eyelids), and characteristic postures (Kavanau, 1997, p. 258). By contrast, animals in a state of restful waking have unelevated arousal thresholds and remain vigilant. Moreover, they exhibit only brief and intermittent occlusion of the pupils (Kavanus, 1997, p. 248).

In some vertebrate animals (mammals and birds), sleep satisfies certain electrophysiological criteria, in addition to behavioural criteria. This kind of sleep is called **true** or **brain sleep.** Brain sleep is defined by various criteria, including: EEG patterns that distinguish it from wakefulness; a lack of or decrease in awareness of environmental stimuli; and the maintenance of core body temperature (in warm-blooded creatures) (White, 2000). There is a *massive contrast* between the EEG patterns of human and animal patients in states of global unconsciousness (deep unconscious sleep, coma, persistent vegetative state, general anaesthesia and epileptic states of absence) and the EEG of patients in a state of waking consciousness (Shaw et al., 2000). The EEG patterns of animals who are *dreaming* share many similarities with that of animals in a wakeful state (Kavanau, 1997).

Second, Rosenthal's definition of intransitive creature consciousness cannot be usefully applied to most living organisms, because of its implicit assumption that a creature's sensory systems are receptive only when it is awake:

> Being conscious in this sense is, roughly, the opposite of being asleep or knocked out; we describe a person or other animal as being conscious if it's *awake*, and at least some of its *sensory systems* are *receptive* in the way normal for a waking state. Otherwise we say it's unconscious (Rosenthal, 1993, p. 355, italics mine).

In fact, it turns out that most living creatures, despite having sensory capacities, cannot be described as being either asleep or awake. As we shall see in a later chapter, *all* cellular life-forms, including bacteria, can be said to possess a wide range of sensory systems. However, the term "behavioural sleep" has a much narrower range of application. Behavioural sleep has been defined for some invertebrate *animals* (notably arthropods) and also for most *vertebrate* animals – although a few species, such as alligators, appear to lack it - but *not* for bacteria, protoctista, plants, or fungi (see Kavanau, 1997, p. 258), while true or “brain sleep” is found *only in mammals and birds.* Since a creature is said to possess *transitive* consciousness of something if it senses it, then we are forced to conclude that the domain of creatures possessing transitive consciousness is much larger than the domain of creatures possessing intransitive consciousness or wakefulness, however we define the latter.

Third, while Rosenthal's claim that "the property of being conscious that some mental states have and others lack is quite a distinct property from both [intransitive] creature consciousness and transitive consciousness" (1993, p. 355) is entirely correct if one is attempting to formulate a *conceptual* distinction, it is scientifically counter-productive, insofar as it overlooks what appears to be a *nomic connection* between an animal's satisfying the *brain-based criteria for wakefulness* and its being able to give an *accurate report* of its surroundings - which is how scientists routinely assess the presence of consciousness, as they define it. Indeed, some neuroscientists believe brain sleep to be intimately related to having conscious states with a subjective feel, or what philosophers call *phenomenal consciousness* (Cartmill, 2000; Baars, 2001; White, 2000), and a few (Baars, 2001; Cartmill, 2000) have even suggested that wakefulness - defined according to brain criteria - is a reliable indicator of phenomenal consciousness across *all* animal species. If they are right, then it may turn out to be a *law of nature* that animals possessing the right kind of *intransitive* creature consciousness – namely, *brain* wakefulness – are also *phenomenally* conscious. Although the phenomenon of *dreaming* may tempt us to regard phenomenal consciousness as something quite distinct from intransitive creature consciousness, it would be unwise to ignore the strong affinities between the EEG patterns of animals that are dreaming and those in a state of brain wakefulness (Kavanau, 1997), as well as the fact that brain wakefulness (which is a form of intransitive creature consciousness) appears to be a sufficient condition for phenomenal consciousness in human beings. The current philosophical practice of compartmentalising these two varieties of consciousness therefore needs to be questioned.

On the other hand, the condition of persistent vegetative state (PVS), which has been defined as "chronic wakefulness without awareness" (JAMA, 1990), demonstrates that *behavioural* wakefulness is not a sufficient condition for phenomenal consciousness. PVS patients display a variety of wakeful behaviours, all of which are generated by their brain stems and spinal cords. Studies have shown that activity occurring at this level of the brain is not accessible to conscious awareness in human beings (Rose, 2002a, pp. 13-15; Roth, 2003, p. 36).

Laureys (2005) has written extensively about the vegetative state. The following points emerge from current research:

- The vegetative state is quite different from a *coma.* Patients in pathological or pharmacological coma (that is, general anaesthesia) are unconscious because they cannot be awakened. By contrast, patients in a vegetative state manifest behavioural wakefulness: they may grimace, cry or smile (but not in response to specific external stimuli), and they also display characteristic movements of their eyes, head and limbs.
- The vegetative state is not to be confused with *minimal consciousness* – a state in which the patient is aware but unable to communicate, due to perceptual, attentional and motor deficits. Brain scans reveal some important differences between minimal consciousness and the vegetative state. For instance, PET and fMRI reports using complex auditory stimuli have demonstrated large-scale network activation in the minimally conscious state, which is normally not observed in vegetative patients. Sometimes a minimally conscious patient is misdiagnosed as vegetative – mainly owing to the difficulty of detecting awareness in patients with fluctuating levels of mental arousal and a combination of physical impairments. Laureys suggests that minimally conscious

patients might be identified more easily in the future, by scanning their brains while they are asked to perform a mental imagery task.

- According to Laureys, "[t]he common hallmark of the vegetative state seems to be metabolic dysfunction of a widespread cortical network encompassing medial and lateral prefrontal cortices and parietal multimodal associative areas" (2005, p. 558).
- Patients who are *properly* diagnosed as being in a vegetative state lack phenomenal consciousness: they *do not actually feel anything.* "Studies using high-intensity electrical stimulation (experienced as painful in controls) showed robust post-stimulus activation in [the] brainstem, thalamus and primary somatosensory cortex in each of 15 well-documented vegetative patients ... Importantly, higher-order areas of the pain matrix (that is, secondary somatosensory, insular, posterior parietal and anterior cingulate cortices) were *not* activated" (Laureys, 2005, p. 557, italics mine). Likewise, "auditory stimulation in unambiguously vegetative patients activated primary auditory cortices, but not higher-order multimodal areas from which they were disconnected" (ibid). These results suggest that neural activity in the brain's primary cortices cannot result in phenomenal consciousness unless it also activates the brain's frontal-parietal network.

The above findings lend strong support to the view that there is a *nomic* connection between co-ordinated activity in the brain's higher-order multi-modal areas and the occurrence of phenomenal consciousness.

1.5.3 Transitive creature consciousness

It has now been established that an animal can possess *transitive consciousness* in the absence of phenomenal consciousness (*pace* Dretske, 1995). The vomeronasal system, which responds to pheromones and affects human behaviour, but is devoid of phenomenality (Allen, 2003, p. 13) is one good example; the phenomenon of blindsight in humans and monkeys (Stoerig and Cowey, 1997, pp. 536-538; p. 552) is another.

My research has shown that “transitive consciousness” is a vague philosophical concept, as the term “sense” can be used broadly or narrowly. If we define transitive creature consciousness broadly, it could be said to be a property of all cellular organisms, as they all possess a range of *sensory capacities*. According to a narrower definition, however, only organisms with a nervous system can be said to possess "true" senses (Cotterill, 2001). I shall discuss this question in a subsequent chapter.

Additionally, the *criteria* for possession of transitive creature consciousness need to be more clearly specified. For instance, blindsight varies across patients in its degree of severity, and the specificity of the responses shown by these patients varies accordingly (Stoerig and Cowey, 1997, pp. 536-538). Which of these responses count as *bona fide* instances of transitive creature consciousness?

1.5.4 Varieties of state consciousness

As scientific research into *reflexive consciousness* is still too tentative for us to draw any conclusions about its occurrence in non-human animals (Nissani, 2004), I shall refrain from discussing it here, and confine myself to the two other varieties of state consciousness discussed by philosophers.

1.5.4.1 Access consciousness

The most striking feature of the debate regarding *access consciousness* is that the motivation for distinguishing it from *phenomenal consciousness* has changed from a philosophical one to a scientific one. According to Block's original definition, "a representation is access-conscious if it is actively poised for direct control of reasoning, reporting and action" (1998, p. 3). Block argued that philosophers had erred in confusing this access-related function of consciousness with the role played by phenomenally conscious states, which he defined as states with a subjective feel or phenomenology, which we cannot define but can immediately recognise in ourselves (1995). Nonetheless, he regarded the two kinds of consciousness as virtually co-extensional: only in rare instances could one be found without the other.

More recently, however, Block (2001, 2005) has proposed a new basis for distinguishing between access and phenomenal consciousness. The distinction is now made on neurological-cum-behavioural grounds. The recent discovery that different visual perceptions in the brain appear to compete with each other for control of the neurological systems that govern an animal's actions is now taken as evidence for the existence of two kinds of consciousness: (i) a weaker, phenomenal variety which has its own special flavour or content (e.g. red as opposed to green) without necessarily having any role in an animal's actions, and (ii) a stronger, access-related variety, which is hypothesised to occur when one of these phenomenal states gets to dominate a "global workspace" in the brain, broadcast its information across this workspace, and thereby control the creature's actions. The key idea here is that the neural correlates of phenomenal consciousness compete for domination of the workspace, and the winner, which gets to broadcast its signal globally, and govern actions, becomes the neural correlate of access consciousness (Block, 2005). Experiments with stimuli near threshold level in monkeys and humans indicate that there is “an intermediate level representation that can be disconnected from access either by raising the perceptual decision criterion or by decreasing saliency of the stimulus” (Block, 2005, p. 49). Block suggests that this representation is part of the neural correlate of phenomenal consciousness.

It appears that access consciousness can exist in the absence of phenomenal consciousness, in certain situations. The strongest evidence for this claim comes from recent studies of the *mammalian visual system* (discussed in Carruthers, 2004b). Research by Milner and Goodale (1995) suggests that each human brain has two visual systems: a *phenomenally conscious system* that allows the subject to select a course of action but which she cannot attend to when actually executing her movements, and an *access-conscious system* that guides her detailed movements but is not phenomenally aware. However, the case of the *distracted driver*, who is supposedly able to navigate his car home despite being oblivious to his visual states, is *not* a convincing example of access consciousness in the absence of phenomenal consciousness (for a discussion see Wright, 2003, who cites three driving studies showing that driving requires a certain minimum level of attention to the road).

Several comments are pertinent here. First, I would like to point out that the distinctions proposed by Block, while neurologically well-supported, are currently limited to one sensory modality (vision), and a limited class of animals (primates).

Second, Block appears to be inconsistent in claiming that not only do some non-linguistic animals (e.g. chimps) have access consciousness states (1995, p. 238), but "very much lower animals" are access-conscious too (1995, p. 257), while at the same time insisting in other passages that access consciousness presupposes *rationality.* According to Block, "a representation is access-conscious if it is actively poised for direct control of reasoning, reporting and action" (1998, p. 3). Direct control occurs "when a representation is poised for free use as a premise in reasoning and can be freely reported" (1998, p. 4). On such a definition, very few animals would qualify for access consciousness. In his most recent writings, Block defines “rational control of action” very broadly: apparently it includes “systems of memory, perceptual categorization, reasoning, planning, evaluation of alternatives, decision-making, [and] voluntary direction of attention” (Block, 2005, p. 47). But this definition is now too broad: as we shall see in a subsequent chapter, memory can be found even in bacteria, while perceptual categorisation and attention to salient stimuli have been verified in flies.

Finally, Bayne (2003) highlights a gap in Block’s definition of access consciousness: it fails to account for cases where the content of consciousness is poised for the direct but *partial* (as opposed to global) control of thought and action. What should we say when the “global workspace” isn’t fully global? Bayne documents several kinds of brain diseases where the content of a patient’s conscious states is only *partially* access-conscious – that is, available for the rational control of only a restricted subset of the full range of actions that the patient is capable of performing. Additionally, experiments have shown that even “normal” subjects give inconsistent responses when reporting in various modes (blinking, pressing a button, and saying “yes”) on the presence of a transient stimulus (a light that flashes for less than one fifth of a second). Verbal responses turned out to be the *least* accurate. Strangely, subjects were completely unaware of the inconsistency of their responses. Bayne suggests that “access” may vary along two separate dimensions: not only the *degree* to which it is poised to control actions (already conceded by Block), but also the *range* of actions that it is capable of controlling (which may vary from local to global).

1.5.4.2 Phenomenal consciousness

While Block (1995) defines *phenomenally conscious* states as states with a subjective feel or phenomenology, which we cannot define but can immediately recognize in ourselves, Van Gulick (2004) defines phenomenal consciousness in a narrower, more technical sense, as applying to the *overall structure* of experience: as such, it involves far more than sensory *qualia* (raw subjective feelings, such as the experience of seeing red). In this thesis, I shall use the term “phenomenal consciousness” in the broader sense defined by Block (1995), in conformity with the more common philosophical usage.

In a recent article, Block (2001) ditched the term "phenomenal consciousness" in favour of what he called *phenomenality.* Rosenthal (2002a) has criticised Block's (2001) account for its *ambiguity* between two very different mental properties, which Rosenthal refers to as *thin phenomenality* (the occurrence of a qualitative character without a subjective feeling of “what it's like”, e.g. when the subject is presented with subliminal stimuli which she may not be consciously aware of) and *thick phenomenality* (the subjective occurrence of mental qualities). Rosenthal considers only the latter to be truly conscious.

Block (2005) now prefers to use the term “phenomenal NCC” to denote the *neural correlates* of phenomenal consciousness. Block cites research by Super *et al.* (2001) for the occurrence of *two distinct modes* of sensory processing in the visual cortex of monkeys as evidence for an intermediate representation in the monkey visual cortex, which he tentatively identifies with the phenomenal NCC. The strength of this representation normally varies in proportion to the saliency of the stimulus in the monkey's field of vision. However, when the saliency of the stimulus is very low, or when the percentage of "catch trials" (where experimenters try to trick the monkey into making a false response by occasionally presenting it with a homogeneous pattern instead of a stimulus in its field of vision) is raised, the representation's strength is *no longer connected* to the monkey's ability to *access* the stimulus (i.e. saccade to it, in order to obtain a reward). Block argues that the representation must therefore be distinct from the access NCC, and that it could be part of the *phenomenal* NCC. Interestingly, this representation also disappears under anaesthesia.

Intriguing as these results are, it should be kept in mind that they apply to only *one* sensory modality in one kind of animal (a monkey). It would be premature to jump to conclusions on the basis of these data. However, my research for this thesis has uncovered a more promising source of material: the work of neuroscientists, who have been investigating the behavioural and neurological conditions for consciousness for the past 70 years.

1.5.4.2.1 Are primary and higher-order consciousness phenomenal?

Neuroscientists commonly distinguish between *primary* and *higher-order* forms of consciousness (Edelman, 1989). Both forms appear to qualify as *phenomenal* in the philosophical sense. Higher-order consciousness "includes awareness of one's self as an entity that exists separately from other entities" (Rose, 2002a, p. 6), while neurologists use the term *primary consciousness* (also called "core consciousness" or "feeling consciousness") to refer to "the moment-to-moment awareness of sensory experiences and some internal states, such as emotions" (Rose, 2002a, p. 6). The latter definition could easily be interpreted as synonymous with a lower grade of phenomenal consciousness, with one caveat: the word "of" in the definition appears to imply the claim that animals need to be conscious *of* their experiences, in order to qualify as being *conscious* at all. Certainly, HOR theorists such as Rosenthal (1986, 2002a) and Carruthers (2000, 2004b) would readily agree with this claim, as does Lurz (2003), who has put forward his own, "same-order" account of consciousness; on the other hand, defenders of *first-order* representational accounts of consciousness (e.g. Dretske, 1995) flatly reject the claim. We might do better to re-define primary consciousness as "the moment-to-moment awareness *that characterises* sensory experiences and some internal states, such as emotions". Rose (2002a) adds that "[m]ost discussions about the possible existence of conscious awareness in non-human animals have been concerned with primary consciousness" (2002a, p. 6).

1.5.4.2.2 Can there be phenomenal consciousness in the absence of primary consciousness?

The majority of neurologists consider *primary consciousness* to be the most basic form of subjective awareness. However, a few authors such as Panksepp (1998, 2001, 2003f) and Liotti and Panksepp (2003) have proposed that we possess *two* distinct kinds of consciousness: (i) *cognitive consciousness*, which includes perceptions, thoughts and higher-level thoughts about thoughts and requires a *neocortex* (a six-layered structure in the brain which comprises the bulk of the brain's outer shell or cerebral cortex – the neurological consensus (Nieuwenhuys, 1998; Rose, 2002a, p. 6) is that only mammals possess this laminated structure in its developed form), and (ii) *affective consciousness* which relates to our feelings and arises within the brain's *limbic system*, with the *anterior cingulate cortex* playing a pivotal role. Panksepp considers affective consciousness to be the more primitive form of consciousness. It is certainly true that the *neural processing* for cognitive and emotional reponses in humans and other animals is quite distinct, which refutes the view that emotions are simply (conscious or unconscious) *cognitions* (LeDoux, 1999, p. 69). The term "limbic system" has been criticised as outdated by some neurologists (e.g.

LeDoux, 1999, pp. 98-103); however, Panksepp defends it as a useful heuristic concept (1998, pp. 57, 71, 353). If Panksepp is correct, then even animals that do not satisfy the neurological requirements for what he calls cognitive consciousness may still possess some form of phenomenal awareness. For instance, Cabanac (1999, 2003) contends that the *hedonic behaviour* of certain reptiles, which are willing to expose themselves for a short time to an aversive stimulus in order to procure some attractive stimulus, indicates that they experience emotions. Amphibians, on the other hand, display no such behaviour.

While the cumulative evidence Panksepp cites for two kinds of consciousness in the brain is impressive, the distinction between them is somewhat over-drawn: affective consciousness still involves crude, low-level processing of sensory inputs and hence minimal cognition. (The notion of phenomenal consciousness in the complete absence of cognition is philosophically problematic.) For instance, when a hiker encounters a snake while walking in the woods, what happens first is that some low-level "quick and dirty" cognitive processing of sensory data occurs in the *thalamus* (which identifies "gross" features such as objects that look vaguely like a snake) before the information passes to the *amygdala* (which mediates the emotional response) and independently, to the *visual cortex* (which handles the "fine-grained" cognitive task of identifying the stimulus as a snake) (LeDoux, 1999, p. 166). (The thalamus and amygdala are both traditionally classified as part of the limbic system – a term LeDoux eschews.) In any case, both cognitive and affective consciousness fall under the definition of *primary consciousness* proposed above: “the moment-to-moment awareness that characterises *sensory experiences* and some internal states, such as *emotions*”. Nevertheless, Panksepp’s distinction alerts us to the possibility that some phenomenally conscious individuals (including non-human animals) who are unable to describe what they see may still be capable of manifesting their awareness of their surroundings through their emotional responses to events in their environment.

1.5.4.2.3 Criteria for primary consciousness and their relevance to the Distribution Question

1.5.4.2.3.1 Behavioural criteria for primary consciousness

Criteria used by neuroscientists to verify an individual's possession of primary consciousness may be said to fall into two broad categories: observational (or behavioural) and neurological. I shall discuss neurological criteria below. The standard observational criterion used to establish the occurrence of primary consciousness in animals is *accurate report* (AR). In humans reports do not have to be verbal; pressing a button, or any other voluntary response, is routinely accepted as adequate in research (Baars, 2001, p. 35). The philosophical questions that are of relevance here are: whether having primary consciousness is indeed a *sufficient* condition for having phenomenal consciousness; whether the behavioural criteria used to identify primary consciousness are actually *suitable* for that purpose, and if not, what other criteria are; and finally, to what degree they can be applied to *non-human* animals.

The first question can be answered in the affirmative, even if we replace Rose's (2002a) definition of primary consciousness with the weaker definition which I have proposed. Both definitions make use of the term "moment-to-moment awareness" – a phrase which is clearly meant to convey a subjective feeling.

Turning to the second question, some neurologists (Seth, Baars and Edelman, 2005) have recently acknowledged problems with the use of *accurate report* as a criterion for primary consciousness, and have argued that purely behavioural criteria are simply inadequate for the task of identifying it, as *non-verbal* accurate report is difficult to distinguish from mere *sensory* *discrimination* or *stimulus categorization:*

> [B]ehavioural measures risk a slippery slope. In principle, it is difficult to make a distinction between AR and other behavioral indices of sensory categories… Even computers can produce an output that resembles AR, though few scientists would call them conscious on this basis. Further, stimulus categorization can take place unconsciously in humans (Seth, Baars and Edelman, 2005, p. 120).

Other behavioural criteria have been proposed by Panksepp (1998, 2001, 2002b, 2003c, 2003d, 2003f), who considers the most primitive form of consciousness to be *affective* rather than cognitive. While the collective evidence marshalled by Panksepp for conscious feelings in animals is highly persuasive, the notion of a separate "affective consciousness" remains highly controversial in neurological circles. Of greater significance is the fact that Panksepp himself (2002b) makes no claim to identify affective consciousness on the basis of behavioural criteria alone: he also employs neurological and psycho-pharmacological criteria, which I shall discuss below.

Aversive behaviour is often regarded by non-specialists (including some philosophers) as an indicator of the phenomenally conscious experience we call *pain.* The term “aversive behaviour” may include any of the following: *stress* responses; *nociceptive* responses to noxious stimuli; the release of pain-killing *opiates* which are found within the brainstem; the ability to learn to avoid certain “unpleasant” flavours *(flavour aversion learning);* the ability to undergo *classical* and *instrumental conditioning* and learn to avoid noxious stimuli; self-administration of *analgesics;* and *pain-guarding.* However, it has been established by neuroscientists that most of the responses described above – including behaviours such as withdrawal of the stimulated body part, leg locomotion, struggling, facial grimacing, and even vocalisation (Rose, 2002a, pp. 16-17) – are regulated at levels of the brain *below the level of primary consciousness.* In all vertebrates, the fundamental behavioural reactions to injurious stimuli are generated by neural systems in the spinal cord and brainstem. The same reactions occur in people who are *unconscious* - for example, people with massive cortical damage and children born without cerebral hemispheres (Rose, 2002a, pp. 13-14, 17). Rose (2002a, p. 14) discusses six human patients (first described in Jouvet, 1969), who had suffered the complete loss of their cerebral cortex. Some of these decorticate patients exhibited behaviours such as grimacing and cries evoked by noxious stimuli, and pushing at the hands of the examiner. Moreover, it has been shown experimentally that human beings are *never* aware of the neural activity taking place below the level of the cortex – be it in the spinal cord, brainstem or cerebral regions beneath the neocortex (Rose, 2002a, p. 6; see also Roth, 2003). Since the International Association for the Study of Pain (1999; see Rose, 2002a) defines "pain" as an intrinsically *conscious* experience, clinical assessments of pain in human patients are never made solely on the basis of these behavioural criteria.

Finally, *hedonic behaviour* has been proposed as an indicator of the phenomenally conscious experience we call pleasure. It has been suggested that behaviours such as a willingness to make *hedonic trade-offs* whereby an individual will expose themselves for a short time to an aversive stimulus in order to procure some attractive stimulus (Cabanac, 2003), or the occurrence of "rational" and "irrational" forms of pursuit (described in Berridge, 2003a, 2003b) can be treated as yardsticks of pleasure. *Irrational* pursuit occurs when an individual seeks something it neither likes nor expects to like, and can be identified when the individual, under the influence of some drug (e.g. dopamine), is suddenly presented with a "rewarding" stimulus, which then triggers hyperactive pursuit of that stimulus. However, none of these behaviours necessarily indicates phenomenal awareness. The ability to make trade-offs could merely indicate an ability to rank stimuli on a scale of attractiveness and evaluate the net attractiveness of a positive and a negative stimulus, while studies have shown that irrational desires *need not be conscious:* humans can be influenced to like or dislike something simply by subliminal exposure to stimuli which they report being unaware of (Berridge, 2001, 2003a, 2003b).

All of these behavioural indicators (nonverbal accurate report, affective behaviour, aversive behaviour and hedonic behaviour) which have been proposed for phenomenal awareness have been extensively investigated in *non-human animals.* Since the criteria for accurate report include *nonverbal* communication, there is nothing to prevent them being applied to other animals. Recent experiments by Stoerig and Cowey (1997, p. 552) have shown that a monkey can be trained to respond to a stimulus in its visual field by touching its position on a screen, and to a blank trial (no stimulus) by touching a constantly present square on the screen that indicates "no stimulus". The monkey's ongoing responses fit the requirements for a nonverbal "accurate, verifiable report" (Baars, 2001) indicating "sustained awareness of the environment" (Rose, 2002a, p. 6). Moreover, recent experiments by Logothetis with *binocular rivalry* have demonstrated that the humans and monkeys make identical reports about what they see when conflicting data is presented to their left and right two visual fields (Block, 2003). According to Stoerig and Cowey (1997, p. 552), *lack* of awareness has also been experimentally verified in studies of monkeys with *blindsight,* a condition in which patients with damage to the visual cortex of the brain lose their subjective awareness of objects in a portion of their visual field, but sometimes retain the ability to make visual discriminations between objects in their blind field. It is also interesting to note that pigeons can respond variably to the ambiguity in figures like the

Necker cube, which suggests that their subjective impression of its orientation can "flip" (Butler, Manger, Lindahl and Arhem, 2005).

Two philosophical questions are pertinent here. First, should these nonverbal responses be regarded as *introspective reports* by the monkeys on the content of their phenomenal experiences, or simply as *environmental reports* on what they see in front of them, as Block (2005) suggests? The latter suggestion is more parsimonious and accounts for the behaviour equally well. Second, is the ability to report on the objects in one's environment a *sufficient* condition for having phenomenal consciousness? Carruthers thinks not; he argues that phenomenal awareness requires an ability on the subject's part to "draw a distinction between the way things *are* and the way they *seem* or *appear"* (Carruthers, 2004b). Other philosophers (Allen, 2005) have also proposed that any animals which can learn to correct their perceptual errors are phenomenally conscious, but what sets Carruthers apart is that he regards this ability as a *necessary* condition for possessing subjective awareness. The only findings I have been able to uncover in this field are negative: experiments with monkeys fitted with glasses inverting the retinal image showed that in the monkeys (unlike human beings) this completely disrupted their behavior, and they entered a long period of inactivity (Leontev, 1978).

An additional reason for caution on the use of nonverbal accurate report as an indicator of phenomenal consciousness is the admission by Seth, Baars and Edelman (2005, p. 120) that "it is difficult to make a distinction between AR [accurate report] and other behavioral indices of sensory categories". Moreover, "the ability to generalize across classes of stimuli is extremely widespread in the animal kingdom" and has been demonstrated in insects (Seth, Baars and Edelman, 2005, p. 120).

My own view on the controversy regarding the use of nonverbal acts as indicators of phenomenal awareness is that Carruthers' negative point is well-taken: *nonverbal* accurate report is not only inadequately defined, but also inherently incapable of serving as a sufficient criterion for the possession of phenomenal consciousness. On the other hand, Carruthers' positive contention, that the linguistic capacity to distinguish between the way things *are* and the way they *seem* is a pre-requisite for having subjective awareness, remains an unproven assertion, and a highly contentious one at that, as it entails that not only monkeys but also human beings (e.g. pre-linguistic infants and people with severe mental disabilities) who lack the conceptual and linguistic wherewithal to formulate the abstract notions of appearance versus reality, possess *no* phenomenal consciousness whatsoever.

If nonverbal reports cannot establish the presence of phenomenal consciousness in animals, can *affective behaviour* do so? Panksepp (2002b) contends that several converging lines of evidence point to the occurrence of a primitive affective consciousness in animals: "(i) behavioral reinforcement studies; (ii) place preference-aversion studies; (iii) manifest and ubiquitous emotional vocalizations; (iv) neuro-ethological studies evoking the same emotional behavior from the same human/animal brain analogs and (v) the coherent translations between human and animal psychopharmacological work" (Panksepp, 2002b). Panksepp has certainly performed some remarkable research in this area. Recently, Panksepp and Burgdorf (2003c), in an article entitled "'Laughing' rats and the evolutionary antecedents of human joy?" discussed their recent discovery of play- and tickle-induced ultrasonic vocalisations in rats which are analogous to laughter in human children. The authors identified no less than twelve points of resemblance between rat "laughter" and children's laughter and argued that alternative non-mentalistic explanations were poorly supported by the evidence. However, the point I wish to make here is that since affective behaviour is simply one component of a suite of evidence cited by Panksepp and Burgdorf (2003c), pointing to the occurrence of feelings in these animals, it cannot be a sufficient condition for phenomenal consciousness. Panksepp (2002b) is well aware of this: he additionally cites extensive neurological and pharmacological analogies between human and animal brains and nervous systems in order to cement his case for animal consciousness.

As we saw above, various kinds of *aversive behaviour* have been proposed by philosophers as indicators of the subjective experience we call *pain.* My survey of the scientific literature yielded the following results on their prevalence in animals:

- *stress responses* occur in all cellular organisms, including bacteria (Smith, 1991);
- nearly all animals display *nociceptive responses* to noxious stimuli (sponges and sharks being two notable exceptions) (Smith, 1991; Rose, 2002a, 2002b, 2003a, 2003b);
- *opiates,* which kill pain in human beings, can be found in the brainstems of various kinds of animals, but are thought to have originally evolved for a quite different purpose, as part of the immune system's response to noxious stimuli (Stefano, Salzet and Fricchione, 1998);
- *classical and instrumental conditioning* occur in all animals with a true nervous system, including worms (Abramson, 1994);
- *flavour aversion learning* is found in reptiles, mammals and birds, but not in amphibians (Paradis and Cabanac, 2004);
- *self-administration of analgesics* is known to occur in mammals and also in birds (Grandin and Deesing, 2003; see also Church, 2000, p. 14);
- the same holds true for *pain-guarding* (Grandin and Deesing, 2003).

However, most of these responses are generated by neural systems in the spinal cord and brainstem. Since the same responses occur in people who are not phenomenally conscious, and since research has shown that human beings are *never* aware of the neural activity taking place below the level of the cortex (Rose, 2002a; Roth, 2003), we cannot use these criteria as evidence of phenomenally conscious pain in animals. The same goes for reactions such as facial grimacing and vocalisation, which occur in people who are not conscious - for example, people with extensive cortical damage and children born without cerebral hemispheres (Rose, 2002a, pp. 13-14, 17). Likewise, although the tests that are routinely employed by analgesia researchers to assess the efficacy of analgesics (such as the rodent tail-flick test, the hot-plate test, the paw-pressure test and the paw-withdrawal test) make use of *mammalian models of pain* that attempt to quantify the degree of discomfort experienced by laboratory animals, the aversive reactions of these animals to noxious stimuli (heat and pressure) are actually controlled by their brainstems. Thus they do not measure pain as such, but *nociception.*

Rose (2002a), after conducting an exhaustive review of the literature relating the neurology and behaviour of fish and the clinical indicators used by neurologists to assess pain, concluded that consciousness in fish is "a neurological impossibility" (2002a, p. 2). More recently, Rose (2003a) has written a devastating critique of a much-publicised report by Sneddon, Braithwaite and Gentle (2003) which claimed to have identified evidence of *pain guarding* in fish.

There are other behavioural responses to pain which *do* indicate the presence of phenomenal consciousness: the *cognitive-evaluative* components of pain (e.g. attention to the pain as a perceived threat to the individual, and conscious generation of strategies for dealing with the pain) and *affective* components (experiencing the pain as emotionally unpleasant). However, Rose (2002a) maintains that the occurrence these components of pain can only be definitively established by the patient's *verbal report.* For instance, people who have had surgery to their anterior cingulate gyrus to alleviate chronic pain report that the pain is still there (sensory-information component) but that it no longer bothers them (Rose, 2002a, pp. 19-21).

Hedonic behaviour has also been investigated in animals. Cabanac (1999, 2002, 2003) has performed experiments suggesting that reptiles and higher vertebrates exhibit this kind of behaviour, while fish and amphibians do not. However, it was argued above that an alternative, more parsimonious explanation is possible, requiring only that an organism is able to *remember* stimuli previously encountered, *rank* their relative attractiveness and *evaluate* the net attractiveness of an activity containing positive and negative components. Attraction, as we argued in the Introduction, need not indicate awareness; nor does summation. I shall discuss memory in a subsequent chapter.

Since primary consciousness is the most rudimentary form of phenomenal consciousness, our failure to find behavioural criteria which are sufficient to establish the occurrence of primary consciousness means that there are no purely *behavioural* criteria that can serve as indicators of *phenomenal* consciousness, in humans or other animals.

1.5.4.2.3.2 Neurological criteria for primary consciousness

Neurologists have nevertheless managed to find *three neurological identifying traits* of what they call *primary consciousness:* (i) a distinctive EEG signature of irregular, low-amplitude, and fast electrical activity in the brain ranging from 12 to 70 Hz; (ii) a thalamus, a cortex and recursive (or reentrant) pathways between the two (and probably also a third region, the basal ganglia); and (iii) widespread activation of the brain's *neocortex* or isocortex (a laminated structure which surrounds the brain), not just in the local areas of the sensory cortex, but also in the parietal, prefrontal and medial-temporal areas of the cortex (Seth, Baars and Edelman, 2005). (See Figures 1.1 and 1.2.) Human consciousness appears to require brain activity which is *diverse, temporally conditioned and of high informational complexity.* The neocortex satisfies these criteria because it has two unique structural features: *exceptionally high connectivity* within the neocortex and between the cortex and thalamus; and enough *mass* and *local functional specialisation* to permit regionally specialised, differentiated activity patterns (Rose, 2002a, p. 7, italics mine). According to Rose, "sources such as clinical neuropsychology (Kolb and Whishaw, 1995), neurology (Young et al., 1998; Laureys et al., 1999, 2000a-c), neurosurgery (Kihlstrom et al., 1999), functional brain imaging (Dolan, 2000; Laureys et al., 1999, 2000a-c), electrophysiology (Libet, 1999) and cognitive neuroscience (Guzeldere et al., 2000; Merikle and Daneman, 2000; Preuss, 2000)" lend convergent support to the view that the human capacity for conscious awareness depends on the neocortex (Rose, 2002a, p. 6). It is now believed that slow-wave sleep, coma and PVS cause a loss of primary consciousness precisely because in these states, the *ability to integrate information* between different regions of the cerebral cortex is greatly reduced (Tononi, 2004; Baars, 2003).

The neocortex is divided into *primary and secondary regions* (which process low-level sensory information and handle motor functions), and the *associative regions* (see Figure 1.2). Brain monitoring techniques indicate that in human beings, only processes that take place within the associative regions of the cortex are accompanied by consciousness; activities which are confined to the primary sensory cortex, or processed outside the cortex, are *inaccessible* to consciousness (Roth, 2003, pp. 36, 38; Rose, 2002a, p. 15). Consciousness thus depends on the functions of the *associative cortex*, not primary cortex. The associative regions are distinguished by their *high level of integration and large number of connections with other regions of the brain* (Roth, 2003, p. 38).

Other sites in the brain have been proposed for primary consciousness. The *cerebellum*, located at the back of the brain (see Figure 1.1), "contains probably more neurons and just as many connections as the cerebral cortex, receives mapped inputs from the environment, and controls several outputs"; however, "lesions or ablations indicate that the direct contribution of the cerebellum to conscious experience is minimal" (Tononi, 2004), and "removal of the cerebellum does not severely compromise consciousness" (Panksepp, 1998, p. 311) - apparently because different regions of the cerebellum tend to operate independently of one another, with *little integration of information between regions* (Tononi, 2004).

The *limbic system* (and in particular, the *anterior cingulate cortex*) (see Figure 1.3) has been proposed by other authors (Panksepp, 1998, 2001, 2003f; Liotti and Panksepp, 2003) as the site of a primitive *affective consciousness,* as distinct from the *cognitive consciousness* (generated by the cerebral cortex) that processes sensory inputs. However, the notion that the brain has an autonomous centre of consciousness residing in the limbic system is a highly contentious one, as it appears to conflict with brain monitoring data cited above (Roth, 2003, p. 36). Additionally, the very term “limbic system” has been attacked as outdated by some scientists (LeDoux, 1998, pp. 98-103), although Panksepp (1998, pp. 57, 71, 353) defends it as a useful heuristic concept. Moreover, Panksepp’s assertion that the anterior cingulate cortex (ACC) forms part of a “limbic region” which is separate from the cerebral cortex has been contested by Allman, Hakeem, Erwin, Nimchimsky and Hof (2001), who argue on anatomical grounds that the ACC is actually part of the cerebral cortex, as it also has a complex layered structure. Finally, the anterior cingulate cortex, like the neocortex, is peculiar to mammals – a fact which creates difficulties for the hypothesis that the emergence of affective consciousness in evolutionary history *predated* the appearance of the mammalian neocortex.

On the other hand, it needs to be kept in mind that there are currently several competing theories as to how and where consciousness is generated in the brain. Butler, Manger, Lindahl and Arhem (2005) describe *four classes of models* that attempt to locate the brain's consciousness generators: a bottom-up sensory approach (Crick, Koch), a top-down sensory approach (Edelman, Tononi), a bottom-up motor approach (Eccles) and a top-down motor approach (Cotterill). The only thing the models seem to have in common is that *some* regions of the cortex are required for the generation of consciousness, but exactly which regions are critical remains a subject of vigorous controversy (see Figure 1.4). Both of the top-down models include regions *outside* the cortex as part of the brain's generating mechanism for consciousness. In these two models, the limbic system – and other regions as well, in Cotterill's model – is *part* of the mechanism, although neither of these models posits the limbic system as the site of an independent centre of consciousness, as Panksepp does.

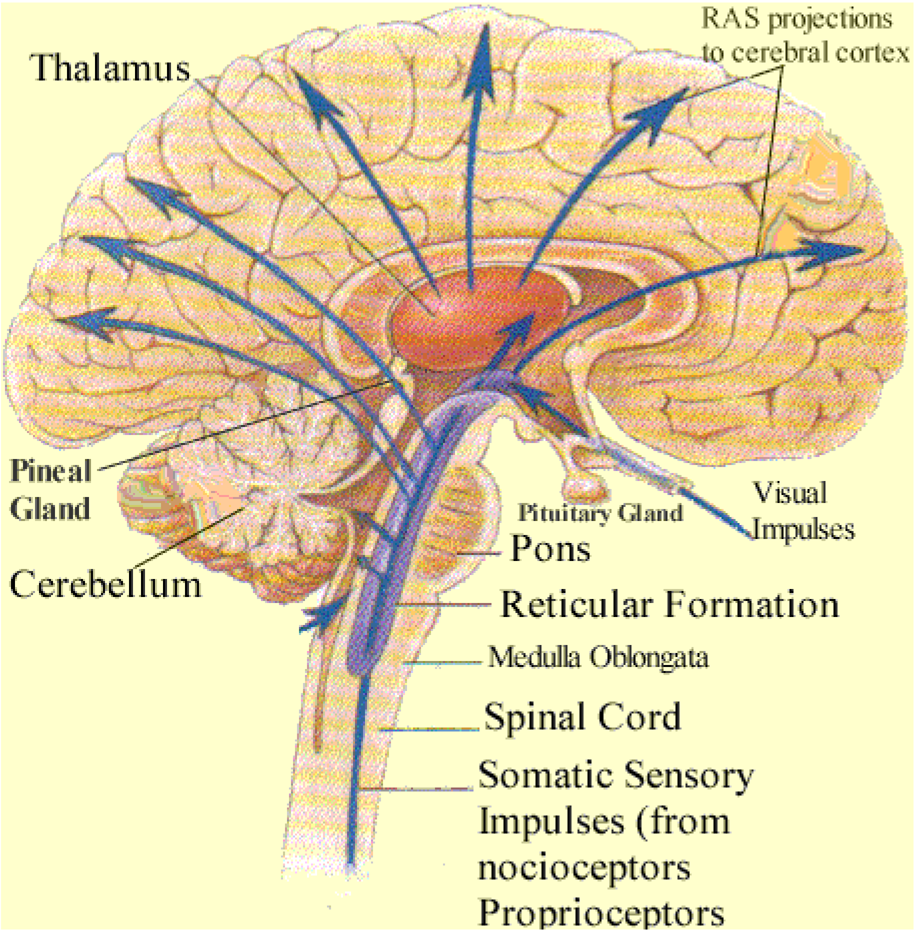


Figure 1.1 How primary consciousness is generated in the brain. The reticular activating system (RAS) – which comprises parts of the medulla oblongata, the pons and midbrain – receives input from the body's senses, *excluding* smell. When the parts of the RAS are active, nerve impulses pass upward to widespread areas of the cerebral cortex, both directly and via the thalamus, effecting a generalised increase in cortical activity associated with *waking or consciousness.* Image courtesy of Dr. Rosemary Boon, founder of Learning Discoveries Psychological Services

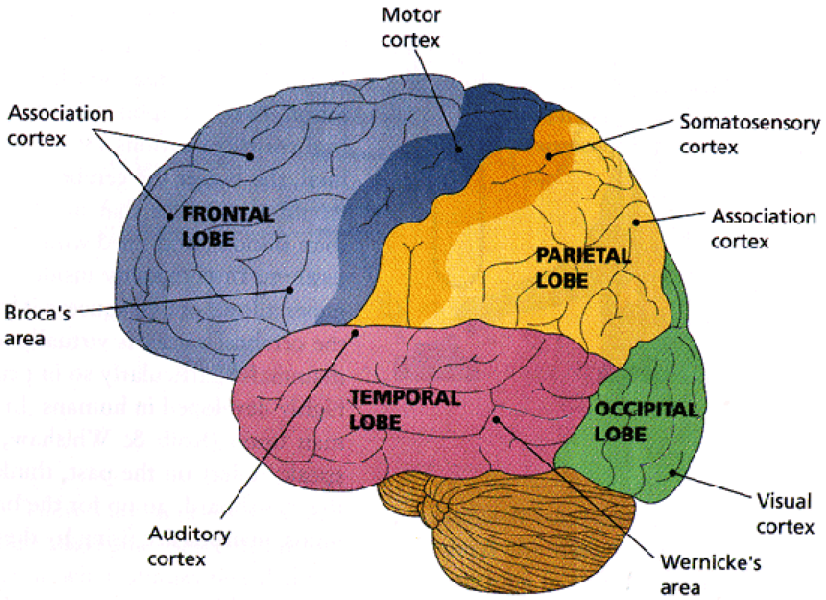


Figure 1. 2 Divisions of the cerebral cortex. Brain monitoring techniques indicate that in human beings, only processes that take place within the *associative* regions of the cerebral cortex are accompanied by primary consciousness (Roth, 2003, pp. 36, 38; Rose, 2002a, p. 15). Courtesy of Dr. Gleb Belov, Department of Mathematics, Technical University of Dresden, Germany

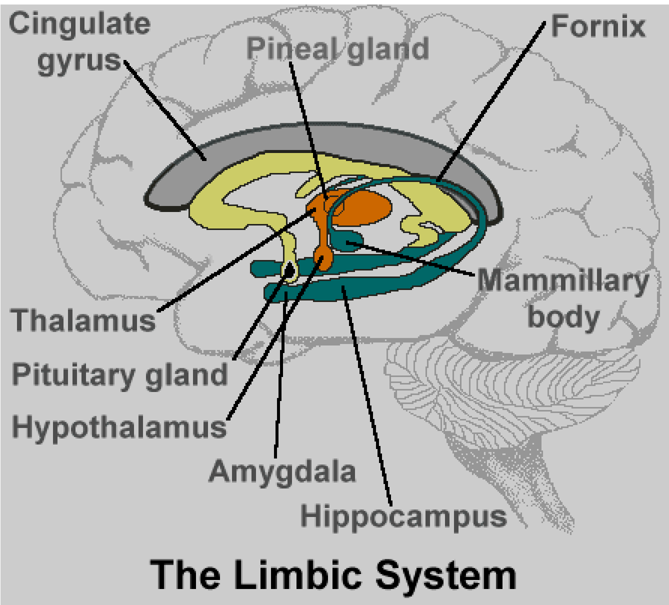


Figure 1.3 The limbic system. Some neurologists think it is part of the brain's mechanism for generating primary consciousness. Courtesy of Sandhills Community College, North Carolina.

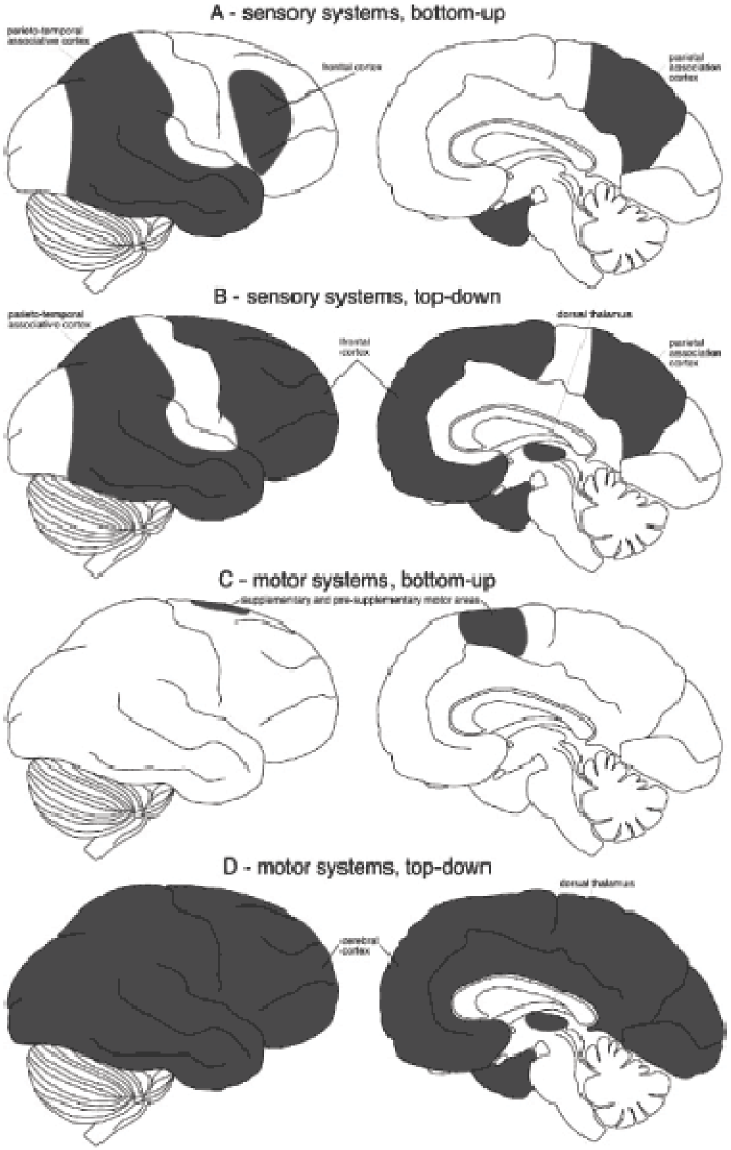


Figure 1.4 Four leading theories of how and where the brain generates primary consciousness. A side view of the human brain is shown on the left, and a cross-section through the brain is depicted on the right. In theory A, the neurons that actually generate consciousness are limited to the temporal, parietal and prefrontal regions of the neocortex. Theory B includes limbic system structures – septal regions, amygdala, and hippocampus – that are related to emotion and learning, as part of the generator mechanism. In theory C, consciousness is associated with activity in a small region of the neocortex: pre-supplementary or supplementary motor areas. Theory D includes several regions outside the neocortex, located in the limbic system, as part of the actual generator mechanism for consciousness. Courtesy of Butler, Manger, Lindahl and Arhem, 2005. Reproduced in blog by P.Z. Myers, 9 September 2005.

Unlike the data cited by Block (2005) on the visual cortex of monkeys, the three neurological identifying traits for primary consciousness listed by Seth, Baars and Edelman (2005) can be generalised over various sensory modalities and applied to a wide variety of animals. A strong case can be made for the occurrence of primary consciousness in *mammals* (Baars, 2001); whereas in lower vertebrates such as *fish*, the requisite brain structures are absent and there is no other region of the brain in which activity of comparable complexity can be said to occur (Rose, 2002a). *Birds* are only known to satisfy the first of the three neurological criteria for consciousness. More research needs to be done to ascertain whether the reentrant interactions between thalamus, cortex and basal ganglia occur in birds (Edelman, personal email, 19 July 2004). Most authorities are still disposed to deny that birds and reptiles possess primary consciousness, on the grounds that they lack a true neocortex, which is unique to mammals (Rose, 2002a, p. 10) and do not appear to have any brain structures possessing the special features of the association cortex - a high level of integration and a large number of connections with other regions of the brain. However, some authorities believe that birds' brains possess structures analogous to the neocortex (Seth, Baars and Edelman, 2005). Butler, Manger, Lindahl and Arhem (2005) also contend that “most of the critical structures presumed necessary for consciousness in mammalian brains have clear homologues in avian brains” (2005, p. 923), and hypothesise that the neuro-anatomical features of the forebrain common to mammals and birds may give rise to both *complex cognition* (evidenced by birds' ability to perform sophisticated tasks) and *consciousness.* They speculate that consciousness may have arisen in the ancestors of today's reptiles, mammals and birds. The only *invertebrates* that are thought to possibly possess the neurological wherewithal for consciousness are cephalopods, such as the octopus (Seth, Baars and Edelman, 2005). Despite their impressive cognitive feats, it is generally believed (David Edelman, personal email, 19 July 2004) that honeybees cannot be conscious, as their brains are too small (but see Koch, 2001, for a contrary view).

The fact that the three neurological criteria for primary consciousness are fairly well-defined allows us to hope that the Distribution Question (which animals are phenomenally conscious?) may turn out to be soluble. For philosophers with an interest in animal awareness, the crucial question concerns the relation between *primary* consciousness and *phenomenal* consciousness: in particular, is the former a sufficient condition for the occurrence of the latter?

1.5.5 Which varieties of consciousness should we distinguish?

I argued above in sections 1.5.2 to 1.5.4 that the varieties of consciousness distinguished by philosophers are highly problematic when applied to non-human animals. This raises the question: which varieties should we then distinguish? Based on my reading of the scientific literature, I would like to propose the following "natural categories" of consciousness:

- the different grades of *sensitivity* found in cellular organisms can be considered as levels of *sensory consciousness.* True senses, as we shall see in the next chapter, are confined to organisms with nervous systems;
- the three kinds of *memory* that can be found in organisms (procedural, semantic and episodic) are so distinct from one another that they could be regarded as manifestations of three underlying varieties of consciousness;
- *behavioural wakefulness*, as opposed to torpor – a distinction which is found in a wide variety of animals, including many insects, but not in bullfrogs, sea turtles, tortoises, alligators or sightless cave dwelling animals (Kavanau, 1997, p. 258);
- *brain wakefulness*, which is unique to mammals and birds (Kavanau, 1997);

- what I will call *integrative consciousness*, or the kind of consciousness which gives an animal simultaneous access to multiple sensory channels and enables it to integrate information from all of them. Mammals and birds possess this kind of consciousness, while reptiles lack it (Sjolander, 1993, p. 3; Dennett, 1995b, p. 691; Grandin, 1998);
- what I will call *object consciousness*, or awareness of Piagetian object constancy; ability to anticipate that an object which disappears behind an obstacle will subsequently re-appear. Reptiles lack this kind of awareness (Grandin, 1998), but at least some birds – doves, magpies, parrots and ravens – possess it (Butler, Manger, Lindahl and Arhem, 2005);
- what I will call *anticipatory consciousness*, or the ability to visually anticipate the trajectory of a moving object. Mammals can "lead" moving prey they are attacking by anticipating their trajectories - an ability that depends on their visual cortex (Kavanau, 1997, p. 255). Pigeons also possess this ability (Wasserman, 2002, p. 180). Although some fish and amphibians can snap at moving prey with their projectile tongues, "there is no evidence that fish and amphibians … attacking moving prey can 'lead' them by anticipating trajectories" (Kavanau, 1997, p. 255);
- *primary consciousness*, which appears to be confined to mammals and birds, and possibly cephalopods (Seth, Baars and Edelman, 2005);
- two distinct *kinds* of primary consciousness: an *affective* (and minimally cognitive) consciousness which is strongly associated with the limbic system, and a more fine-grained *cognitive* consciousness, associated with the neocortex; and
- *higher-order* consciousness, whose occurrence in non-human animals has yet to be established (see Nissani, 2004 for an overview of the evidence).

It is going to be a difficult matter for scientists and philosophers to decide which of these varieties of consciousness that occur in the animal kingdom deserve to be called "phenomenal" in the sense defined by Block (1995), although as we saw above, sensitivity and behavioural consciousness certainly do not. It has been argued that *primary consciousness* is inherently phenomenal; but the fact that several other varieties of consciousness described above *(brain wakefulness, integrative consciousness, object consciousness* and *anticipatory consciousness)* are restricted to the same kinds of animals in which primary consciousness is found (i.e. mammals and birds), constitutes *prima facie* evidence of a *nomic* connection between these varieties of consciousness, and the phenomenal consciousness beloved of philosophers.

1.6 Philosophical issues relating to intentionality

1.6.1 How should we define intentionality?

The philosophical notion of intentionality has been defined in various ways, by appealing to the notions of "aboutness" (or directedness towards an object), propositional content, and conditions of satisfaction (see Siewert, 2003, for a comprehensive discussion of the relevant literature; see also Searle, 1999; Jacob, 2003). While these definitions may turn out to complement one another, they are superficially quite different, and do not necessarily coincide with one another as regards the scope of the phenomena they include. I therefore propose to discuss the strengths and weaknesses of each definition in terms of the following five criteria, which I have selected for a good definition of "intentionality".

The first criterion we might expect a good definition of "intentionality" to satisfy is *fidelity* to the *historical usage* of the term, as the word itself is of medieval Scholastic origin and was revived in the nineteenth century by Brentano, who explicitly equated what Scholastic philosophers of the Middle Ages referred to as "the intentional inexistence of an object" in the mind with the property of “direction towards an object”, which he used to define intentionality (Jacob, 2003).

Clarity is another important criterion for evaluating rival definitions of intentionality: while some terms in common usage, such as “game” and “bald”, are inherently vague, philosophers are expected to define their own terms with a fair degree of rigour.

The fact that Brentano claimed to *define* mental phenomena in his work, *Psychology from an Empirical Standpoint*, as “those phenomena which contain an object intentionally within themselves” *(*1874 / 1911 / 1973, pp. 88-89) suggests a third criterion for judging the merits of competing definitions: their ability to pick out *all* mental states, and *only* mental states, as falling under the definition of intentionality. In other words, a good definition of intentionality should *define and unify the domain of the mental.*

A fourth criterion by which one might assess competing definitions of intentionality is their *biological scope:* are they specific to human beings, or can they also be applied to other creatures that are possible candidates for having mental states?

Finally, one might ask whether any of these definitions sheds light on the relation between *consciousness* and intentionality.

1.6.2 Common definitions of intentionality

Siewert (2003) discusses three ways in which philosophers have tried to elucidate the concept of intentionality. First, intentionality is most commonly defined in terms of “aboutness”: it is “that aspect of mental states or events that consists in their being *of* or *about* things” (Siewert, 2003).

For instance, one can ask, “What are you thinking about?” More precisely, intentionality is “the *aboutness* or *directedness* of mind (or states of mind) to things, objects, states of affairs, events” (Siewert, 2003). It is easy to show that beliefs and desires fall under this definition, as they have to be *about* something. I may believe that the food in front of me is delicious: I have a belief about the food, and a desire relating to it (a desire to eat it). The food is the *intentional object* of my belief and desire. The definition can also be applied to other mental states besides beliefs and desires: perceptions, emotions, memories and intentions are all *about* something, too. In Brentano’s words:

> In presentation, *something* is presented, in judgment *something* is affirmed or denied, in love loved, in hate hated, in desire desired and so on *(Psychology from an Empirical Standpoint,* 1874 / 1911 / 1973, pp. 88-89, italics mine).

A second way of explaining the concept of intentionality is by invoking the notion of *mental* (or intentional) *content*, which figures prominently in the analytical tradition of philosophy. Put simply, for a mental state to possess intentionality is for it to have content. Different *kinds* of intentional states may have the same content: Sophie may *believe* that her husband, Tom, is going bald, she may also *fear* that Tom is going bald, *hope* that he is going bald, or simply *see* that he is going bald. Sometimes, a “failure of substitutivity” may occur when a term (e.g. “Tom”) used to describe the content of an intentional state is replaced by another term which refers to the same thing (i.e. have the same *extension)*. Sophie’s husband may be the former president of the local bowling club, but if Sophie is unaware of this fact, then she will neither fear nor believe that the former president of the local bowling club is going bald. In this case, we can say that the two terms (“Tom” and “the former president of the local bowling club”) have a different *sense;* thus a sentence in which one term is substituted for the other will no longer have the same *content.* The content of an individual's intentional state can be expressed as a proposition *p*. In the above example, *p* is the proposition that *Tom is going bald.* Sophie may be said to be "acquainted" (as Russell (1905/1956) put it) with *p* insofar as *p* is not only what is understood when she understands the reference and truth conditions of the expressions she uses to report her state of mind, but it is also the content of her state of mind. Since so many of the states that we describe as "intentional" can be construed propositionally (Sophie may believe *that p*, fear *that p*, hope *that p*, see *that p*, and so on), it seems reasonable to define intentional states as "propositional attitudes".

Another account of intentionality, advocated by Searle (1999), equates intentional states with those states that have *conditions of satisfaction,* or at least presuppose other states that do. Beliefs are either true or false – they are beholden to facts in the real world, and if the world differs from the way it is described in someone's belief, then the subject must amend her belief accordingly. We can thus speak of them as having a "mind-to-world" *direction of fit.* Desires and intentions, on the other hand, have a world-to-mind direction of fit: "their function is not to represent how things are but how we would like them to be or how we plan to make them be" (Searle, 1999, p. 102). "Beliefs can be true or false, desires can be fulfilled or frustrated, intentions can be carried out or not carried out", but the common feature they all share is that in each case, "the intentional state is *satisfied* or not depending on whether there is indeed a match between the propositional content and the reality represented" (Searle, 1999, p. 103, italics mine). Thus beliefs, desires and intentions can all be said to possess conditions of satisfaction, because they all have a direction of fit. According to Searle, it is a general feature of intentional states with a propositional content that they have conditions of satisfaction. However, there are some states with a propositional content which lack a direction of fit. For instance, my being proud that I won a race has no direction of fit because the fitting has already taken place – the propositional content of the intentional state ("I won the race") is already satisfied. But even states like this with a *null direction of fit* have conditions of satisfaction: "if I am proud that I won the race, then I must at least (a) believe that I won the race and (b) find it desirable or want it to be the case that I won the race" (Searle, 1999, p. 104). Finally, there are intentional states, such as love and hate, which have an *object*, but *no* propositional content. However, Searle argues that even these states are to some degree constituted by accompanying beliefs and desires, which do have conditions of satisfaction:

> So, one cannot love a person, for example, without having a set of beliefs and desires regarding that person. And those beliefs and desires are, in large part, constitutive of the love that one has for that person. Thus, although superficially love does not have conditions of satisfaction, any actual case of one human loving another is constituted in large part by a set of intentional states that do have conditions of satisfaction (Searle, 1999, p. 104).

Normativity is thus a central feature of Searle's account of intentionality: beliefs are either correct or mistaken, desires are either satisfied or not, and even intentional states that lack a direction of fit presuppose beliefs and desires which have conditions of satisfaction.

1.6.3 An evaluation of philosophical accounts of intentionality

1.6.3.1 Fidelity to historical usage

Defining "intentionality" in terms of "directedness" or "aboutness" is certainly in keeping with the historical roots of the term: the word “intentionality” derives from the Latin word *intentio*, which in turn derives from the verb *intendere*, which means being directed towards some goal or thing, as an arrow is towards a target (Jacob, 2003). However, the notion of mental *content* also figures in Brentano’s celebrated definition:

> Every mental phenomenon is characterized by what the Scholastics of the Middle Ages called the intentional (or mental) inexistence of an object, and what we might call, though wholly unambiguously, reference to a *content*, direction toward an object…, or immanent objectivity *(*1874 / 1911 / 1973, pp. 88-89, italics mine).

Searle’s notion of “conditions of satisfaction”, on the other hand, is a philosophically novel way of construing the meaning of intentionality. Although Searle, in his exposition of intentionality, starts with the traditional concept of intentionality as “that feature of mental states whereby they are directed at or about objects and states of affairs other than themselves” (1999, p. 99), and goes on to claim that typically, intentional states have a propositional content, he nevertheless maintains that “the key to understanding intentionality is conditions of satisfaction” (1999, p. 103). According to Searle, the more traditional accounts of intentionality fail to capture its essentially normative character.

Searle’s innovative exposition of intentionality may not fare well when judged by purely historical criteria; however, it may prove to be more successful than other accounts when judged by the other criteria listed above.

1.6.3.2 Clarity

One major drawback of defining intentionality in terms of "aboutness" or directedness towards an object is its lack of clarity. Siewert (2003) rhetorically asks:

> But what kind of "aboutness" or "of-ness" or "directedness" is this, and to what sorts of things does it apply? How do the relevant "intentionality-marking" senses of these words ("about", "of", "directed") differ from: the sense in which the cat is wandering "about" the room; the sense in which someone is a person "of" high integrity; the sense in which the river's course is "directed" towards the fields? (Siewert, 2003, online).

In particular, the occurrence of a kind of "aboutness" even in *simple systems* suggests that it is too vague a concept for defining intentionality:

> [T]hings we're not prepared to credit with thought - for example, heat-seeking missiles and sunflowers - also exhibit directedness towards objects. The challenge then is to find a way to distinguish the special sort of directedness possessed by *bona fide* thinkers from the more primitive kinds exhibited by these simpler systems (Beisecker, 1999, p. 282).

I will discuss the "aboutness" of artifacts at further length below. The point I wish to make here is that even if it turns out that there are some philosophically significant differences between the "aboutness" of artifacts and that of thoughts, we shall need to draw upon *additional* concepts in order to explain what those differences are.

The notion of *propositional content* is arguably less opaque than that of “aboutness”. However, it should be noted that some philosophers have argued that the concept of “content” is in need of further refinement, and have claimed to distinguish different senses of content (see Siewert, 2003, for a discussion). It has also been suggested that whereas thoughts and prepositional attitudes do indeed have a propositional content, some intentional states (e.g. *qualia*, such as pains) may have a more primitive kind of content, which does not require propositional thought or even concepts (see Jacob, 2003, for a discussion).

Finally, the term “conditions of satisfaction” appears to be a fairly straightforward one, which also ties in well with the notion of propositional content:

> If states of mind contrast in respect of their satisfaction (say, one is true and the other is false), they differ in content… And if one says what the intentional *content* of a state of mind is, one says much or perhaps all of what *conditions must be met* if it is to be satisfied – what its conditions of truth, or veridicality, or fulfillment, are (Siewert, 2003, online).

1.6.3.3 Ability to define and unify the domain of the mental

It was suggested above that the criterion of “aboutness” fails to pick out *only* mental states. Some artifacts exhibit a kind of “aboutness”: Dennett (1995a) is famous for claiming that since the shape of the bimetallic spring in a thermostat embodies information about room temperature, a thermostat qualifies as a minimal belief-holder. Dennett (1995a) construes "beliefs" in a "maximally permissive" sense as "information-structures" that are "sufficient to permit the sort of intelligent choice of behavior that is well-predicted from the intentional stance".

The case of the thermostat is but one of many natural devices which satisfy Fred Dretske's (1980, 1981) *information-theoretic* proposal for naturalising intentionality:

> In essence, the information-theoretic proposal is that device S carries information about instantiations of property G if and only if S's being F is nomically correlated with instantiations of G. If S would not be F unless property G were instantiated, then S's being F carries information about, or as Dretske likes to say, indicates G-ness. A fingerprint carries information about the identity of the human being whose finger was imprinted. Spots on a human face carry information about the disease... A compass carries information about the location of the North pole. In all cases, a property of a physical device nomically covaries with some physical property instantiated in its environment (Jacob, 2003, online).

One significant disanalogy between the "aboutness" of beliefs and that of physical devices that carry information is that the former can *misrepresent*, in a way that the latter cannot (Jacob, 2003). Beliefs can be false. Of course, compasses can fail to point north. But as Dretske (1988) has argued, this can be considered as a misrepresentation only in a derivative sense. So long as the compass is free from magnetic interference, the needle will infallibly point north, regardless of the purposes and intentions of the designers and users. If interference occurs due to a local magnetic anomaly or a nearby TV set, a compass can "misrepresent" the location of the north magnetic pole, but we refer to this as a misrepresentation only because the compass's designated *function* is to indicate the direction of the pole, rather than the direction of the local anomaly or the TV set. The compass only has this function because of the purposes and attitudes of the designers.

Another major difference between beliefs and information-carrying devices is that beliefs are sensitive to the *meaning* of the terms they relate in a way in which devices are not. In philosophical jargon, beliefs exhibit a stronger kind of *intensionality.* It is a law of nature that if there is a current flow between two points, then there is a voltage difference between the two. Thus the position of a needle on a galvanometer carries information not only about the flow of current but also about voltage differences between two points. One can, however, believe that there is a current flow between two points, without believing that there is a voltage difference between the two (Jacob, 2003).

A further peculiar feature of thoughts and other mental states, noted by medieval philosophers, is that they can be about objects that *do not exist:* I may believe that the food in front of me is delicious, but if the food in question is really plastic that has been molded, painted and sprayed with volatile chemicals, in order to make it look and smell like delicious food, then my belief is about a non-existent object. The information-carrying devices described above are incapable of instantiating this property.

Confronted with these dissimilarities, one has two choices. The more conservative choice is to continue to define intentionality in terms of “aboutness”, while distinguishing the strong “aboutness” of beliefs from the much weaker “aboutness” of information-bearing devices. Perhaps weak intentionality, combined with some extra conditions, will give us the strong intentionality that distinguishes mental states. For instance, one widely adopted proposal for dealing with the difficulties left by the information-theoretic approach is Millikan’s (1984, 1993,

2000) *teleosemantic* approach, which states that having a *biological function* that arises from some historical process of natural selection is a necessary condition for “aboutness” as Brentano defined it. However, Millikan’s proposal illustrates some of the difficulties inherent in the search for sufficient conditions for “aboutness”. First, extra conditions may still not constitute *sufficient* conditions:

> …[W]hereas biological organs have functions that may fail to be fulfilled, they do not *ipso facto* exhibit intentionality in Brentano’s full sense: neither a heart nor a stomach are *of* or *about* anything. Millikan’s claim, however, is not that having a function is *sufficient* for aboutness, but that it is *necessary* (Jacob, 2003, online, italics mine).

Second, the terms used to generate sufficient conditions for “aboutness” may turn out to be more amenable to *other accounts* of intentionality. For instance, one merit of Millikan’s proposal is that it explains misrepresentation: for a device to misrepresent what it is about is simply for it to malfunction. However, the notion of a malfunction is an essentially *normative* one, which seems better suited to Searle’s account of intentionality.

Finally, the search for sufficient conditions for “aboutness” (in the strong sense of the term) begs the question of how we can be sure that this property is indeed the hallmark of mental states.

The alternative and more radical choice is to abandon the project of defining intentionality in terms of “aboutness”, and look for another way of distinguishing beliefs and other mental states from thermostats and compasses. Definitions of “intentionality” which invoke *propositional content* and *conditions of satisfaction* do not appear vulnerable to the difficulties raised above:

we do not ordinarily speak of artifacts as having propositional content, or conditions of satisfaction. While we do speak of databases as storing and manipulating data which can be used to generate propositions (e.g. Ted Jones lives at 46 Bay Street), that is no reason to impute *propositional attitudes* to databases as such; rather, it is the users of the database who extract meaning from the data, which they express in the form of propositions in their language. As these propositions may be true or false, they also have *conditions of satisfaction.*

We have seen that "aboutness" fails to pick out *only* mental states. A further failing of this term is that it fails to pick out *all* mental states: for instance, feelings of elation, anxiety or depression don't seem to be "about" anything. On the other hand, the other definitions of "intentionality" proposed above appear equally incapable of accounting for these recalcitrant cases.

Although the property of "intentionality" applies to most but not all mental phenomena, one could still plausibly claim that it encompasses all *kinds* (types) of mental states, even if it fails to cover all *instances* (tokens) of these states. For instance, we may feel depressed for no particular reason, but typically we are depressed *about* something.

One advantage of defining intentionality in terms of *propositional content*, as Russell does, instead of "aboutness", is that this definition applies only to mental states. However, one major drawback of this account is that it is unable to encompass a wide variety of intentional states, such as love and hate, which have an object but lack any propositional content. Thus Russell's construal of intentional states as *propositional attitudes* can only serve as a *partial* account of intentionality, at best.

As noted in section 1.6.3.2, some philosophers have suggested that pains and other sensory experiences possess a *primitive* kind of intentionality that does not presuppose the ability to grasp propositions, or even concepts. If they are right, this *sensory* intentionality would have a special *non-conceptual* kind of content, and Russell's propositional account of intentionality would not apply to the "lower echelon" of mental states either. If the notion of a "non-conceptual content" makes sense, it would certainly resolve the problem of how pre-linguistic infants and non-human animals could be said to have intentional mental states. On the other hand, other philosophers (Anscombe, 1965; McDowell, 1994) would argue that the notion of a non-conceptual content is simply unintelligible. Siewert (2003) remarks that the main difficulty in evaluating the idea of a "non-conceptual content" is that philosophers still lack a clear idea of what "having a concept" means. I shall revisit the intentionality of sensory experiences below. My own view is that while the idea of a "non-conceptual content" is dubiously coherent, there is no good reason to construe the content of sensory experiences in *propositional* terms, as Russell's account requires us to. When a dog sees a tree, there is no reason to suppose that it sees *that* there is a tree.

Searle's account manages to circumvent the difficulty posed by intentional states (such as love and hate) that lack *propositional content*, by showing that even these states have *conditions of satisfaction.* But while this account explains most or all of the phenomena we call mental states – with the possible exception of odd cases like feelings of depression, which we mentioned above – it does so at the expense of marginalising many of these states. Love and hate are surely paradigmatic intentional states, but in Searle's account they are considered to be intentional only because these states presuppose the occurrence of beliefs and desires regarding the object of one's love or hate. Indeed, by arguing that these beliefs and desires are "in large part, constitutive of the love one has for that person" (1999, p. 104), Searle is effectively assimilating love to two other intentional states which are in reality quite distinct from it. The same goes for the large number of intentional states – such as being glad or sorry, ashamed or proud – which have propositional content but lack a direction of fit. Why should these states be regarded as intentional only because of the beliefs and desires that accompany them?

At this point, it seems that Searle has no choice but to reject our intuition that intentional states such as love and pride are just as fundamental as beliefs and desires. He could respond that these states *seem* to be equally fundamental because they are obviously "about" something, but then argue that "aboutness", taken by itself, fails to clarify the notion of intentionality, for reasons we discussed above. Thus if Searle is right, we should ignore intuitions about intentionality based purely on "aboutness".

Searle's account also needs to handle intentionality at the *sensory* level. Searle claims (1999, pp. 96-97) that conscious visual experiences possess intentionality in their own right – or *intrinsic intentionality,* as he calls it. David Thompson (1985, online) has criticised Searle (1983) on this point, accusing him of equivocating between visual *experiences* (which lack intentional content and do not refer to anything) and visual *perceptions* (which possess an intentional content):

> Visual experiences are only empirically related to Intentional content. We could imagine a non-human creature whose perception of a tree, while having the same content, is realized in a radically different psychological structure… As psychological events, visual experiences are neutral with respect to reference… A visual experience is not itself, strictly, *of* anything, it just is. Only when it takes on logical properties, only when it

> acquires an Intentional content can it be *about* an object. Perception, and other intentional states logically determine conditions of satisfaction and thereby refer to an object. Visual experiences do so only secondarily, empirically, contingently, in so far as they are related to Intentional content by what Searle calls “realization” (Thompson, 1985, online).

Thompson goes on to argue that Searle fails to explain how “realization” can occur. Searle (1999, pp. 28-31) undercuts this objection by questioning the alleged dichotomy between *experiences* and *perceptions* of the real world, and by defending naïve perceptual realism (the view that when I perceive a tree, it is the tree that I perceive, as an independently existing material object) against two well-known arguments – the argument from science and the argument from illusion – that claim we need to posit intermediary “sense data” or “percepts” to explain perception. Having demolished these arguments, Searle turns to a simple case – the visual experience of a computer screen – and argues:

> There is no way that I could have *this* visual experience that I am having without it at least seeming to me that there is a computer screen in front of me (1999, pp. 96-97).

Thompson’s argument that other creatures could experience the same thing in a radically different manner is beside the point here: even if true, all it shows is that there are many different kinds of experiences that can refer to the same thing.

I conclude that bodily sensations are not counterexamples to the claim that intentionality is necessary for mentality, and that Searle’s account can accommodate these examples satisfactorily. However, I see no reason why the content of these sensations needs to be construed propositionally. Instead of saying, “It seems to be *that* there is a computer screen in front of me”, one could equally well say, “I am looking at *what* appears to be a computer screen” (here, the content is construed as an object). I would also suggest that the latter approach provides us with a more straightforward way of attributing intentionality to animals’ bodily sensations.

We saw above that for Searle, intentional states (such as love and hate) which lack propositional content have *conditions of satisfaction*, but only because these states presuppose the occurrence of beliefs and desires regarding their object. Thus even if we choose to construe bodily sensations as intentional states that simply have an *object* rather than propositional content, we can only provide them with Searle’s conditions of satisfaction by reference to their accompanying beliefs and desires. There seems to be no room, on Searle’s account, for a more primitive kind of intentionality that does not require the capacity for holding beliefs.

This entailment of Searle’s account may strike some as counter-intuitive. Beliefs presuppose a certain level of cognitive sophistication; indeed, some philosophers (e.g. Davidson, 1975, 1999) have argued that creatures lacking language are simply incapable of holding beliefs. If we accept Searle’s account, it follows that the sensory experiences of many non-human creatures are not intentional states after all, as they lack the cognitive wherewithal for having beliefs. This raises the question: where do we draw the line? Searle himself (as we shall see below) maintains that a capacity for *consciousness* is the critical requirement, and since he makes it clear that he regards the “inner, qualitative and subjective nature” of consciousness as “essential features” (1999, p. 41), one can only assume that he is referring to *phenomenal* consciousness. Those wishing to attribute intentional states to creatures that are capable of having sensory experiences but incapable of phenomenal consciousness therefore have two choices: they must either explain how internal states which are not even *potentially* conscious can still possess conditions of satisfaction, or reject Searle's demand for conditions of satisfaction and construct an *alternative account* of intentionality, which is broad enough to include these creatures' sensory experiences as mental states while at the same time excluding non-biological states (such as compass and thermostat readings) which, as I argued in section 1.1, are mindless. Beisecker's (2000) account of intentionality, which I shall discuss later in this thesis, is a thought-provoking example of the former approach. By contrast, the latter approach does not look at all promising: the only account of intentionality that has been proposed to date which is broad enough to encompass non-conscious sensory experiences is Brentano's original account, which is based on "aboutness", but as we saw above, it seems to suffer from the defect of being too broad: artifacts can be said to possess a kind of "aboutness" too.

We have seen how Searle's account of intentionality can accommodate most if not all kinds of mental states, including sensory experiences. It remains to enquire whether Searle's account applies *only* to mental states, or whether it could apply to other states as well. Failure to exclude states and events which are clearly not mental would render Searle's account of intentionality incapable of unifying and defining the domain of the mental.

Searle himself considers *consciousness*, not intentionality, to be the true hallmark of the mental. Many of our brain processes are "nonconscious" and hence "not mental at all" (1999, p. 86), while others are *bona fide* mental states. The latter may be either conscious or "unconscious" – by which Searle means "the sort of state that could be conscious" (1999, p. 86).Thus consciousness (or at the very least, availability to consciousness) is the vital feature of these brain processes which "makes them *mental,* makes them have the features of intentional mental states" (1999, p. 88). Searle adds that "we can only understand intentionality in terms of consciousness" (1999, p. 65). However, it makes little difference for our purposes whether Searle regards consciousness or intentionality as the distinguishing feature of mental states: what matters here is whether he can provide us with a unifying account that explains why mental states alone possess the attribute of intentionality.

I have two main criticisms of Searle's attempt at unifying the domain of the mental via the notion of consciousness. First, although he does a good job of discrediting rival accounts of how intentionality might arise, he fails to make a convincing positive case that (potential or actual) consciousness is what gives brain processes their intentionality. Second, in making the assertion that intentionality presupposes consciousness, he appears to rely on *traditional* definitions of intentionality; his own notion of "conditions of satisfaction" does little if any work here.

Searle's case appears to boil down to the following two claims: only brain processes which are *conscious* (or at least, potentially conscious) can *represent* objects or states of affairs in the world (1999, pp. 64-65), and only processes which are at least potentially conscious can count as following the *intentional content* of a rule (1999, p. 89).

The following quote illustrates Searle's first claim:

> [T]ypically a conscious state such as an intention or a desire functions by representing the sort of event it causes. For example, I want to drink water, so I drink water. Here,

> the effect, drinking water, is consciously represented by the cause, the desire to drink water… At this point, I just want to remark on the amazing property that conscious beings have to represent objects and states of affairs in the world and *to act on the basis of those representations.* It is a general feature of most, though not all, conscious phenomena that they represent objects, events, and states of affairs in the world. Indeed, the most important feature of consciousness is that there is an essential connection between consciousness and the capacity that we human beings have to represent objects and states of affairs in the world to ourselves….[T]hough not all conscious states are intentional, and not all intentional states are conscious, there is an essential connection: we can only understand intentionality in terms of consciousness (1999, pp. 64-65).

There is an apparent leap in logic here. Granted that most conscious states *represent* something, it does not follow that our human capacity to "represent objects and states of affairs in the world to ourselves" presupposes consciousness. It is also noteworthy that in making this claim, Searle invokes the traditional concept of a "representation", without mentioning his own notion of "conditions of satisfaction" (which he defines in a later chapter).

Later, however, Searle uses the illustration of his belief [in 1999] that Clinton is the President of the United States, to develop his case that *availability to consciousness* is required to explain the intentionality of belief:

> [I]f a state is a genuine unconscious *mental* state, then it must be at least the sort of state that could be conscious. We need, therefore, to distinguish *nonconscious* states

> of the brain, such as the secretion of the neurotransmitter norepinephrine into the synaptic cleft, from *unconscious* mental states that are realized in the brain – such as my belief, when asleep, that Clinton is president. Now, since when I am totally unconscious the only occurrent reality of the brain is nonconscious, what fact about those nonconscious mental states makes some of them into mental states? The only answer is that certain nonconscious states of the brain are capable of causing mental phenomena (1999, pp. 86-87).

There is no positive argument here: Searle is simply *asserting* that only the potential for consciousness can account for the intentionality of some of our nonconscious brain states – namely, those which correspond to our unconscious beliefs. Later (1999, pp. 90-91), he goes on to demolish rival explanations of intentionality which attempt to reduce it to "dumbed-down" forms at lower levels in the brain (Dennett's "army of progressively stupider homunculi") or to some kind of causation (Fodor). But discrediting these positions does not make Searle's own position correct.

Finally, Searle claims that although we can follow a rule without consciously thinking about it, "unconscious rule-following, like conscious rule-following, must be a matter of following the *intentional content* of the rule" (1999, p. 89, italics mine) – which is why it makes no sense to suppose that there are rules that could not become conscious even in principle. This is a more substantial argument. Once again, it is curious that in order to make his argument that intentionality presupposes consciousness, Searle chooses to invoke the concepts of "intentional content", rather than his own notion of "conditions of satisfaction". We can however bridge the gap between the two notions by defining the *satisfaction* of an intentional state according to “whether there is indeed a match between the *propositional content* and the reality represented”, as Searle does (1999, p. 103, italics mine). It then follows that non-mental phenomena are ineligible for having conditions of satisfaction. However, in order to complete his case that all mental phenomena are at least potentially conscious, Searle would have to also show that mental states can only be ascribed to creatures that are capable of following *rules.*

We have seen that Searle’s conception of “conditions of satisfaction” is tied to that of intentional content. However, if we focus exclusively on the notion of a *condition of satisfaction*, *per se,* we can imagine it being satisfied by states that are not even potentially conscious. For instance, all human beings – including those who are in a vegetative state – have various instinctual *drives.* Since each drive is towards some goal, the drive could be said to be *satisfied* by the attainment of the goal, regardless of whether this attainment is “available to consciousness” as Searle stipulates.

To sum up: none of the three accounts of intentionality examined above appears capable of defining and unifying the domain of the mental, on its own. Russell’s account is too narrow as it excludes states (such as love) that have no propositional content, while an account that is based on “aboutness” seems to be too vague, as well as too broad. Searle’s account has the virtue of encompassing the vast majority of mental states, while excluding states that are not mental. However, it implicitly relies on the other two accounts in order to rule out states that are not even potentially conscious. This definitional inter-dependence need not be a bad thing; what it suggests is that a *combination* of the strengths of the above three accounts offers the best hope for our being able to define and unify the domain of mental phenomena, using the notion of intentionality.

Interestingly, Searle himself considers *consciousness*, not intentionality, to be the true hallmark of the mental, as he contends that consciousness is what makes it possible for processes to be "about" anything, and that intentionality therefore has to be understood in terms of consciousness (1999, pp. 64-65). I shall revisit this claim in section 1.6.3.5, where I discuss the relation between consciousness and intentionality.

One final feature of intentional states that emerges from Searle's account of intentionality is that some of these states – namely, *beliefs, desires* and *intentions* – occupy a privileged position in any exposition of intentionality. We can only attribute intentionality to mental states (e.g. love and hate) which have an object but no content, insofar as these states presuppose the occurrence of beliefs and desires regarding the object.

Thus Searle's account seems to imply that even a *minimal mind* has to be capable of having beliefs and desires of some sort.

1.6.3.4 Biological scope

"Aboutness" is certainly a concept which can be applied to creatures whose biology is different from our own, so it meets our fourth criterion for a satisfactory definition of "intentionality". On the other hand, the property of "aboutness" also appears to be instantiated by many biological systems that we might consider mindless – such as magnetosome bacteria (to be discussed in a later chapter), which align themselves with the earth's magnetic field – as well as by *non-biological* systems (thermostats), which cannot possess mental states if the arguments in section 1.1 are valid.

The ascription of *propositional content* to other creatures' mental states is much more problematic. Some philosophers argue that the ascription of any mental state with a *propositional content* (such as a belief) to a non-human animal is absurd, either because (i) the object of a belief is always *that* some sentence S is true, and lacking language, an animal cannot believe that any sentence is true (Frey, 1980), or because (ii) nothing in an animal's behaviour allows us to specify the content of its belief and determine the boundaries of its concepts (Stich, 1979, refers to this as the "dilemma of animal belief", p. 26), or because (iii) none of our human concepts can adequately express the content of an animal's belief, given its lack of appropriate linguistic behaviour that would confirm that our ascription was correct (Davidson, 1975).

An example from Dennett (1997, p. 56) illustrates this point. What does a dog think, just as it is about to eat? Does it think the thought *that* "My dish is full of beef", or the thought *that* "My plate is full of calves' liver", or even the thought *that* "The red, tasty stuff in the thing that I usually eat from is not the usual dry stuff they feed me"?

The common assumption underlying the above objections is that the content of a thought must be expressible by a *that*-clause, in some human language. Carruthers (2004a) rejects this assumption on the grounds that it amounts to a *co-thinking constraint* on genuine thoughthood: "In order for another creature (whether human or animal) to be thinking a thought, it would have to be the case that someone else should also be capable of entertaining that very thought, in such a way that it can be formulated into a that-clause." This is a dubious proposition at best: as Carruthers points out, some of Einstein's more obscure thoughts may have been thinkable only by him.

A more reasonable position, urges Carruthers, is that an individual's thoughts can be characterised equally well *from the outside* (by an indirect description) as from the inside (by a that-clause which allows me to think what the individual is thinking):

> In the case of an ape dipping for termites, for example, most of us would ... say something like this: I don't know how much the ape knows about termites, nor how exactly she conceptualizes them, but I do know that she believes *of* the termites in that mound that they are there, and I know she wants to eat them (Carruthers, 2004a, p. 3, online PDF version).

Dennett makes a similar point:

> The idea that a dog's "thought" might be inexpressible (in human language) for the simple reason that expression in a human language *cuts too fine* is often ignored, along with its corollary: the idea that we may nevertheless exhaustively describe what we can't express, leaving no mysterious residue at all (1997, p. 56).

The point I wish to make here is not that animals are capable of having beliefs, but that the arguments against animal belief we examined so far contain questionable philosophical assumptions. I conclude that (a) there are no convincing grounds for restricting mental states to creatures that are capable of having states with a clearly specifiable propositional content, and (b) the Russellian account of belief is too narrow to apply to other creatures.

Even if the requirement that we should be able to specify the content of animal beliefs is an unreasonable one, it can still be argued that beliefs presuppose the possession of *concepts.* There are two principal arguments against the possibility of attributing concepts to other animals. First, Davidson (1975, 1984, 1999) makes the strong claim that any individual holding a belief needs to possess a large number of background concepts, including the *concept of a belief.* Davidson refers to this claim as the “holism of the mental”. The following example illustrates Davidson’s line of argument:

> [T]o have a belief about a cat, one must have mastery of the concepts that are involved in this judgment or belief. A creature does not have the concept of a cat merely because it can discriminate cats from other things in its environment… To have the concept of a cat, you must have the concept of an animal or at least of a continuing physical object, the concept of an object that moves in certain ways, something that can move freely in its environment, something that has sensations. There is no fixed list of things you have to know about, or associate with, being a cat; but unless you have a lot of beliefs about what a cat is, you don’t have the concept of a cat (1999, p. 8).

Elsewhere Davidson addresses the question of whether a creature is capable of having any beliefs at all if she lacks the *concept of a belief* altogether. In his article, “Thought and Talk”, he writes: “It seems to me that it cannot” (1975; 1984, p. 170). Stephan draws attention to a counterintuitive entailment of Davidson’s claim:

> It seems clear that neither animals, nor infants, nor demented adults would belong to the set of systems that can have beliefs and concepts in Davidson’s sense (1999, p. 81).

Even if Davidson's claim is too strong, however, one can formulate a more modest argument against animal concepts, based on the *negative* claim that merely being able to *discriminate* between Xs and non-Xs is not a sufficient condition for having the *concept* of an X. Since the only way we can investigate the possession of concepts by other animals is to look at their discriminative capacities, it seems that we can never know for sure if animals have concepts – from which it follows that we can never know if they have beliefs. If one accepts that beliefs are part-and-parcel of even a minimal mind, then it seems we can never be sure that other animals have minds.

Allen (1999) has attempted to meet this objection by specifying a more robust set of conditions for having a concept. According to Allen (1999, p. 37), it is justifiable to attribute the having of the concept X to a system S if it can (i) systematically discriminate between Xs and non-Xs, (ii) recognise its own discrimination errors, and (iii) thereby learn to better discriminate between Xs and non-Xs. Stephan (1999) argues that if an animal possesses the additional ability to distinguish between *true* and *false calls* signalling the presence of an X (e.g. a leopard), we are surely warranted in describing the animal's behaviour in intentional terms.

What is lacking from the foregoing philosophical discussion is a recognition of the variety of usages of the word "concept" in studies of animal cognition. As Ristau (1991, p. 82) points out, there are many different kinds of concepts that animals are alleged to possess: abstract concepts, perceptual concepts, and natural concepts, to name but a few. Scientists who study animal cognition do not use the same kinds of procedures to investigate all these concepts; they use various kinds of criteria used to determine whether animals possess the abstract concepts of "same-different", or of a temporal interval, or a number; and different methods again when investigating animals' possession of perceptual concepts (e.g. red; table; Van Gogh painting). The question of whether and what kinds of concepts animals have is not likely to have a simple answer. Non-human animals may be capable of having some kinds of concept but not others, and we should expect to find various levels of cognitive sophistication in the animal kingdom.

That leaves us with Searle's account. Given that desires have conditions of satisfaction on Searle's account, it might seem that the identification of intentional states in animals and other organisms would be a straightforward affair: we could look for goal-driven behaviour in animals, and identify its conditions of satisfaction with the attainment of the goal. Bacterial *phototaxis* is one example of a non-conscious process that *seems* to have a condition of satisfaction: finding a source of light. The same could be said for non-conscious drives in animals: since these drives are towards some goal, they could be said to be "satisfied" by its attainment. However, as we saw above, such an approach would be a complete misreading of Searle's views: according to Searle, "we can only understand intentionality in terms of consciousness" (1999, p. 65). Searle holds that only processes which are *conscious* (or at least, potentially conscious) can represent objects or states of affairs in the world (1999, pp. 64-65). Any goal-driven animal behaviour that is *not* generated by a conscious mental state within the animal cannot possess intentionality in its own right (intrinsic intentionality), but merely exhibits what Searle describes as an "as-if" or metaphorical intentionality – the kind we use when we say, "The plants in my garden are hungry for nutrients" (1999, p. 93). Thus if we define conditions of satisfaction as Searle intends, then we cannot impute "conditions of satisfaction" to processes or behaviour occurring in other animals without first ascertaining whether these animals are capable of phenomenal consciousness.

On the other hand, Searle himself seems to take a broad view regarding the extent of consciousness in the animal kingdom: “in the real world, the way humans and higher animals typically cope is by way of conscious activities” (1999, p. 63).

1.6.3.5 Ability to explain the relation between consciousness and intentionality

It is by no means clear that “aboutness” sheds any light on the relation between *consciousness* and intentionality. As we saw above, physical devices which are devoid of consciousness certainly instantiate a weak form of “aboutness”. Even the stronger version of “aboutness” may not presuppose phenomenal consciousness: we do, after all, speak of an individual’s having unconscious beliefs.

Russell’s propositional account does not appear to shed any light on the question either. Even if it could be shown that the ability to entertain propositional attitudes with a specifiable content was required for intentionality, it would not necessarily follow that phenomenal consciousness was required.

According to Searle’s account, consciousness is required to make mental states refer in their own right: that is, *intrinsic intentionality* is possible only because of consciousness. Yet as we saw above, Searle’s positive arguments fail to establish his claim. While Searle’s criticisms of popular “reductive” accounts of intentionality are far more telling, they fail to preclude the possibility that philosophers may be able to formulate a non-reductionist account of intentionality that does not require consciousness (see for instance Beisecker, 1999).

Siewert (1998, 2004) suggests that the philosophical debate over whether the phenomenal features of mental states explain the intentional ones, or *vice versa*, focuses on the wrong questions. He argues that phenomenal consciousness is essentially intentional, without being explained by that fact. Siewert contends that many of the phenomenal features of consciousness are assessable for accuracy. At the same time, he doubts whether phenomenal consciousness can be built up from non-phenomenal intentionality.

My own comment here is that the foregoing discussion of phenomenal consciousness suggests that it is best understood as an emergent *neurological* property. I see no particular reason to suppose that it will be reducible to a higher-level property such as intentionality.

I conclude that no single definition of intentionality can do justice to the term, and that we need to harness the strengths of all three existing approaches. Searle has done philosophy in drawing attention to the normative character of intentionality, but he fails to demonstrate his claim that intentionality presupposes phenomenal consciousness. At this stage of our investigation, I suggest that we keep an open mind regarding the relation between the two.

1.6.4 Dennett's intentional stance

In his influential book, "Kinds of Minds" (1997) and elsewhere, Dennett defines an intentional system as an entity whose behaviour can be more easily predicted from an *intentional stance*, where the entities are treated as if they were agents who choose to behave in a certain way, because of their underlying *beliefs* about their environment, and their *desires* – or (in what Dennett regards as an alternative description), because of their *information states* that enable them to achieve their *goals.* Insofar as intentional systems are said to have beliefs and desires, they exhibit the philosophical property of *aboutness:* beliefs and desires have to be about something. Dennett suggests that we can usefully regard living things and their components from an intentional stance, because their behaviour is "produced by *information*-modulated, *goal*-seeking systems" (1997, p. 34).

The chief merit of Dennett’s intentional stance is that it provides a perspective within which we can situate mental states, and telling us where to start looking for them: on his theory, we should start by looking for behaviour that can be described by the intentional stance.

In basing his account of intentionality on the philosophical notion of “aboutness”, Dennett lays himself open to the charge that his intentional stance is too *broad* in its scope: as we saw above, Dennett applies it to thermostats, and even opioid receptors in the brain, to use one of Dennett's examples, are "about" something else: they have been "designed" to accept the brain's natural pain-killers, endorphins. Anything that can "embody information" possesses intentionality (1997, p. 48). However, even Dennett's severest critics, such as Searle (1999), do not dispute that the intentional stance can be applied to almost any kind of mental state (objectless moods being a possible exception), and to every kind of mind that may exist in the natural world.

Searle concedes that mindless systems may exhibit what he calls "as-if intentionality": they behave as if they had genuine (i.e. mindful) intentionality, and can be metaphorically described as such (1999, p. 93). The real point at issue between Searle and Dennett is whether the intentionality of our mental states is a basic, intrinsic feature of the world, or whether it can be reduced to something else.

1.6.4.1 Is Dennett’s intentional stance tied to reductionism?

At the outset of my quest for mental states in animals and (possibly) other organisms, I committed myself to an open-ended investigation, which avoided making philosophical assumptions about the nature of "mind" or "mental states". If Dennett's intentional stance turned out to be wedded to a particular, contentious account of "the mind", then its legitimacy would be open to challenge from the outset.

Dennett makes the highly contentious reductionist claim (1997, pp. 27, 30-31) that intentional agency in human beings is grounded in the pseudo-agency of the macromolecules in their bodies. This claim has been contested by Searle, who argues (1999, pp. 90-91) that it is vulnerable to the *homunculus fallacy.* In its crudest version, the homunculus fallacy attempts to account for the intentional "aboutness" of our mental states by postulating some "little man" or "spectator" in the brain who deems them to be about something. Although Dennett does not account for the intentional "aboutness" of our mental states in this way, he does attempt to solve the problem by taking it down to a lower biological level, where the problem of "aboutness" is said to be disappear: the intentionality of our mental states is the outcome of the mini-agency of the macromolecules in our bodies, and the intelligent homunculus is replaced by a horde of "dumb homunculi", each with its own specialised mini-task that it strives to accomplish (Dennett, 1997, pp. 30-31). Searle (1999, pp. 90-91) argues that this move merely postpones the problem: what gives our macromolecular states the intentional property of "aboutness"?

Nor does Searle think much of causal accounts of "aboutness", where the intentionality of our symbols is said to be due to their being caused by objects in the world. The fatal objection to causal accounts is that the same causal chains may generate non-intentional states as well (1999, p. 91).

The fundamental divide between Dennett and Searle on intentionality concerns whether there is such a thing as "intrinsic intentionality" (whereby our mental states have a basic property of "aboutness"), as distinct from "derived intentionality" (whereby "words, sentences, books, maps, pictures, computer programs", and other "representational artifacts" (Dennett, 1997, pp. 66, 69) are endowed with an agreed meaning by their creators, who intend them to be "about" something). For Dennett, the distinction is redundant because the brain is itself an artifact of natural selection, and the "aboutness" of our brain states (read: mental states) has already been determined by their "creator, Mother Nature", who "designed" them (1997, p. 70).

This move by Dennett is something of a fudge: "Mother Nature" (to borrow Dennett's anthropomorphism) does not "design" or "intend" anything; it merely *causes* things to happen, and as Searle has pointed out, causation is insufficient to explain intentionality. Searle (1999, pp. 89-98), while agreeing with Dennett that intrinsic intentionality is a natural, biological phenomenon, insists that there is an irreducible distinction between constructs such as the sentences of a language, whose meaning depends on what other people (language users) think, and conscious mental states such as thirst, whose significance does not depend on what other people think. Mental states, and not human constructs, are the paradigm cases of intentionality, and it is just a brute fact about the natural world that these conscious states (which are realised as high-level brain processes), refer intrinsically. An animal's conscious, intentional desire to drink, to use one of Searle's examples, is a biologically primitive example of intrinsic intentionality, with a natural cause: increased neuronal firing in the animal's hypothalamus. "That is how nature works" (1999, p. 95). Searle thus eschews both mysterian (dualist) and eliminative (reductionist) accounts of intentionality.

Despite the fierce controversy that rages over the roots of intentionality and the reducibility of mental states, it is admitted on all sides of the debate that a wide variety of entities can be treated as if they were agents in order to predict their behaviour. This, to my mind, is what makes Dennett's intentional stance a fruitful starting point in our quest for bearers of mental states. The issue of whether mental states can be reduced to mindless, lower-level processes is independent of the question of whether the intentional stance can be used to search for mental states.

1.6.4.2 Two intentional stances?

It is my contention that our intentional discourse comes in different "flavours", some richer (i.e. more mentalistic) than others, and that Dennett's intentional stance can be divorced from the use of terms such as "beliefs" and "desires". It is important, when describing the behaviour of an organism, to choose the right "flavour" of discourse - that is, language that is just rich enough to do justice to the behaviour, and allow scientists to explain it as fully as possible. It was argued in the Introduction that we should use mental states to explain the behaviour of an organism if and only if doing so allows us to describe, model and predict it more comprehensively, and with as great or a greater degree of empirical accuracy than other modes of explanation. If we can explain the behaviour of an intentional system just as well without recourse to talk of mental states such as "beliefs" and "desires", then the ascription of mental states is scientifically unhelpful.

Dennett's use of terms such as "information" (1997, p. 34) and "goals or needs" (1997, pp. 34, 46) to describe the workings of thermostats (1997, p. 35), shows that intentional systems *do not* always have to be described using the mentalistic terminology of "beliefs", "desires" and "intentions", in order to successfully predict their behaviour. An alternative "language game" is available. There are thus at least two kinds of intentional stances that we can adopt: we can describe an entity as having *information*, *or* ascribe *beliefs* to it; and we can describe it as having *goals*, *or* ascribe *desires* and *intentions* to it.

What is the difference between these two intentional stances? According to Dennett, not much: talk of beliefs and desires can be replaced by what Dennett (1995a) characterises as less colourful but equally intentional talk of semantic information and goal-registration. *Pace* Dennett, I would maintain that there are some important differences between the "information-goal" description of the intentional stance and the "belief-desire" description.

1.6.4.3 A goal-centred versus an agent-centred intentional stance

One difference between the two stances is that the former focuses on the *goals* of the action being described (i.e. what is being sought), while the latter focuses on the *agent* - in particular, what the agent is trying to do (its intentions). The distinction is important: often, an agent's goal (e.g. food) can be viewed as extrinsic to it, and specified without referring to its mental states. All the agent needs to attain such a goal is relevant information. A *goal-centred* intentional stance (which explains an entity's behaviour in terms of its goals and the information it has about them) adequately describes this kind of behaviour. Other goals (e.g. improving one's character, becoming more popular, or avoiding past mistakes) cannot be specified without reference to the agent's (or other agents') intentions. An *agent-centred* intentional stance (which regards the entity as an agent who decides what it will do, on the basis of its beliefs and desires) is required to characterise this kind of behaviour.

1.6.4.4 Narrowing the search for mental states: the quest for the right kind of intentional stance

It was suggested above that we should use mental states to explain the behaviour of an organism if and only if doing so allows us to describe, model and predict it more comprehensively, and with as great or a greater degree of empirical accuracy than other modes of explanation. Using Dennett's intentional stance, we can now clarify the task at hand in our search for entities with mental states. Having identified "mind-like" behaviour - i.e. behaviour that can be described using the *goal-centred* intentional stance, our next question should be: what kinds of mind-like behaviour, by which entities, are most appropriately described using an *agent-centred* intentional stance? The goal-centred stance is thus our "default" position. A switch to a mentalistic account (i.e. an agent-centred stance, which explicitly refers to *beliefs* and *desires*) is justified *if and only if we conclude that it gives scientists a richer understanding of, and enables them to make better predictions about, the organism's behaviour.* I shall henceforth refer to this supposition as my investigative hypothesis.

Dennett approvingly cites the example of a logger who told him: "Pines like to keep their feet wet" (1997, p. 45). Describing the behaviour of pines from a mentalistic perspective is wholly appropriate in the domain of poetry. However, I maintain that the use of such mentalistic language by scientists is justified only if it furthers their understanding of how an entity functions, in a way that mind-neutral language could not. The sentence "Pines thrive on moisture", by contrast, implicitly acknowledges that pines have a good of their own, while avoiding unnecessary mentalism.

In the Introduction, I noted that Dennett’s *intentional stance* raises three issues that are relevant to my quest for a minimal mind. First, how realistically should we construe the beliefs described in Dennett's intentional stance? According to the approach being suggested here, we can define the search for "mental states" in organisms as a search for behaviour that can *only* be explained by attributing *beliefs* and *desires* to the entities engaging in that behaviour – in other words, in terms of an *agent-centred* intentional stance, rather than a merely *goal-centred* stance.

Second, is Dennett correct in equating "belief-talk and desire-talk" with "their less colorful but equally intentional alternatives; semantic information-talk and goal-registration-talk" (1995a)? If I am right, the answer is no: systems whose behaviour can be adequately described using the latter stance can be regarded as mindless. Beliefs and desires of some variety appear to be a *sine qua non* for having mental states.

Finally, is there a fundamental difference between the intentionality of devices like thermostats and that of human agents? If my proposal is correct, then there must be a relevant difference between the intentionality of devices, which can be satisfactorily described using a *goal-centred* intentional stance, and the *agent-centred* stance required to adequately characterise human behaviour and some forms of non-human animal behaviour. For instance, the essential difference between the goal-centred stance and the agent-centred stance might turn out to be that events best described by the latter stance exhibit a distinct kind of normativity, which warrants the use of mentalistic terminology.

Section B: The Ingredients of Intentional Agency

Section B – The Ingredients of Intentional Agency

In this section, I focus on several capacities which are thought to be pertinent to the possession of mental states: sensory capacities and discriminative ability; memory, flexible behaviour, the ability to learn and represent; self-directed movement and control; and the ability to correct one's mistakes.

A Biological Prelude

The following diagrams and accompanying explanations may be of assistance, as the discussion in section B assumes that the reader is familiar with the major groups into which organisms are classified.

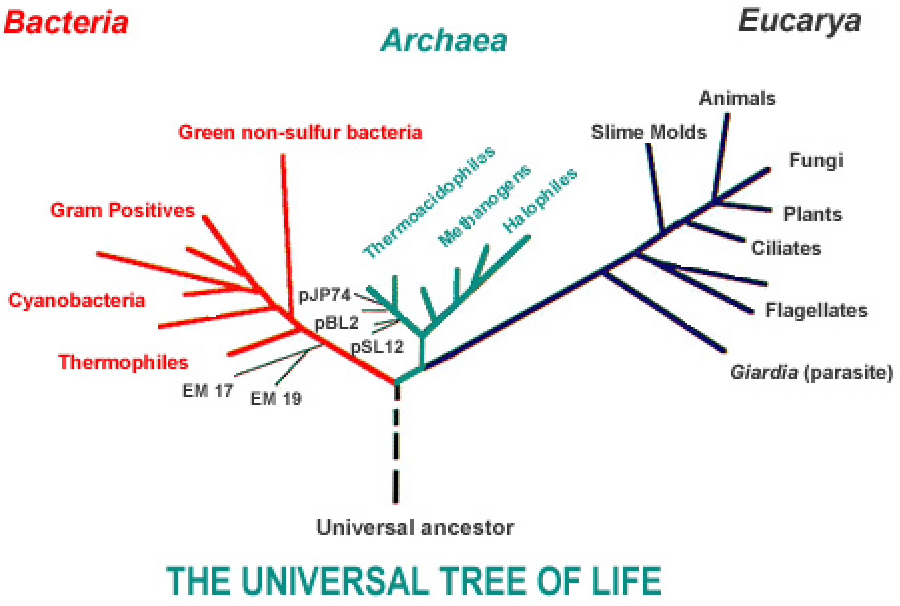
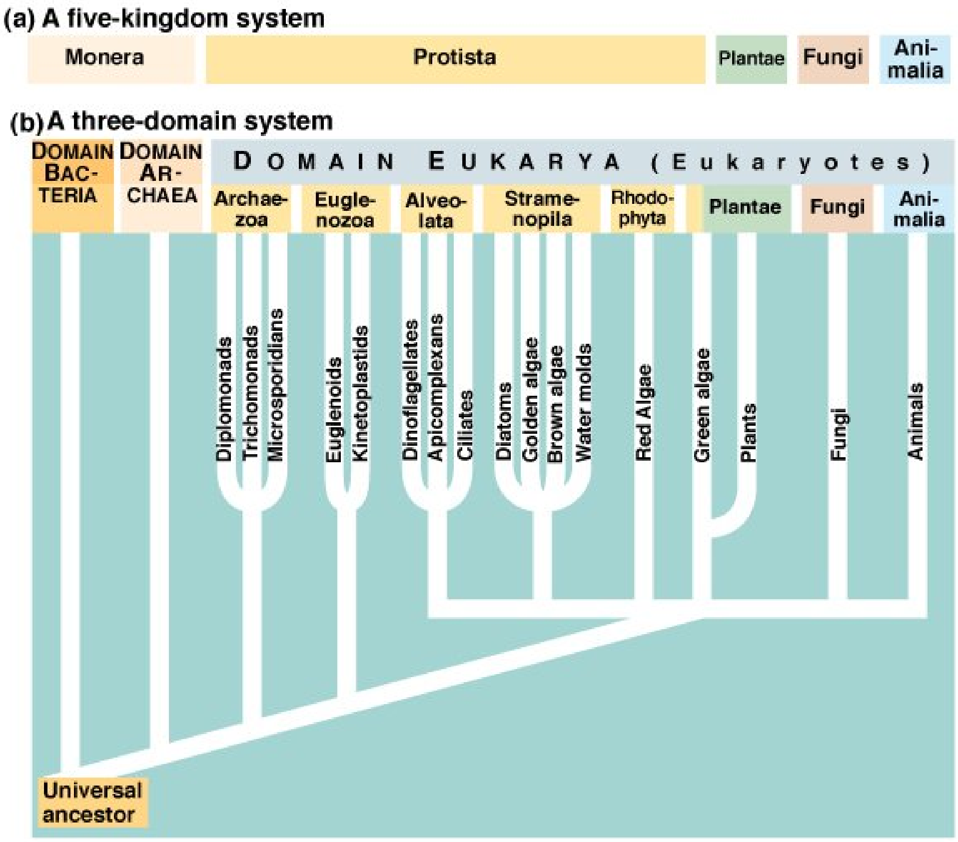


Figure 2.1 The tree of life, showing the three main groups (domains). Note that *animals, plants* and *fungi* are just *tiny twigs* on *one* of the main branches (eucarya, also known as eukaryotes). The branching order is an approximation. Source: Astrobiology 202 lecture notes by Dr. Michael Meyer, Steward Observatory, University of Arizona



©1999 Addison Wesley Longman, Inc.

Figure 2.2 The three domains and five kingdoms of life. The branching order is an approximation only. Copyright Addison Wesley Longman, Inc. and courtesy of Dr. Steve Carr, Department of Biology, Memorial University of Newfoundland

Historical background

Living things were originally classified by Linnaeus into two kingdoms (plants and animals). The two-kingdom classification gradually fell into disfavour, as evidence began to accumulate in the 19th century, following the work of Darwin and Haeckel, that the two-kingdom scheme was too narrow to express the diversity of life. The five-kingdom system, first proposed in 1969, placed organisms whose cells lack a nucleus and whose DNA floats freely within the cell *(prokaryotes)* in the kingdom Monera, while all other living things were classified into four kingdoms: *protoctista* (formerly known as protists), *plants, fungi* and *animals.* Thanks to the pioneering work of Dr. Carl Woese, biologists now realise that the differences between various kinds of prokaryotes dwarf those between the other four kingdoms. Prokaryotes are now placed in two distinct domains: *(eu)bacteria* and *archaea.* All other organisms belong in a third domain: *eukaryotes.*

Definitions of domains:

The two domains of *Bacteria* and *Archaea* are now recognised as fundamentally distinct. *Archaea* are a recently discovered group of microorganisms that can live in extreme habitats such as thermal vents or hyper-saline water.

All other organisms (protoctista, plants, fungi and animals) are placed in a third domain: *Eukaryotes*, or organisms whose DNA is stored in a nucleus. Eukaryotic cells possess an extra level of complexity which prokaryotes lack: they contain specialised structures called *organelles*. Eukaryotes share anatomical similarities which make their information transfer pathways more complex as well as faster than those of prokaryotes. All eukaryotes make use of *rapid electrochemical signalling* to transmit information, in addition to the slow process of chemical diffusion used by bacteria. The processes whereby materials (e.g. nutrients) and information are diffused within a eukaryotic cell are more complex than those in a prokaryotic cell, simply because eukaryotic cells are so much bigger (Kaiser, 2001; Illingworth, 1999; Cotterill, 2001, p. 5).

Definitions of kingdoms used in Whittaker's five-kingdom system (still used by some authors):

1. *Monera:* includes bacteria (as well as archaea), whose cells lack nuclei. Most but not all are single-celled (unicellular) organisms. Monera are *prokaryotes;* the members of the other kingdoms are *eukaryotes.*

2. *Protoctista* ("first builders"): includes single-celled microbes with nuclei, that were previously labelled protista (protozoa and protophyta), as well as certain multi-celled organisms, such as kelp, that do not belong to the plant, animal or fungi kingdoms. Amoebae, algae, seaweeds, slime moulds, ciliates, diatoms, paramecia and forams belong in this kingdom. Although many protoctista are commonly lumped together with bacteria and viruses as "microbes", they are actually much more like plants, animals and fungi, at a cellular level.

3. *Fungi:* placed in a kingdom apart from plants, because their life-cycle, feeding habits (they do not photosynthesise) and ecological adaptation are so different. Most fungi (unlike plants) build their cell walls out of a hard material called chitin. The living body of a fungus, known as the *mycelium* (usually hidden beneath the surface), is made out of a web of tiny filaments called *hyphae.* The part of the fungus that we see is only the "fruit" of the organism.

4. *Plants:* multicellular organisms which develop from embryos and whose cells contain chlorophyll, enabling them to make their own food. Photosynthesis is not unique to plants; it is found in certain bacteria (cyanobacteria) as well. The cell walls of plants are made of cellulose. Different authorities define plants on the basis of either chemistry (restricting plants to the groups where the cells contain a membrane-bounded *chloroplast*, where the products of photosynthesis are stored, with a particular type of *chlorophyll*) or the structure of the wall of the chloroplast.

5. *Animals:* multicellular organisms that are heterotrophic (feeding on others), diploid (with two sets of chromosomes), and that develop from the fusion/fertilization of an egg and a sperm cell (as do plants). Unlike plants and fungi, all animals (except for a few sponges) follow certain stages during their embryonic development: a morula appears first, then a blastula and finally a gastrula - a hollow sac that forms the embryonic precursor to the digestive tract, by means of which animals ingest nutrients and excrete waste. However, the only feature that appears common to all animals and only animals is a special kind of *extracellular matrix,* which plays a key role in the development of animals. (Sources: McCourt, Chapman, Buchheim and Mishler, 1995; McGraw-Hill Encyclopedia of Science and Technology, 2005; Fun Facts about Fungi, 1998; Morris, 1993).

There is a fundamental difference between Aristotle's *teleological* definition of animals, and the contemporary scientific definition. For Aristotle, the existence of *sensory capacities* was a defining characteristic of animals: without these discriminatory abilities, animals could not survive, avoid danger or acquire what they need (*De Anima* 3.9, 432a16, 3.12, 434a30 - 434b3). Locomotion was a sufficient but not a necessary condition for being an animal (*De Anima* 3.9, 432a16; 3.9, 432b19-20, 3.12, 434b8). By contrast, the modern scientific definition of "animal" is based on the fundamental similarities, in their structure and their bio-molecules - especially at the cellular level - between organisms that were formerly classified as animals on the basis of their sensory capacities.

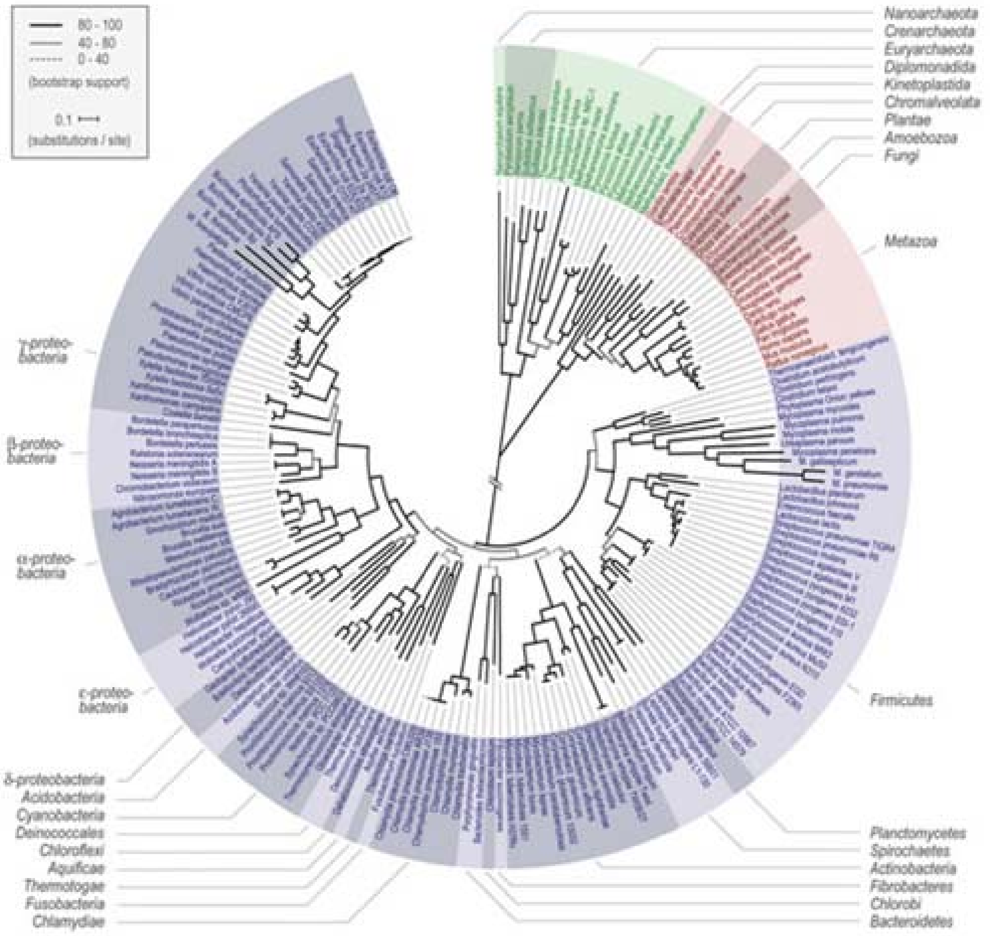


Figure 2.3 A new family tree of life developed by a group at the European Molecular Biology Laboratory (EMBL) in Heidelberg, 2006. Courtesy of European Molecular Biology Laboratory. Web address: http://www.embl.org/aboutus/news/press/2006/02mar06/index.html. Source article: F. D. Ciccarelli, T. Doerks, C. von Mering, C. J. Creevey, B. Snel & P. Bork. “Towards automatic reconstruction of a highly resolved tree of life.” In *Science*, 3 March 2006.

Bacteria are shown in blue, archaea in green and eukaryotes in red. The key points here are: the predominance of microbes, which make up the bulk of the branches of the tree; and the fact that the various groups of animals (here referred to as *Metazoa*, the largest group shown in red) account for only a small part of the diversity of life.

Chapter 2. Sensory Criteria for Identifying Mental States in Organisms

The main themes that have emerged from my investigation of sensory capacities in living things are the need for definitional clarity, more careful distinctions between different grades of sensitivity and a greater understanding of the relation between sensitivity and motion in organisms.

2.1 What is a sensor?

There is currently no agreed definition of the term "sensor" in the literature. I propose to employ the following definition:

> Any device that receives a signal or stimulus (as heat or pressure or light or motion etc.) and responds to it in a distinctive manner (taken from WordNet, Cognitive Science Laboratory, Princeton University, 2006).

The above definition of “sensor” highlights three significant points. First, a sensor is an information-encoding device, or *receptor.* However, a sensor is more than a mere receptor; it also has to generate a *response.* Second, the response has to be *specific* to the stimulus detected: the ability to generate a random response to a stimulus does not make a device a sensor. Third, *any kind of material* is capable in principle of acting as a sensor: camera film is photosensitive, as are metals which release electrons when exposed to light (the photoelectric effect). The bimetallic strip in a thermostat is a temperature sensor.

In chapter one, it was argued that being *alive* is a necessary condition for having mental states, Some devices with sensors are manifestly not alive. We may therefore conclude:

> Conclusion 2.1: The possession of sensors cannot be a sufficient condition for having mental states.

Any sensor can be described using Dennett's *goal-centred intentional stance:* it is a "micro-agent, ... an utterly minimal intentional system whose life project is to ask a single question, over and over - 'Is my message coming in NOW?' ... - and spring into limited but appropriate action whenever the answer is YES" (Dennett, 1997, p. 108). It is intentionality at this level, Dennett argues, that allows an organism to process information about its surroundings.

Although we can describe a sensor (and by extension, any entity possessing sensors) as being *sensitive* to the stimulus it responds to, I shall use the verb “detect” rather than “sense” to denote what a sensor *does.* The latter verb will be reserved for any organisms that can be said to possess *senses* (a term we have yet to define).

Case Study 1: the lysis-lysogeny decision in viruses

A parasitic lambda-phage virus is able to respond to changes in its environmental conditions by adopting a bet-hedging strategy in order to cope with fluctuations in the availability of their hosts *(E. coli* bacteria). This strategy is commonly referred to in the scientific literature as the *lysis-lysogeny* decision. When the virus invades its host bacterium *(E. coli),* it may kill its host immediately by multiplying until the host's cell walls burst *(lysis)* or it may remain dormant and may confer immunity to infection upon its host *(lysogeny).* The important point here is that the strategy a virus adopts on a particular occasion is a *random* one, which is triggered by thermal background noise at the molecular level (Preuss P., 2000).

The lambda-phage virus exhibits what scientists call *phenotypic plasticity*, which Ancel Meyers (2003) has defined as “the ability of organisms with the same genotype to vary their developmental pattern, phenotype or behaviour in response to varying environmental conditions.” If we merely defined a sensor as "a device that responds to a physical stimulus", then we might conclude that these viruses possess sensors, as they respond to environmental change. But in fact, viruses, which are little more than living molecules, have no built-in signal *receivers* and therefore possess no *information* that enables them to realise their goals. Thus their behaviour, when making the lysis-lysogeny decision, does not even meet the criteria required for the adoption of Dennett's *goal-centred intentional stance*. The viruses do not possess (i.e. encode or store) *information* about environmental conditions that would help them to achieve their *goal* (replication), but simply respond to changing conditions in a random, non-specific manner.

> Conclusion 2.2: An organism's possession of sensors, which enable it to encode information about its environment and thereby achieve its goals, is a necessary condition for our being able to describe the organism's behaviour according to Dennett's *goal-centred* intentional stance.

> Conclusion 2.3: An organism must be capable of encoding and storing information about its environment before it can be said to possess mental states (in particular, beliefs and desires). (Corollary of my Investigative Hypothesis, plus Conclusion 2.2.)

2.2 What grades of sensitivity are found among organisms?

Cotterill (2001) provides an excellent overview of the key distinctions drawn by biologists investigating the different levels or *grades* of sensitivity found in various kinds of organisms.

Case study 2: bacteria

Bacteria, unlike viruses, certainly possess sensors. According to John S. Parkinson, a professor of biology at the University of Utah, "most organisms - even bacteria - can sense sound, light, pressure, gravity and chemicals" (University of Utah, 2002). *E. coli* bacteria "can sense and respond to changes in temperature, osmolarity, pH, noxious chemicals, DNA-damaging agents, mineral abundance, energy sources, electron acceptors, metabolites, chemical signals from other bacteria, and parasites" (Meyers and Bull, 2002, p. 555). Bacteria are very sensitive to chemicals - for instance, *E. coli* bacteria have five different kinds of sensors which they use to detect food. As Di Primio, Muller and Lengeler (2000, pp. 4 - 5) explain, common bacteria like *E. coli* swim in chemical gradients towards attractants (e.g. glucose) or away from repellents (e.g. benzoate) - a phenomenon known as chemotaxis. Other bacteria display phototaxis and magnetotaxis, or movement in response to light and magnetic fields, respectively (Martin and Gordon, 2001, p. 219).

Bacteria possess specialised "receptors" or information-encoding devices, which are sensitive to light, chemicals, magnetic fields and so on. These receptors may or may not be activated, depending on the local environment. A bacterium has two kinds of motion: *directed movement* (a "run", which occurs when a bacterium's rotary motors or flagella, rotate in a counter-clockwise direction) and *random tumbling* (which occurs when a bacterium's flagella suddenly change direction and rotate clockwise).

When the external section of a bacterial receptor recognises and binds its target, a signal passes through the rest of the receptor and causes sequential changes in two proteins inside the bacterium. (This two-protein sensing system is found in all bacteria and in many other life-forms, but not in animals.) The first protein is called a kinase and sits next to the receptor. Normally, when there is no signal, this protein activates a second protein, the regulator, which interacts with the gear shift of a bacterium's flagella, causing them to turn clockwise and the bacterium to tumble randomly, about once every second. However, when there is a signal from the receptor, the kinase cannot activate the regulator protein. Thus, the flagella continue to turn counterclockwise, and the bacterium, instead of tumbling, swims smoothly towards the target or *goal* (Aegerter, 1997). What is more, these receptors can even store *information* about their objects over a short period of time - in other words, they possess a kind of chemical "memory" (to be discussed later).

> Conclusion 2.4: All cellular organisms (including bacteria) possess sensors that can encode various states of information about their surroundings that helps them achieve their goals.

Dennett (1997, p. 84) regards the sensitivity displayed by phototactic bacteria as fundamentally the same as the photosensitivity of light meters in cameras. I would argue that he is profoundly mistaken here, for reasons discussed in chapter one: in the former case, but not the latter, the sensors are part of the body of a living individual, which has a good of its own (intrinsic finality) and therefore benefits from its ability to sense changes in its environment.

Cotterill (2001) describes four evolutionary milestones in the evolution of what he calls “true senses”. Below, I discuss four major evolutionary milestones highlighted by Cotterill, discuss the different meanings of "sense" and highlight the implications for intentional agency.

Evolutionary milestone one: the appearance of electrochemical signalling in eukaryotes

Case study 3: Protoctista

Sensory capacities such as chemotaxis, thermotaxis (movement in response to heat), phototaxis, geotaxis (movement in response to gravity) and an ability to identify suitable mates are well-attested for protoctista (Martin and Gordon, 2001, p. 409).

There are some important differences between bacteria and protoctista which determine the way they sense objects in their environment. Protoctista, unlike bacteria, are *eukaryotes.* There are anatomical similarities shared by all eukaryotes which make their information transfer pathways more complex as well as faster than those of prokaryotes. Eukaryotic cells not only contain a nucleus, but are also about 10 times larger than bacterial cells. For that reason alone, the processes whereby materials (e.g. nutrients) and information are diffused within a eukaryotic cell are more complex than those in a prokaryotic cell (Kaiser, 2001; Illingworth, 1999; Cotterill, 2001, p. 5).

A motile eukaryotic cell also travels much faster than a prokaryotic cell (e.g. a bacterium), and therefore encounters danger or opportunity far more frequently. It needs a way to communicate *rapidly* between the forward end of the cell and the flagella, which usually propel it from the rear. Chemical signals, whether transmitted by diffusion or circulation, do not move fast enough from one end of a eukaryotic cell to another to meet its needs for rapid communication of information.

Accordingly, all eukaryotes (protoctista, plants, fungi and animals) make use of rapid *electrochemical signalling* to transmit information, in addition to the slow process of *chemical diffusion* used by bacteria.

Case study 4: Plants

Plants possess an impressive range of sensory capacities, which have been described in an informative article by Coghlan (1998). For instance, research on a humble weed known as thale cress *(Arabidopsis thaliana*, a plant with a relatively simple genome) reveals that some plants can "see". They have proteins attached to light-sensitive compounds. Proteins called phytochromes enable plants to work out the quality of light and compete with neighbouring plants. Other proteins called crytochromes enable plants to work out whether it is night or day, the length of the day, the quantity of light, and the direction it is coming from.

Plants are also known to have analogues of our other senses: *taste* (the roots of some plants "taste" the soil and find out where vital nutrients are most abundant, thereby saving energy, as the roots grow towards the source of the nutrients, rather than randomly); *smell* (some plants exude a substance when wounded, which alerts neighbouring plants to produce chemicals that repel insects or attract predators); *touch* (the responsiveness of the Venus fly trap, or Mimosa, to touch is well-known, but these abilities are simply amplifications of what all plants can do); and even *hearing* (prolonged exposure to 2 kilohertz frequencies at 70 to 80 decibels can increase the germination rate and the growth rate of certain plants) (Coghlan, 1998).

Evolutionary milestones two, three and four: nervous systems, movement mediators and reflexes

Case study 5: Cnidaria

Nerve cells are only found in animals. In fact, they are unique to so-called "true" animals (the subkingdom *Eumetazoa*, which excludes sponges). The simplest of these "true" animals are the *Cnidaria* - commonly known as coelenterates, including animals such as jellyfish, sea anemones, corals and freshwater hydra, which possess the most rudimentary *nervous systems* found in nature. Fossil evidence indicates that cnidaria were present in the Ediacaran period, 635 to 542 million years ago (Prescott, 2007, pp. 3 - 4). Cotterill describes the transition:

> Instead of merely being deployed on the organism's bonding surface, the receptor molecules became incorporated in the membranes of the multicellular creature's specialized receptor cells (2001, p. 5).

Prescott (2007, pp. 5-6) regards cnidaria as an important advance over sponges:

> Whilst the most primitive metazoans, the sponges, lack neurons and respond only to direct stimulation (usually with a very slow, spreading contraction), cnidarians have quite complex nervous systems, composed, for the most part, of distributed nerve nets,and show both internally generated rhythmic behaviour, and coordinated patterns of motor response to complex sensory stimuli (2007, p. 5).

Cotterill (2001) describes two types of jellyfish, *Aglantha digitale* and *Amphogona apicicata*, which exhibit dual reponse patterns, mediated by two different types of nerve impulses: a slow feeding mode, and a fast reaction mode which enables these jellyfish to rapidly escape from predators.

To Cotterill, the distinctive fast reaction mode in these jellyfish suggests the presence of two components which serve to distinguish what he considers to be *true senses* from those found in "lower" organisms:

(i) a neuron-based *movement mediator* which is capable of activating different motor programs in response to environmental feedback; and
(ii) a genuine autonomous *reflex*, characterised by "a receptor cell's reaction to an unprovoked stimulus leading unaided to a motor response" (2001, p. 5). For Cotterill, the most significant feature of this response is that it is independent of the creature's internal state.

Prescott (2007) cites findings which appear to lend further support to Cotterill’s proposal. According to Prescott, *Aglantha digitale* “is thought to have the most complex neural circuitry of known Cnidaria” (2007, p. 7), with twelve distinct neuronal circuits and two cell-to-cell condiction systems. Additionally, many other cnidarians are now known to have fast-signal conduction mechanisms, which support fast attack or escape reactions: the hydrozoan jellyfish, whose nerve net is arranged in a circular tract; the scyphozoan jellyfish *Aurelia aurita* has two distinct nerve networks, which support rapid swimming and feeding respectively; while box jellyfish, which actively pursue their prey, have a condensed nerve ring that connects their ganglia together, presumably for faster transmission of signals (Prescott, 2007, pp. 6-8).

Recent research suggests that some cnidarians may be able to see images. Cubozoans, or killer box jellyfish, are known to have complex eyes, similar in their basic design to those of vertebrates, despite the absence of a brain or central nervous system. The eyes connect into the neural network of the jellyfish. It has been suggested that for cubozoans, vision may play a role in their feeding and reproductive behaviour, as some of them are capable of pursuing and targeting specific individuals (Martin, 2000; see also Nilsson *et al.,* 2005).

The occurrence of reflexes and *bona fide* senses in other, more "advanced" phyla of animals (e.g. arthropods, molluscs and annelids) is universally accepted (Cotterill, 2001, p. 5).

Case study 6: Aplysia

The aquatic mollusc *Aplysia* exhibits a further evolutionary innovation which goes beyond even the most advanced cnidaria: a rudimentary form of *sensory processor.* The chief advantage of a sensory processor is that it enables an animal's nervous system to find *correlations* between different sensory inputs, these being tied to *motor output* after undergoing some intermediate processing (Cotterill, 2001, pp. 5-6). Although its entire nervous system consists of only a few hundred neurons, *Aplysia* is capable of associative learning (i.e. classical conditioning as well as instrumental conditioning). I shall discuss associative learning in a subsequent chapter.

2.3 Which organisms have true senses?

If bacteria are sensitive to sound, light, pressure, gravity and chemicals, should we then say that they possess *bona fide* senses? Or is there a distinction between sensing an object, and being sensitive to it? For that matter, what are senses, anyway?

2.3.1 Only animals have senses: the Aristotelian perspective

Aristotle maintained that senses are not found in all living things. *Aesthesis* - which in his writings "is capable of bearing the meanings both of sensation and of perception" (Lawson-Tancred, 1986, p. 78; see also Sorabji, 1993, pp. 8, 15) - is found in animals, and only animals *(De Anima* 2.2, 413b1ff; 3.12, 434a30; 3.13, 435b1). Aristotle stresses that senses exist for a practical, teleological reason: they are *discriminative* capacities *(De Anima* 3.9, 432a16), which enable animals to survive. Without these capacities, animals cannot avoid danger or acquire what they need *(De Anima* 3.12, 434b - 1ff). This is particularly true of animals that move around: "If any body that travels did not have perception, it would be destroyed and so not achieve nature's function by reaching its purpose" (*De Anima* 3.12, 434a33-34). Aristotle regarded locomotion as a sufficient but not a necessary condition for being an animal (*De Anima* 3.9, 432a16; 3.9, 432b19-20, 3.12, 434b8).

Aristotle was unaware of the existence of bacteria, which also travel and are sensitive to their surroundings. He was also unaware of the sophisticated sensors possessed by many plants. Had he known what we know now, would he have attributed perception to bacteria and plants? To answer this question, we need to carefully examine the distinctions drawn by Aristotle between different kinds of receptivity in his writings on perception.

First, Aristotle argued that there is more to sensing an object than merely being affected by it:

> For perception is being affected *in a certain way.* Thus the active thing *[the sense object - V.T.]* makes that *[the sense organ - V.T.]* which is potentially like it, like it in actuality...[T]he sense is the recipient of the perceived forms without their matter, as the

> wax takes the sign from the ring without the iron and gold... And it is also clear why it is that plants do not perceive, though they have a psychic part and are in some way affected by the touch-objects. After all, they become cold and hot. The reason, then, is that *they do not have a mean, nor such a principle as can receive the forms of the sense-objects,* but are affected by the matter as well *(De Anima* 2.11, 424a1ff; 2.12, 424a17-19, 424a34-424b3, italics mine).

What Aristotle seems to be arguing is that although a living thing which senses an object (with one of its sense organs) is altered by that object, it is not altered by taking that object into itself, but rather by taking on the object's form, without its matter. Rephrasing Aristotle's insights in contemporary terminology (which is etymologically rooted in the form-matter distinction), we might say that for Aristotle, to be able to sense or perceive an object means, roughly, to be a living thing with a sensory organ that can encode *information* about that object.

There remains the question of what Aristotle meant by saying that *aesthesis* requires the possession of a "mean", which Aristotle seems to think plants lack. Aristotle elaborates in the following passage:

> [T]he sense itself is a "mean" *[mesotes]* between any two opposite qualities which determine the field of that sense. It is to this that it owes its power of discerning objects in that field. What is "in the middle" is fitted to discern; relatively to either extreme it can put itself in the place of the other (*De Anima* 2.11, 424a4-7).

It is important to grasp that for Aristotle the *sense* and the *sense organ* are not the same thing:

> Aristotle does not say that the sense is “the power of the sense-organ to change” … but rather that it is “the power of *discerning* the objects in that field.” [italics mine] Thus Aristotle’s distinction between the sense and the sense-organ is quite marked. The sense-organ is that which takes on the form of the object that is acting on it, whereas the sense is a “mean” between two extremes which can perceive (measure) the change that occurs in the sense-organ (Fowler, 1995, online, section 4).

This interpretation appears to be borne out by the following passage, in which Aristotle carefully distinguishes between sense and sense organ:

> Each sense, then, is of the sensible thing that is subject to it, is present in the sense organ *as* sense organ and discerns the variations in the sensible thing that is its subject. For instance, sight discerns variations in whiteness and blackness… (*De Anima* 3.2, 426b9ff).

If this interpretation of Aristotle is correct, then sensing is a two-step process for Aristotle. First, an organ of a creature that senses is "informed" by the stimulus it detects (that is, receives the form of its sense-object without the matter); second, the degree of change or variation in the organ’s state when it encounters the stimulus must be somehow measured.

There are thus two ways in which sensitivity in an organism can fail to qualify as a *bona fide* sensory capacity for Aristotle: either the actual state of the object sensed may not be formally encoded as information, or there may be no measurement of how much change the sensory organ has undergone. A chemical reactant fails the first of Aristotle's conditions: it receives the matter as well as the form of its object. A binary sensor that merely detects the presence or absence of the stimulus fails the second condition: because there is no measurement of the degree of change it has undergone, it cannot properly be said to sense the stimulus.

It has been suggested that Aristotle would have regarded bacteria and other microbes as animals, since they are capable of locomotion, which he regarded as a capacity possessed only by some animals *(De Anima* 3.9, 432b19). In his own words: "No non-stationary body has a soul without perception...If, then, any body that travels did not have perception, it would be destroyed... After all how is it to be nourished?" *(De Anima* 3.12, 434b2-3, 434a33, 434b1). It should also be pointed out that van Leeuwenhoek, the first person to observe bacteria through his microscope, called them "little animals" or "animalcules" (Waggoner, 1996). Moreover, bacteria possess *receptors* that can encode information about their objects (attractants and repellents), at different actualisations (i.e. concentrations of attractants), spanning five orders of magnitude (Illingworth, 1999). Thus they certainly satisfy Aristotle’s first condition.

The bacterial cell can also compare its present state with its state a short time ago. The number of receptors stimulated by attractive or repellent molecules – apparently this number is an average of measurements taken over a period of about one second – is "compared" with the number of receptors stimulated in the previous measurement (stored as an internal signal representing the average of measurements taken 3 seconds earlier). Although a bacterial cell can store just *one* set of intermediate results, its ability to *compare measurements* taken at the beginning and end of a three-second interval arguably meets Aristotle’s second condition for *bona fide* perception: the ability to discern changes in the sense organ.

Given that the capacity of protoctista and plants to store information and measure changes in their sensors is considerably more sophisticated than that of bacteria, it is even harder to see how these discriminatory abilities could fail to meet Aristotle’s criteria for having sensory perception. Indeed, a modern commentator, Charles Kahn, has argued that had Aristotle known what we know now, he "would have been obliged to" grant one-celled animals "a share in perception proper... since the possession of a sense faculty is included in the definition of an animal" (1979, p. 25).

It was noted above (Coghlan, 1998) that plants use certain proteins to work out the quality of light and compete with neighbouring plants, while other proteins enable them to discern whether it is night or day, the length of the day, the quantity of light, and the direction it is coming from. Although Aristotle denied that plants can perceive *(De Anima* 2.3, 414a31, 3.13, 435b1), it appears to follow from his own criteria that he would have revised his conclusion, in the light of what we now know.

2.3.2 Cotterill's arguments for denying true senses to bacteria

Cotterill (2001) draws attention to the fact that organisms' sensors differ vastly, not only in their degree of sophistication, but also in the manner in which they move organisms to obtain their goals.

Cotterill denies the existence of proper senses in bacteria, because the order of *stimulus* and *response* is reversed: instead of environmental changes acting as the stimulus which causes a motor response in a bacterium, the bacterium initiates its own random tumbling movements and thereby gains information about its surroundings, using a short-term memory that informs it as to whether the concentrations of certain molecules in its environment have changed in the last few seconds. As Cotterill puts it:

> The stimulus in this case is thus the motor movement, while the response is that of the impinging substances. This is just the opposite of a reflex... There are no senses, of the type found in more advanced species, and the internal state of the creature is embodied in the concentrations of various molecules. These concentrations dictate the creature's movements (2001, pp. 3-4).

In a creature with reflexes, by contrast, the motor response is "independent of the creature's internal state" (2001, p. 5), and the reaction of a *specialised* receptor cell in the creature's body to "an unprovoked stimulus" (2001, p. 5, italics mine) leads to a rapid, automatic motor response.

Cotterill's main argument can be recast in an Aristotelian form. He acknowledges that bacteria can store information relating to the concentrations of substances in their surroundings (2001, p. 6), but insists that sensing an object means something more than that. In Aristotle's account, sensing an object meant *being affected* by it in a certain way, and an animal's desire of the sensed object produces locomotion towards it *(De Anima* 3.10, 433a16). In bacteria, by contrast, "sensing" an object means *acting* upon it, and it is the perpetual movement (locomotion) of bacteria that enables them to "sense" chemicals. That is, locomotion is prior to "sensation".

We can express Cotterill's objection another way, by saying that sensations are a form of "feedback", whereas bacteria seem to use "feed-forward" instead to navigate around their environment. Certainly, bacteria are much more active in probing their environment than other sensitive organisms, because of their size: since they are too small to gauge spatial variations in the concentrations of molecules in their environment (e.g. differences between concentrations at their anterior and posterior extremities), they have to actively "sample" their surroundings, relying on a form of short-term chemical memory to alert them to changes.

However, there are in fact two kinds of bacterial motion: (a) the random tumbling movements which bacteria initiate in order to probe their surroundings, and (b) the directed "runs" which they make along chemical gradients towards attractants. Cotterill is quite right in characterising movement of the former kind as a stimulus and its feedback about its environment as a response. But it is only movement of the *latter* kind that Aristotle would characterise as locomotion, or "movement started by the object of desire" *(De Anima* 3.10, 433a16). This kind of motion is subsequent to, not prior to, the act of sensing the attractant. Moreover, the change in a bacterium's pattern of movement (from random tumbling to directed swimming) is produced by a combination of events occurring both inside and outside its body: its internally driven propulsion which allows it to probe its environment, and the variations in the concentrations of attractant or repellent molecules in its environment. This does not sound so different from the Aristotelian notion of perception as "being affected in a certain way" *(De Anima* 2.11, 424a1).

Nevertheless, Cotterill has performed a valuable service to philosophy by highlighting the link between sensory perception and *locomotion* in organisms. The manner in which sensing brings about locomotion varies greatly between different kinds of organisms. Cotterill restricts the term “true sense” to organisms in which the sensing of an object brings about a specialised pattern of locomotion towards it.

2.3.3 Eukaryotes vs. bacteria

The difference in *locomotion patterns* between bacteria and eukaryotes could be considered to be relevant to the philosophical definition of a *sense.* Whereas in bacteria it can be argued that locomotion is *prior* to "sensation", in eukaryotes it is the other way round: information received by these organisms' sensors makes them move towards their goals. Whereas bacteria rely on measurements taken at successive 3-second intervals to find food when probing their surroundings, eukaryotic cells, which are much larger than bacterial cells, are able to detect tiny spatial concentration gradients from one side of the cell to another when searching for food. It thus seems that locomotion in eukaryotes is *caused by* changes that they sense in their surroundings. (The distinction between prokaryotes and eukaryotes is not black-and-white: some one-celled eukaryotes, like bacteria, propel themselves by means of their flagella – long whip-like structures used in cellular locomotion – and obtain sensory information about their surroundings as a result of their own movements.)

Another difference between bacteria and eukaryotes is that eukaryotes use *electrochemical signalling* because their cells are larger than those of bacteria, and therefore require rapid transfer of information. Despite these differences, Cotterill (2001, p. 5) does not regard protoctista and sensitive plants as possessing "what we would call senses", as there are some important features which they lack. Because they do not possess nerve cells, nervous systems or reflexes, their behavioural repertoire is limited, and their motor responses are simply a function of their internal state (Cotterill, 2001, p. 5). The main practical difference between these organisms and bacteria is simply the speed-up of information transfer.

2.3.4 The manner in which true senses bring about locomotion

For Cotterill, the key criteria which distinguish organisms with true senses are:

- the presence of *dedicated* receptor cells (as opposed to receptor molecules on the surface of the organism);
- the presence of a neuron-based *movement mediator* which is capable of activating different motor programs in response to environmental feedback; and
- the presence of *reflexes*, which produce motor responses *independent* of the creature's internal state.

Earlier, we defined a sensor as a device that receives a signal and responds to it in a distinctive way. What Cotterill is arguing here is that in an organism with true senses, the response is a dedicated, specialised one: there is a specific *motor pattern* which is triggered when receptor cells of a certain kind are activated. This response is rapid and automatic (Cotterill, 2001, p. 5).

Although the specialised characteristics identified by Cotterill are highly significant, I think it would be unhelpful to restrict the term “senses” only to animals possessing these attributes. Instead, I propose to distinguish between broad and narrow usages of the term “sense”:

> Definition - "sense"
>
> (a) On a *broad* definition of "sense", any organism possessing (i) sensors that can encode and store information relating to a stimulus, and (ii) a built-in capacity to measure the degree of change in the sensor’s state when it encounters the stimulus, can be said to *sense* the stimulus.

(b) On a *narrower* definition, the verb "sense" can be restricted to organisms satisfying the conditions in (a), whose sensors are dedicated *receptor cells* which trigger a *distinctive, built-in, rapid-response motor pattern* which is *specific* to the signal and *independent* of the organism's internal state.

Conclusion 2.5: On the broad definition used above, all cellular organisms (including bacteria) can be said to possess senses. On the narrower definition, senses are confined to organisms with nervous systems, and appear to exist in two species of jellyfish, as well as all "higher" phyla of animals.

2.4 Sensation and perception

Sensing and *perceiving* are commonly distinguished from one another in both ordinary and philosophical parlance. The term “perception” suggests an interpretational and/or cognitive refinement, which goes beyond sensation: thus “concept” is one definition of the term “perception”; “physical sensation interpreted in the light of experience” is another (Merriam-Webster, 2006, online, entry “perception”). It is often assumed that first we experience a stream of raw sense-data, which we then process and upon which we subsequently impose our perceptual categories. Philosophers are also accustomed to drawing this distinction. For instance, Lawson-Tancred (1986) criticises the Aristotelian term *aesthesis* as a “bogusly wholistic concept”, which conflates the “two distinct activities of sensation and perception” (1986, p. 78). The former can be characterised as a transformation of a bodily organ (the sensor); the latter, however, is “closer to an act of judgement” (1986, p. 77).

This picture has been largely discredited by scientific findings in the fields of vision and olfaction (smell). Federmeier (1997) investigates the neurobiology of vision in her technical monograph, “Perceiving a New Category: The Neurobiological Basis of Perceptual Categorization”, in which she overturns the unexamined assumption, common to many current models of vision, that “categorization is separate from perception” (1997, p. 11). To avoid confusion, I should point out that Federmeier uses the terms “perception” and “categorization” for what most philosophers and laypeople would call “sensation” and “perception” respectively. The following quotes convey the thrust of her findings:

> Current models and theories of categorization have tended to assume that *categorization and perception are separable processes, with perception preceding categorization.* In contrast, this paper argues that categorization is a solution to a conflict faced by all information processing systems and gives evidence that the visual system faces this conflict and solves it by categorizing. Neurobiological data suggest that *perceptual categorization begins to take place in the earliest stages of visual processing* and is highly developed in visual areas such as the inferotemporal cortex. Attention and experience can be shown to affect the neurophysiology of [the] visual cortex in a manner analogous to their effects on categorization behavior. Together, these sources of evidence support an *inherent relationship between visual perception and perceptual categorization* (1997, p. 1, italics mine).

> *From the earliest stages of sensory processing,* neurons and groups of neurons respond to some kinds of inputs and not to others, creating perceptual categories that are the roots of all human experience (1997, p. 2, italics mine).

> By the time information reaches higher visual processing areas, a *complex categorical structure* seems to exist in the physical organization of neurons with particular response properties (1997, p. 50, italics mine).

> Evidence from *both* bottom-up and top-down sources, therefore, strongly supports a close relationship between perception and categorization. Perceptual categorization is an inherent part of bottom-up visual processing; *one must categorize in order to see* (1997, p. 21, italics mine).

Olfaction also imposes categories from the outset. Freeman (1991) describes the process:

> [W]hen an animal or a person sniffs an odorant, molecules carrying the scent are captured by a few of the immense number of receptor neurons in the nasal passages; the receptors are somewhat specialized in the kinds of odorants to which they respond. Cells that become excited fire action potentials, or pulses, which propagate through projections called axons to a part of the cortex known as the *olfactory bulb.* The number of activated receptors indicates the intensity of the stimulus, and their location in the nose conveys *the nature of the scent.* That is, each scent is expressed by a *spatial pattern* of receptor activity, which in turn is transmitted to the bulb (1991, p. 79).

Again, the research supports the view that the brain imposes categories at a very early stage in sensory processing – even before the signals from the olfactory bulb have reached the olfactory cortex, from which they are broadcast to different parts of the brain. Freeman goes on to describe a subsequent stage in the entorhinal cortex, where the signals are combined with old memories to create a meaning-laden *gestalt,* which is unique to each individual:

> For a dog, the recognition of the scent of a fox may carry the memory of food and expectation of a meal. For a rabbit, the same scent may arouse memories of chase and fear of attack (1991, p. 79).

However, the process of creating a gestalt takes us beyond perceptual categorization, and into the realm of subjective meaning. As such, it is irrelevant to the question we are addressing here: does science support the notion of a distinction between “mere” sensation and a categorising perception that subsequently imposes order on raw “sense-data”? The answer appears to be negative; any philosophical distinction that is drawn between sensation and perception must therefore be a purely logical one.

2.5 The relation of the senses to different varieties of consciousness

In chapter one, I discussed various philosophical definitions of the term “consciousness”, including *transitive creature consciousness*, or the ability to *perceive* and respond to objects, events, properties or facts, thereby making one conscious *of* them. The broad and narrow definitions of “sense” above imply the occurrence of many different grades of transitive creature consciousness in the animal world. As we have seen that creatures with a brain and central nervous system do indeed perceive objects according to categories, we may conclude:

> Conclusion 2.6: All creatures with “true” senses (as defined in section 2.3.4) also possess “transitive creature consciousness”, even if we confine the term “transitive creature consciousness” to those organisms that are capable of perceptually

categorising stimuli.

Recent scientific discoveries (Block, 2001, 2005) suggest that different visual perceptions in the animal brain compete with one another for control of the neurological systems, and that the winner gets to dominate a “global workspace” in the brain, broadcast its information across this workspace, and thereby control the creature’s actions. When one of these representations achieves “rational control of action”, Block describes it as access conscious, where “rational control of action” is defined to include “systems of memory, perceptual categorization, reasoning, planning, evaluation of alternatives, decision-making, [and] voluntary direction of attention” (Block, 2005, p. 47). We saw in section 2.4 that “perceptual categorization” is an integral part of the process of sensing an object: the brain imposes these categories from the beginning. We may conclude that those creatures possessing a central nervous system and endowed with “true” senses, automatically qualify for having what Block calls *access consciousness.* Senses are also necessary for access consciousness: without sensory capacities, the perceptions that compete in the animal brain for dominance and the rational control of action would never occur in the first place.

> Conclusion 2.7: Possession of “true” senses (in the sense defined in section 2.3.4) is both a *necessary* and a *sufficient* condition for the possession of access consciousness.

In chapter one, we remarked on the vomeronasal sensory system, which responds to pheromones and affects human behaviour, but is devoid of phenomenality (Allen, 2003, p. 13). We also noted that in the human brain, the neocortex is divided into *primary and secondary* *regions* (which process low-level *sensory* information and handle motor functions), and the *associative regions*. Brain monitoring techniques indicate that only processes that take place within the associative regions of the cortex are accompanied by conscious awareness; activities which are confined to the primary *sensory* cortex, or processed outside the cortex, are *inaccessible* to consciousness (Roth, 2003, pp. 36, 38; Rose, 2002a, p. 15).

While there are several competing scientific theories as to how primary consciousness arises in the brain, all of the contemporary scientific accounts (summarised in chapter one) agree that it can only occur in organisms possessing senses of some sort. On some accounts (e.g. Edelman, 1989), vision is considered to play a central role in the generation of consciousness, but this remains contentious. I argued in chapter one that scientific findings regarding primary consciousness could be applied to the philosophical investigation of *phenomenal* consciousness.

> Conclusion 2.8: Possession of senses is a necessary but not a sufficient condition for the possession of phenomenal consciousness.

2.6 The relevance of senses to the intentionality of mental states

In chapter one, we discussed three ways of defining intentionality: in terms of "aboutness", propositional attitudes, and conditions of satisfaction respectively.

If the account of sensing which I have defended here is correct, then the act of sensing has a fairly robust kind of "aboutness". It is not just about colours, tastes, smells and so on; rather it is about objects of a specific kind. It seems that all animals with "true" senses (as defined in 2.3.4) are able to recognise objects in this way:

> [T]he recognition of food items, nest-mates, prey, predators, potential mates, etc. seems almost universal among animals including insects …, fish … and crustaceans … Even mollusks learn to recognize odors … (Franklin *et al.*, 2005, online).

I have not been able to ascertain whether sensing in worms has this robust kind of "aboutness". According to De Bono and Maricq (2005, online abstract), the roundworm *C. elegans* can "dynamically respond to dozens of attractive and repellent odors" even though it has only five olfactory neurons. However, this ability to respond flexibly to odors does not entail that the worm can recognise the objects that generate the odors.

Intentionality can also be described in terms of *propositional attitudes.* However, there seems to be no point in construing animals' sensory capacities as *propositional attitudes*, when a simpler *object-based* account is able to explain them equally well. In the above example, it seems natural to say that what the organism senses is simply an *object* of a certain kind, with physical property X, rather than the fact *that* an object of a certain kind is present. There are, to be sure, cases where it seems more plausible to construe the content of an animal's act of sensing propositionally. An oft-cited example from Aristotle's *Nicomachean Ethics* (3.10) is that of the lion which perceives *that* the ox it is about to eat is near. Thus although Aristotle denied beliefs to animals, he allowed that they could have *perceptions* with a *propositional content*.

Perceptions, whether of objects or propositional states of affairs, can still be correct or mistaken. Dretske (1999, p. 27) describes a case where a naïve foraging bird misperceives a Monarch butterfly as "food". Perceptions thus appear to have *conditions of satisfaction* of some sort. Note that unless an animal is able to *categorise* what it senses, there is no meaningful sense in which its act of sensing can be judged to be correct or mistaken.

> Conclusion 2.9: For many animals (insects, molluscs and vertebrates), perception possesses *intentionality,* according to the first and third of the definitions outlined in chapter one: it is *about* objects of a certain kind, and it has *conditions of satisfaction* to at least a limited degree.

2.7 Which intentional stance best describes the senses?

In chapter one, we distinguished between two versions of Dennett’s *intentional stance:* a goal-centred version which explains an organism's behaviour in terms of its information about its goals, and an agent-centred account which invokes beliefs and desires to explain the behaviour. As sensors encode information of some sort, they can certainly be described using Dennett's *goal-centred* intentional stance. The *adequacy* of this goal-centred account becomes apparent when we consider that from a teleological perspective, the primary significance of an organism's ability to sense is that the act of sensing makes it aware of the presence of one of its *built-in ends* (e.g. food), in order that that it can pursue this end. (Senses may also make an organism aware of *means* to an end, but that is not their essential function.) However, there seems to be no predictive or explanatory advantage in invoking *beliefs*, *desires* and *intentional agency* in order to account for the pursuit of an *end* as such; a *goal-centred* stance, which construes the behaviour as information-guided pursuit of a built-in goal, accounts for the behaviour perfectly well.

It is of course true that that in human beings (and some non-human animals), the act of sensing an object will normally give rise to certain *beliefs* about that object. However, the question at stake here is not whether sensing in certain animals is normally accompanied by the formation of beliefs, but whether the mere possession of senses by an organism implies that it is capable of entertaining beliefs or any other mental states.

Another reason for not invoking an agent-centred stance to account for sensory perception is that it invokes *beliefs,* which in turn presupposes that animals are able to form *concepts.* Although we have seen that many animals are quite capable of categorising and discriminating between stimuli, we have not yet discussed whether they can form *concepts.* Although Federmeier (1997) argues that sensing necessarily involves categorisation, she acknowledges that categorisation does not entail having concepts:

> If an organism treats a set of perceptually different items similarly by eating all of them, for example, these items could be said to define the category “food” for that organism. If that organism also has a representation in which “food” is understood to be anything edible, the organism might then be said to have a concept of food. Categories, not concepts, will be dealt with in this paper (1997, p. 3).

Since the act of sensing an object does not entail phenomenal consciousness and can be described adequately in terms of Dennett’s *goal-centred* intentional stance, we may conclude:

> <u>Conclusion 2.10:</u> An organism’s possession of senses *per se* is an inadequate warrant for saying that it is capable of mental states, even if the organism's senses are of the sophisticated kind found only in "higher" animals.

Chapter 3 – Memory

3.1 What is memory?

Microsoft defines memory as “a *device* where information can be stored and retrieved” (Microsoft Help and Support Glossary, 2006, italics mine), while a leading psychology textbook defines it as “the *process* whereby we encode, store and retrieve information” (Feldman, 2003, italics mine). I would like to propose the following general definition:

> Definition - "memory"
>
> The term memory refers to: (a) any *capacity* for encoding, storing and retrieving information; or (b) any *process* of encoding, storing and retrieving information.

This definition has two merits: it distinguishes between memory as a state and memory as a process, and it can be applied to both natural and artificial systems: indeed, it can be said to occur in "any system that has structures ... that can persist for a long time and affect the behaviour of the system" (Wolfram, 2002, p. 823).

In the interests of clarity, we need to distinguish between various kinds of memory occurring in nature, as we did with the senses.

3.2 What kinds of memory exist in nature, and which creatures have them?

3.2.1 The simplest kind of memory: chemical memory

The following account of the simple memory mechanism used by bacteria is pooled from a variety of sources (Illingworth, 1999; Di Primio, Muller and Lengeler, 2000, pp. 4 - 6; Cotterill, 2001, pp. 3-5; University of Utah, 2002).

The simplest kind of memory found in organisms is a kind of chemical memory found in bacteria, which use it to search for food. Because bacteria are too small to detect any changes in the concentration of nutrients from one end of their one-celled bodies to the other, their only way of deciding which way to move in their search for food is to rely on a very short-term memory mechanism, in which they move around randomly, sample the concentrations of chemicals in their environment at regular short intervals, and compare the current concentration of attractant chemicals in their environment with the concentration during the last measurement. A bacterium has four different varieties of sensors, known as methyl-accepting chemically sensitive proteins, or MCPs, on its surface. A bacterial cell can sense a chemical if even one of its sensors comes into contact with a chemical, and it can detect the change if the number of sensors in contact with a chemical increases by just one. The chemical sensors in a bacterial cell signal changes in the attractant concentrations, rather than absolute concentrations of the attractants. If there is an increase in the concentration of attractants, the bacteria will keep moving in the same direction. In other words, bacteria use *temporal* gradients rather than *spatial* gradients to detect food.

> Conclusion 3.1: Even the simplest cellular organisms possess some kind of memory capacity, which enables them to detect changes in their environment.

A bacterium's memory is a consequence of the fact that its tracking system takes a few seconds to catch up with any alteration in chemical concentrations, enabling the bacterial cell to compare its present state with its state a short time ago. The number of receptors stimulated by attractive or repellent molecules (apparently this number is an average of measurements taken over a period of about one second) is "compared" with the number of receptors stimulated in the previous measurement (stored as an internal signal representing the average of measurements taken 3 seconds ago). Cotterill (2001) describes the process:

> ... the creature's recent history ... determines the instantaneous magnitudes of the various molecular concentrations. No chemical record is kept of the magnitudes of the various concentrations at different times... [T]he information in its environment concerning the spatial distribution of nutrients... is merely lumped into a single number, and the bacterium's cognitive repertoire is telescoped into a single binary choice, clockwise or anti-clockwise rotation of the flagellum (Cotterill, 2001, p. 22).

Bacterial memory is thus constrained in four significant ways, since it is:

- *minimal* (it can store just one set of intermediate results, allowing bacteria to remember any changes in the concentration of attractant chemicals that have occurred in the past 3 seconds);
- *binary* (the bacterium simply “chooses” whether to continue tumbling randomly or keep moving in its present direction);
- *relative* (the bacterium does not remember absolute concentrations of attractants, but changes in concentrations); and
- *ahistorical* (no chemical record is kept of the magnitudes of the various concentrations at different times - instead, the bacterium simply compares its present circumstances with its situation a few seconds earlier).

Leaving aside the chemical memory of bacteria, there are many different kinds of memory found in human beings. The following diagram (taken from Franklin S., Baars B., Ramamurthy U. and Ventura M. 2005) provides a useful overview of the varieties of memory systems found in human beings.

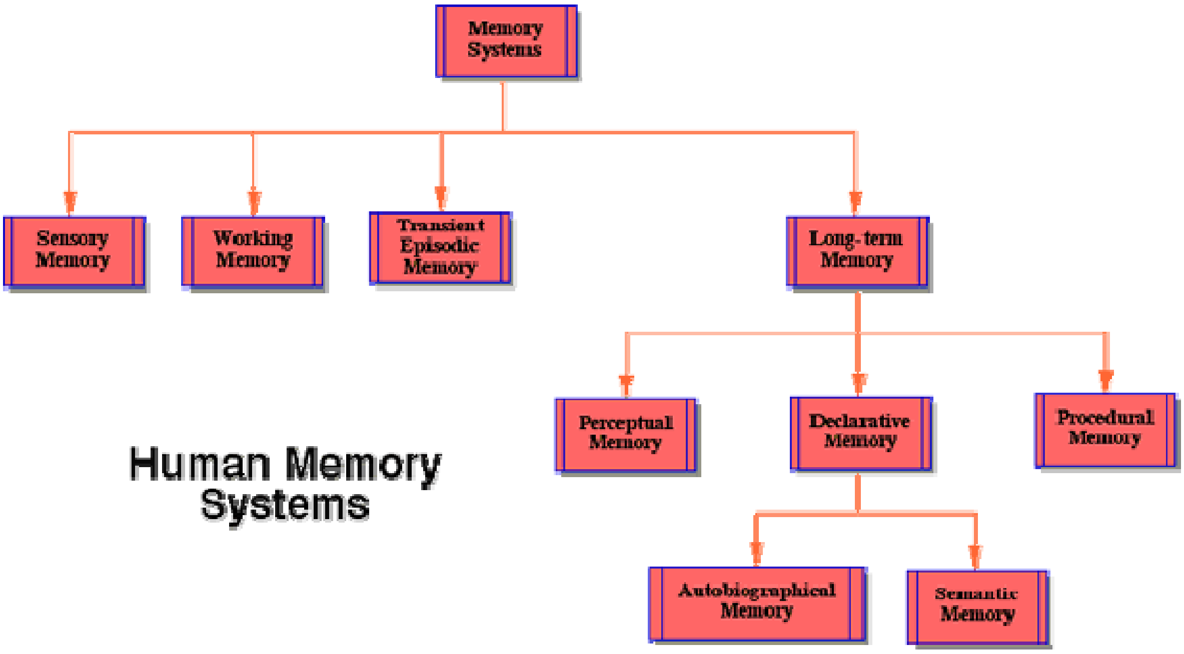


Diagram 3.1: Human Memory systems (Franklin S., Baars B., Ramamurthy U. and Ventura M., 2005). Note that autobiographical memory (or memory for events having features of a particular time and place) is often referred to in the literature as *episodic* memory. Most of the categories shown here are commonly recognised; a few (sensory memory, perceptual memory, transient episodic memory) are recognised by some experts but not others.

Franklin et al. (2005) distinguish three forms of memory from long-term memory:

> *Sensory memory* holds incoming sensory data in sensory registers and is relatively unprocessed. It provides a workspace for integrating the features from which representations of objects and their relations are constructed. There are different

> sensory memory registers for different senses…Sensory memory has the fastest decay rate, measured in hundreds of milliseconds…
>
> *Working memory* is the manipulable scratchpad of the mind (Miyake and Shah 1999). It holds sensory data, both endogenous (for example, visual images and inner speech) and exogenous (sensory), together with their interpretations. Its decay rate is measured in seconds. Again, there are separate working memory components associated with the different senses…
>
> An unusual aspect of the IDA model is *transient episodic memory* (TEM), an episodic memory with a decay rate measured in hours… (Franklin et al., 2005, online).

As these are short-term forms of memory, I shall not discuss them further here. In human beings, these forms of memory are typically conscious; however, we cannot assume that the same holds true for creatures that fail to satisfy the neurological conditions described in chapter 1. Below, I address the question of whether the occurrence of *memory* in other organisms warrants the ascription of mental states to them.

3.2.2 Procedural (Non-declarative) Memory

Roediger, Marsh and Lee (2002) provide a useful summary of the current literature relating to memory. The distinction proposed by Ryle in 1949 between *procedural memory* and *declarative memory* - roughly, between "knowing how" and "knowing that" - is still widely invoked, although some recent authors refer to the former simply as non-declarative memory. The distinction is not as clear-cut as one would like: it seems that most if not all memory tasks have some procedural component, although some procedural tasks (e.g. tying one's shoe-laces) require no declarative component (Roediger, Marsh and Lee, 2002, pp. 5-6). Nevertheless, procedural memory and declarative memory appear to be governed by different mechanisms – shown by the fact that there are patients with impaired declarative memory, whose procedural memory is completely intact (Okano, Hirano and Balaban, 2000, p. 12403) – as well as different brain circuits. Both forms of memory are dependent on frontal cortex activity; however, declarative memory also critically depends on structures in the medial temporal lobe, whereas procedural memory does not (Tamminga *et al.,* 2000, p. 162). The celebrated case of H.M. illustrates this point particularly well. Surgery was performed on him in 1953 to treat his epilepsy, which involved the removal of most of his medial temporal lobes – a region of the brain which includes the hippocampus. Since that time, he has lost the ability to lay down declarative memories, but he can still form procedural memories (Washington University School of Medicine, 1997). On the other hand, the cerebellum is considered to be important for procedural memory (Okano, Hirano and Balaban, 2000, p. 12403). Recent evidence that cerebellar atrophy in epilepsy affects procedural memory (Hermann *et al.,* 2004) supports this hypothesis. We may conclude:

> Conclusion 3.2: The distinction between procedural (non-declarative) and declarative memory – "knowing how" versus "knowing that" – appears to be a fairly robust one.

The term "procedural memory" remains poorly defined in the literature. For instance, Feldman (2003) defines it as “memory for skills and habits, such as riding a bike or hitting a baseball”, Tamminga *et al.* (2000) observe that it “includes the ability to learn motor skills”, and the McGraw-Hill Dictionary of Scientific and Technical Terms (2003) defines it as “the memory of motor, perceptual, and cognitive skills". The common item listed in these definitions is "memory for skills", but the term "skill" is a vague one. Roediger, Marsh and Lee (2002, p. 5) give a more precise definition: procedural memory includes capacities such as *classical conditioning*, *motor skill learning* and *complex* (skill-based) *problem-solving*. As classical conditioning (which I discuss in chapter five) is the most widespread of these capacities in the animal kingdom, I shall treat it as a si*ne qua non* for the possession of procedural memory.

In evolutionary terms, procedural memory is thought to be very ancient, as it is shared in some form by virtually all animals (Tulving, 1985), although I have not been able to locate any evidence for its occurrence in sponges. If, like some authors (Roediger, Marsh and Lee, 2002), we choose to regard classical conditioning and motor skill learning as the most primitive forms of procedural memory, then its occurrence in cnidaria (coelenterates) also appears unlikely, for reasons that will become apparent in chapter 5. There have been claims that plants and protoctista are capable of undergoing classical conditioning (see Hennessey et al., 1979, pp. 417-423; see also Abramson et al., 2002, for a careful evaluation of claims of classical conditioning in plants), which would imply that they possess a primitive form of procedural memory. I discuss these claims below in chapter five, where I conclude that they have not yet been adequately substantiated, and that for the time being, it would be prudent to regard classical conditioning as confined to animals with central nervous systems.

> <u>Conclusion 3.3:</u> Worms appear to be the most "primitive" creatures with a rudimentary form of procedural memory. Procedural memory is confined to creatures with central nervous systems.

3.2.3 Perceptual Memory

Some authors (e.g. Franklin et al., 2005) draw a distinction between *perceptual memory*, which plays a role in the recognition, categorisation and interpretation of stimuli, and declarative memory. Defined in this way, perceptual memory appears to be very ancient:

> We speculate that perceptual memory is evolutionarily older than … declarative memory. The functions of perceptual memory, the recognition of food items, nest-mates, prey, predators, potential mates, etc. seems almost universal among animals including insects…, fish… and crustaceans… Even mollusks learn to recognize odors (Franklin et al., 2005, online).

3.2.4 Declarative Memory

Tulving (1972) has suggested that there are two forms of declarative memory: an *episodic memory* for particular events (e.g. "Where did you go on vacation last summer?"), which involves accessing the time and place of their occurrence, and a *semantic memory* for general facts about the world (e.g. zebras have four legs). This distinction is widely accepted in the literature (see Roediger, Marsh and Lee, 2002). However, the neurological basis of this distinction remains shaky, as there are differing schools of thought as to what extent semantic memory employs the same brain systems as episodic memory.

The assessment of declarative memory in animals is a challenging task, as they cannot verbally declare what they remember. The procedure normally used by scientists to assess animals is *delayed non-matching to sample* (DNMS) tasks, in which animals are presented with a sample object and then after a delay interval are asked to choose between the sample object and a novel object. Choosing the novel, non-matching object is the correct choice. The demonstrated ability of primates, rats, pigeons (Young and Wasserman, 2001) and even honeybees (Giurfa et al., 2001) to perform these tasks indicates that they possess some form of *semantic* memory.

Giurfa (2005) has argued that mastery of such tasks requires the ability to classify stimuli according to a *rule:*

> In rule learning, the animal bases its choice, not on the perceptual similarity between the novel and the known stimuli which may not share… any common feature, but on links that transcend the stimuli used to train it…. An example of rule learning is the learning of the so-called principles of sameness and of difference…In DNMS, the animal has to learn the …[rule],… '*choose always the opposite to what is shown to you (the sample)*' (Giurfa, 2005, online).

Episodic memory is considered to be the most recent form of memory in evolutionary terms, and its occurrence in non-human animals is still disputed. There is evidence that western scrub jays possess an episodic-like memory, but at the present time, the alternative hypothesis that some form of semantic memory can account for the birds' behaviour cannot be ruled out (Shettleworth, 2001; Clayton et al., 2003; Emery and Clayton, 2004). Recently, the case for episodic memory in non-human animals has been bolstered by the discovery (University of Edinburgh, 2006) that Canadian rufous hummingbirds, whose brains are no larger than a grain of rice, possess two abilities which are peculiar to episodic memory – they can remember both the locations of food sources and when they visited them. The hummingbirds in the study were able to recall where certain flowers were located and how recently they were visited. Scientists tracked how often hummingbirds visited eight artificial flowers filled with a sucrose solution in the birds' feeding grounds. They refilled half the flowers at 10 minute intervals and the other half 20 minutes after they had been emptied. The birds' return to the flowers matched the refill schedules: flowers refilled at 10-minute intervals were visited sooner. Susan Healy, of the University of Edinburgh's School of Biological Sciences, commented: "We were surprised that their timing abilities were so good and that they managed to cope so efficiently with as many as eight different flowers" University of Edinburgh, 2006, online).

Significant as these recent findings are, they should not be overstated. Episodic memory is also referred to in the literature as "autobiographical memory", but the memory displayed by the hummingbirds can hardly be termed "autobiographical". The birds' sense of time (do they remember absolute or relative intervals?) also needs to be investigated, as well as the regions of their brains which are activated when retrieving information relating to events in the past.

> Conclusion 3.4: At least some mammals, birds and insects possess a form of semantic memory. It is too soon to say whether any non-human animals possess episodic memory.

3.3 Memory and Consciousness

Although the acquisition and refinement of a skill often involves conscious effort, the learning of a new skill can certainly take place in the absence of phenomenal consciousness. Grau (2002) reports a case in which rats whose spinal cords had been severed at the second thoracic vertebra (T2), leaving them paralysed below their mid-sections, learned to avoid electrical shocks with their severed spinal cords. Additionally, the regions of the brain which are critical for the formation of procedural memories – namely, the cerebellum and basal ganglia – are generally thought to lie outside the brain's generator mechanism for *primary consciousness* (refer to the discussion in chapter one).

I know of no case in the literature of a person learning to follow a rule (such as the same-different rule used to verify the occurrence of *semantic memory* in animals) while unconscious. One might imagine that something akin to *subliminal learning* is occurring when an animal learns the same-different rule. While studies (Berridge, 2001, 2003a, 2003b) indicate that we can acquire information without being consciously aware of it, Franklin et al. (2005) counter that "[a]lthough subliminal acquisition of information appears to occur, the effect sizes are quite small compared to conscious learning." Also, the passive acquisition of information is quite different from the explicit choice involved in adhering to a rule. The behaviour involved here, then, seems to require the occurrence of phenomenal consciousness; yet this seems to be at odds with the discovery that even honeybees are capable of *semantic memory* (Giurfa et al., 2001). The paradox of the honeybee brain is that while it is very small (1 cubic millimeter, less than 1 million neurons), it supports an impressive array of complex behaviours (Giurfa, 2003). The neurological consensus is that the honeybee brain could not support primary consciousness (Edelman, personal email, 19 July 2004). There are two possibilities here: either the bees are not really following a rule, but doing something simpler (hence no phenomenal consciousness is required); or they are following a rule, unconsciously. Further research needs to be done here.

> <u>Conclusion 3.5:</u> Procedural memory can occur in the absence of phenomenal consciousness. The semantic memory found in certain animals (mammals, birds and

honeybees) is strongly suggestive of phenomenal consciousness, but as bees lack the neurological wherewithal for primary consciousness, no firm conclusions can be drawn as to whether the occurrence of semantic memory in other animals indicates phenomenal consciousness.

3.4 The Intentionality of Memory

In chapter 1, we saw that the philosophical term "intentionality" could be variously defined in terms of aboutness, propositional attitudes and normativity. Each of the different kinds of memory could be described as being *of* or *about* something – be it a skill (as in *procedural* memory), an object in one's environment, such as an item of food, a nest-mate, a predator, a potential mate or a potential victim of prey (as in *perceptual* memory), a fact or rule that one has learned (as in *semantic* memory); or an occurrence in one's life (as in *episodic* memory). Of course, it would be a mistake to assume that all of these kinds of memory instantiate "aboutness" to the same degree. For instance, suppose that Jane has a detailed recollection of the time when she learned to ride a bike, while Tom knows how to ride a bike but has no memory of learning how to, as a boy. Jane's episodic memory possesses a much richer kind of "aboutness" than Tom's procedural memory; nevertheless, Tom's memory is clearly *about* something: a practical skill which he acquired in his youth and which he can still execute proficiently.

Of the forms of memory we have examined here, only semantic memory seems to presuppose *propositional attitudes.* The memories for *skills* which are stored in procedural memory are notoriously difficult to express in words; perceptual memory encodes memories of *objects* in an animal's environment, rather than propositions; and episodic memory holds memories of *events*, although its contents may be described propositionally if the individual possessing the memory is capable of language.

Memories also possess a kind of *normativity;* their conditions of satisfaction may fail to obtain. This is especially obvious if we consider procedural memory: one may think that one has properly memorised a skill like driving a car, and then fail to execute it correctly when the occasion demands, with embarrassing consequences. Likewise, a perceptual memory is correct if and only if the object recalled by the subject was indeed the object that he/she perceived.

3.5 Does a creature need to exhibit memory before we can justifiably impute mental states to it?

One might argue that even a creature that was utterly unable to form memories could still have mental states, as we can perfectly well conceive of an organism with conscious feelings but no memories. This objection is irrelevant: as we argued in the Introduction, conceivability does not imply possibility. Even if it were physically possible that phenomenally conscious creatures lacking in memories could exist in our world, this would leave us with the problem of how we could know that they had mental states. The question we should ask instead is whether we could ever be justified in using an agent-centred stance to describe the behaviour of an organism with no memory capacity.

In any case, supposed cases of individuals who have lost all their memories, but still have beliefs and desires, turn out to be nothing of the sort. If we look at the most celebrated example in the psychological literature, the individual known as H.M., it turns out that although he has completely lost his episodic memory of events since his operation, and his semantic memory appears to be frozen in the 1950s (with a few exceptions), his procedural memory remains intact (Rorden, 2005). Likewise, patients with Korsakoff's syndrome are incapable of forming declarative memories but retain the ability to form procedural memories.

To see why the occurrence of memory in an organism is an essential condition for justifiably attributing beliefs and desires to it, I would invite the reader to consider the question of whether there could be an organism whose beliefs were *only* about its *biological ends* - in other words, an organism that was capable of entertaining beliefs of the "food-here-now" variety, but unable to form *new* beliefs of the “press-button-get-food” variety. At first blush, it might seem that the answer is yes: after all, such an organism would certainly be capable of exhibiting behaviour which manifested its desires for its own built-in biological ends, as well as its beliefs about those ends.

However, such a description would be scientifically redundant: there would no way in principle of behaviourally distinguishing an organism which only has beliefs about its *ends* from an organism *lacking* beliefs and possessing only built-in goals, as neither organism (by definition) is capable of forming beliefs about new and better *means* of achieving its ends. A goal-centred intentional stance could account for its behaviour perfectly well.

Beliefs about ends ("food-here-now") are thus unintelligible in isolation; their ascription is redundant unless the individual is capable of learning new ways of obtaining the objects of its desire - in other words, new and useful skills. (The fact that some skills acquired confer no benefit on their possessor in no way weakens the point being made here, that the teleological purpose of memory is the acquisition of useful skills.)

The case we have been considering is a hypothetical *species* of goal-seeking organism that *never* manifests a capacity to remember at *any* stage of its development. Such an organism lacks what might be called a "natural capacity" to form memories: that is, the internal program that directs its biological development does not encode for the creation of structures with a memory capacity. However, if there existed an *individual* possessing mental states, who completely lost its capacity to remember, due to physical deterioration or trauma, but displayed a rich behavioural repertoire consistent with its holding beliefs and having desires, then we might be inclined to say that the impaired individual still possessed some mental states, despite losing its memory.

What the foregoing argument implies is that imputing beliefs and desires to an organism presupposes that it possesses some form of *procedural* memory. This is a strong conclusion, and one might ask why this particular kind of memory should be so fundamental. Why couldn’t there be an organism with perceptual memories but no procedural memories, for instance?

I contend that procedural memory is the most fundamental form of memory relating to agency, and that the intrinsic finality of other forms of memory is derivative upon that of procedural memory. None of the information (whether generic or tied to a specific time and place) that an individual remembers will benefit it unless it is capable of varying its patterns of behaviour in a way that promotes its own well-being. *To do this, it must acquire new skills.*

Thus the major cognitive limitation of an organism lacking memory capacity is that it would be unable to acquire new patterns of behaviour, as new patterns would have to be stored or encoded somewhere in the organism.

Conclusion 3.6: There can be no scientific or philosophical justification for attributing beliefs and desires to an organism lacking procedural memory. In other words, procedural memory is a necessary condition for the attribution of mental states.

Conclusion 3.7: As there have been no credible claims that prokaryotes (bacteria and archaea) possess any kind of procedural memory, we can assume that they do not have beliefs or desires.

3.6 Which intentional stance is appropriate for describing the different kinds of memory found in nature?

At very least, a creature possessing *semantic* memory must be capable of *following a rule* (such as distinguishing “same” from “different”). The act of following a rule would seem to presuppose an agent-centred intentional stance. A *goal-centred* stance appears to be inadequate to explain rule-following, as Giurfa (2005) has argued, the goal itself is not one of the stimuli, or even something associated with the stimuli, but the mastery of a *principle* (e.g. choose the pair that is different), which *transcends* any particular stimulus; only if one has mastered this principle can one use it to obtain the reward.

As the number of animals exhibiting semantic memory is relatively small (mammals, birds and bees), and my thesis deals with the simplest kind of mind an animal could have, I do not propose to discuss semantic memory further here. The question we need to ask here is: if a creature possessed only procedural memory, could we automatically infer that it had mental states?

From a scientific standpoint, there appears to be no reason why an organism's acquisition of new skills should be better accounted for in terms of its beliefs and desires, rather than the information it possesses in order to achieve its goals.

> Conclusion 3.8: Procedural memory, taken by itself, does not warrant the attribution of mental states to an animal.

The upshot of our enquiry into memory is that the existence of a memory for new skills in an organism is a necessary but not a sufficient ground for ascribing cognitive mental states to it.

Chapter 4 - Flexible Behaviour and its Significance for Mental States

4.1 Fixed and flexible patterns of behaviour in organisms: an overview

4.1.1 What is flexible behaviour? A proposed definition

A capacity for *flexible behaviour* in an organism is often regarded as a condition that must be satisfied before we can justifiably impute mental states to it. For instance, Gould (2002, p. 41) writes that “[t]o most minds, cognition implies an ability to step outside the bounds of the innate”. It is of course true that fixed patterns of behaviour may be accompanied by mental states in an organism. For instance, the sight of a predator may trigger a number of fixed behaviour patterns in its prey, and these patterns may be automatically accompanied by a *belief* that a predator is present, a *fear* that the predator will attack, and a *desire* to escape from the predator. But the question at stake here is not whether fixed patterns of behaviour may be accompanied by mental states, but whether fixed patterns are sufficient *by themselves* to warrant the ascription of mental states to animals.

Unfortunately, the terms “fixed” and “flexible” are rarely defined in the philosophical literature. I would like to propose the following simple definition, which was inspired in part by Kilian and Muller (2002, p. 3) who argue that while unicellular organisms appear to respond flexibly to a wide range of attractant chemicals, their response is actually quite rigid, as the behaviour of each of their constituent molecules instantiates the properties of *functional specificity* and *goal specificity.*

The underlying thinking in the definitions which follow is that we can describe the behaviour of an organism in terms of some mathematical *program*, which contains numerous routines and subroutines (which can be mathematically represented as computational *procedures* and *functions).* These routines might be considered as patterns of behaviour. Each function and procedure has one or more *input variables* and *output variables.* Mathematically, each *pattern of behaviour* is equivalent to a *rule* for transforming input variables into output variables. We can thus describe a pattern of behaviour as *fixed* if it always transforms inputs into outputs according to the same rule, and as *flexible* if the transformation *rule* changes over time.

> Definition - "fixed pattern of behaviour"
>
> We can mathematically represent a *pattern of behaviour* in an organism by an *output variable* (say, z). A *fixed* pattern of behaviour can be defined as a pattern where the value of the *output* variable z remains the same, given the same values of the *input* variables.

(Note: the term “fixed pattern of behaviour” is not to be confused with a similar term, "fixed action pattern", which is used in a technical sense by ethologists to denote behaviour which continues in the absence of a stimulus.)

It should be clear from the above definition that *fixed* patterns of behaviour constitute only a *subset* of all possible instances of *programmed* behaviour. An organism's behaviour may be programmed, but the program may *change* over time - whether because of new patterns of output (i.e. new *functions),* new input and/or output variables that are invoked by existing functions and procedures (i.e. new *parameters)* or new conditions under which the output patterns are generated (new prog*ram instructions).* Such changes constitute what I would define as "flexible behaviour".

Definition - "flexible behaviour"

If the program governing an organism's behaviour changes over time, such that the value of an output variable z is no longer the same for the same inputs, whether because of a change in the function(s) which define the value of z, or the parameters of the function(s), or the conditions in the program under which the function(s) are invoked, then the behaviour described by z is flexible.

One surprising implication of the above definition of a *fixed* pattern of behaviour is that even in a fixed pattern, the value of the output variable z may be determined by two or more different functions, depending on the values of the inputs. A simple case would be the following program statement, written in Pascal code:

```
IF (x > 4) THEN z := F(x)
ELSE
BEGIN
        IF (x > -2) THEN z:= (F(x) + G(x))
        ELSE z := G(x)
 END;
```

In the list of functions, we might define F(x) as, say, x + 3 and G(x) as (x - 5) / 2. Here we have two different functions (F and G) being invoked for different values of the input variable x. Some writers might consider this to be an instance of flexibility, but according to my definition it is nothing of the sort: *over the course of time,* the value of the output variable z remains the same for any given value of the input parameter x. If the value of x changes from 5 to -7, the function determining the value of the variable z changes from F(x) to G(x), but because the *program* has not changed, we can still describe the overall pattern as fixed.

So what would a *flexible* pattern of behaviour look like in this case? For the Pascal statement above, a subsequent program change in:

(i) the IF statement conditions (e.g. from (x > 4) to (x > 5)), or

(ii) the *definition* of the function F (e.g. from F(x) = x + 3 to F(x) = x + 5) or the function G, or

(iii) the nu*mber of parameters* F and G require,

would qualify as an instance of *flexible* behaviour. Finally, it should be noted that this definition deliberately sidesteps the question of *how* the program change occurs.

Of course, both of the above definitions assume that the change in pattern of behaviour, if it occurs, takes place *within the same individual.* Over the course of time, evolution may modify information pathways in a *lineage* of unicellular organisms. Mutations that alter DNA or RNA may give certain individuals a selective advantage, bringing about changes in the genetic programs regulating their behaviour, but no there is no flexible behaviour on the part of any given individual. Kilian and Muller (2002) characterise this as evolutionary adaptation or "phylogenetic 'learning'" (2002, p. 3), which they distinguish from *true* learning, on the grounds that the changes in behaviour are not "individually acquired during the life-time of the cell" (2002, p. 2).

In chapter one, I argued that mental states could only be meaningfully ascribed to an entity with a *telos,* or good of its own. Although we can speak of a lineage of organisms as thriving or dying out, it does not, properly speaking, have a *telos,* as there is no co-operation of parts working together for the good of the whole here, as occurs within an individual organism. (On the contrary: the "parts" of a lineage are the individuals which make it up; if two of these individuals happen to exist contemporaneously, they will *compete* rather than co-operate.) I would therefore agree with Kilian and Muller's (2002) contention that true learning cannot be meaningfully ascribed to lineages of organisms, but only to individuals.

4.1.2 Different kinds of fixed behaviour distinguished by scientists

The terms "innate", "instinct", "drive", "reflex" and "motivation" are often used with a lack of precision, so a few clarifications are in order. Thorpe (1956) distinguishes three different senses of the word "innate":

> [T]he word 'innate' may imply any or all of the following: (1) inherited or genetically fixed and therefore characteristic of the species; (2) internally co-ordinated; (3) internally motivated. Both instinct and reflex may be innate in senses (1) and (2). Only instinct has an internal drive or motivation in sense (3) (Thorpe, 1956, p. 14).

Staddon (1983) cautions against the naïve tendency to regard "innate" and "learned" as simple opposites:

> [S]heep dogs are selected for their ability to herd sheep; and training a professional sheep dog takes years. Nevertheless, components of herding, such as circling and attempting to group people or animals, appear spontaneously in pets that have never

> been specifically trained… Young male swamp sparrows … learn their song effectively, but are quite selective about what they will accept as a model. In all these cases, the *learned* ability *is put together with* ingredients provided *innately* (1983, p. 12, italics mine).

Additionally, Staddon points out that "[v]ery complex behavior may be innately programmed, and complexity by itself is an unreliable guide as to whether something is learned or innate" (1983, p. 12).

Epstein (1982) discusses the similarities and differences between the concepts of *instinct* and *motivation.* Both include innate mechanisms as well as acquired components, display a sequential organization of drive-induced appetitive and consummatory phases, and contribute to homeostasis (maintaining a stable internal state). The key differences can be summarized as follows: *instinct* is: (i) species-specific in its sensorimotor organization; (ii) taxonomically common; (iii) non-affective (not accompanied by emotion); and (iv) not modified by expectancy in its appetitive phase. *Motivation,* on the other hand, is: (i) individuated (varies across individuals); (ii) taxonomically uncommon; and (iii) includes the anticipation of affect (emotion).

In ethology, instincts are also differentiated from reflexes. *Reflexes* (such as the patellar reflex) are normally envisaged as fairly rigid responses triggered by specific stimuli; while *instincts* (e.g. nest-building in certain species of birds) are more complex processes motivated at higher levels by suites of internal and external factors, which give rise to an array of patterned output.

The term "drive" was introduced into psychology by Woodworth (1918), who envisaged it as a homeostatic force that instigates action to restore equilibrium, after disruptions or disequilibria in basic needs. The term is also used to refer to bodily states that instigate action and underlie reinforcement of learned responses. For an up-to-date overview of motivational concepts in behavioural neuroscience, see Berridge (2004).

Finally, scientists use the following terms to refer to specific patterns of movement in microbes or plants: a *taxis* is defined as a movement of a cell in response to a stimulus, a *tropism* is the directional growth of a plant organ in response to a stimulus such as light, water, touch or gravity; while a nastic movement *(nasty)* is a movement of a plant organ in response to stimuli, that is independent of the direction of the stimuli (e.g. the opening of flowers in response to changes in light) (Isaacs, Daintith and Martin, 1999).

4.1.3 Examples of so-called "flexible" patterns of behaviour cited by philosophers, which turn out to be fixed

Suppose that we can describe a piece of behaviour in an organism using a mathematical function F and some input variables (or parameters) x_1, x_2, x_3, ... x_n, where the value of the output variable z is $F(x_1, x_2, x_3, \ldots x_n)$. It follows that even when the values of x_1, x_2, x_3, ... x_n vary over time, the behaviour still conforms to a fixed pattern, so long as the function F remains the same. Thus the mere fact that the output variable z of a function F has different values under different circumstances in no way implies that z's behaviour is flexible.

Nevertheless, some authors (Godfrey-Smith, 2001; Carruthers, 2004a) refer to this kind of output variability as "flexible" behaviour, as the following three examples illustrate.

Case study 1: The lac operon system in bacteria

Godfrey-Smith (2001) claims that "plants and bacteria do exhibit some capacities for flexible response to environmental conditions, using environmental cues to control development and metabolism." As an example, he cites the *lac operon* system in *E. coli* bacteria:

> These bacteria can respond to a change in local food type through processes in which the availability of a nutrient affects the regulation of genes which code for enzymes able to digest that nutrient (2001, p. 6).

The *lac operon* allows *E. coli* to use lactose as an energy source, and break it up into its constituent sugars: galactose and glucose. An operon may exist in one of two regulatory states: ON or OFF. The *lac operon* is subject to positive and negative forms of gene regulation. The operon's default state is OFF, but the presence of lactose induces the genes to turn ON (negative regulation). However, if there is glucose in the environment, the *lac operon* is not expressed, as bacteria prefer glucose to lactose as a source of food (positive regulation) (see Bridges, 2002 for more detail).

From the foregoing description, it should be clear that we are not dealing with flexible behaviour here, but with behaviour which can be described by an invariant (fixed) mathematical function, whose input variables are the local concentrations of different nutrients (lactose and glucose).

Case study 2: Plants' ability to adjust to changes in lighting

Another example of flexibility cited by Godfrey-Smith (2001) is the ability of plants to adjust to different lighting conditions:

> For example, many plants can determine not just that they are being shaded, but that they are being shaded by other plants. This is done by detecting the wavelength properties of the reflected light. The plants respond to shading by growing longer shoots... (2001, p. 7).

Here, the length of the shoots can simply be represented as an invariant (fixed) mathematical function of the wavelength(s) of the incoming light.

Case study 3: Singing strategies in crickets

Carruthers (2004a), while rejecting (rightly in my view) a mentalistic interpretation of the singing behaviour of crickets, describes it as flexible. According to my definition, such behaviour is fixed:

> It turns out that even flexibility of behavioral strategy isn't really sufficient for a creature to count as having a mind, indeed. For innate behavioral programs can have a conditional format. It used to be thought, for example, that all male crickets sing to attract mates. But this isn't so; and for good reason. For singing exposes crickets to predation, and also makes them targets for parasitic flies who drop their eggs on them. Many male crickets adopt the alternative strategy of waiting silently as a satellite of a singing male, and intercepting and attempting to mate with any females attracted by the song. But the two different strategies aren't fixed. A previously silent male may begin to sing if one or more of the singing males is removed... (2004a, p. 12, online PDF version).

Here, the output behaviour for a male cricket (wait or sing) can be described as an invariant function of two variables: (a) its "attractiveness rating" and (b) the number of attractive males nearby.

4.1.4 Historical variables: even a pattern of behaviour that changes regularly over time can still be fixed

A naïve reading of my definition of flexible behaviour may give the misleading impression that any behaviour in an individual, where the *pattern* changes over time, automatically qualifies as flexible. In fact, however, the definition states that only if the *transformation rule* changes over time is the behaviour flexible. Moreover, nothing in my definition precludes us from using *time-dependent historical variables,* or even derivatives such as rates of change, as *inputs* to the value of the function determining the output variable. Thus on my account, a pattern of behaviour that varies over time may still be quite inflexible, if the underlying transformation rule remains invariant.

Case study 4: bistable switches in bacteria

According to Wolf and Arkin (2003), bacteria exhibit a wide variety of motifs which regulate cell activity: switches, amplitude filters, oscillators, frequency filters, noise filters and amplifiers, combinatorial logic, homeostats, rheostats, logic gates and memory elements (2003, pp. 125-126). Perhaps the most interesting of these behavioural patterns are the *bistable switches* in bacterial cells, which have a *memory* and exhibit a *history dependence* (known as *hysteresis):* their pattern of responding to variations in the strength of a signal depends on the initial setting of the switch, and they tend to react slowly to changes in the signal value. There is one pattern of responding when the signal *increases* in strength (a "going-up" pattern), and another when it *decreases* (a "coming-down" pattern). In each pattern, the switch tends to remain in its current setting until the signal reaches a certain level that makes it suddenly "flip" settings, over a narrow range. These two patterns make up what is called a *hysteresis loop* - a switching pattern commonly found in ferromagnetic materials. Bistable switches can be set to an "on" or "off" position indefinitely by an environmental stimulus encountered by the cell.

Certainly, the behaviour of a bistable switch can be described using Dennett's *intentional stance.* It remembers its setting, it resists changes and "tries" to "ignore" random noise within the cell that would cause it to flip continually backwards and forwards between on and off states. (This conservative trait is adaptive, as it cuts out disruptive interference.) Most impressively, it is apparently capable of *flexibility:* it can change from one pattern of responding (the "going-up" pattern) to another (the "going-down" pattern). Does this change of patterns qualify as *flexible behaviour*, according to our definition? I would argue that it does not. I would argue that instead of saying that the switch acquires a *new pattern* as its setting fluctuates, we could more economically describe the "going-up" and "coming-down" patterns as part of a *single pattern* (the hysteresis loop) which is built into the chemistry of the switch. The value of the output ("on" or "off") can be defined a function of *two variables:* the strength of the *current* input signal and that of the *previous* input signal. Together, these two pieces of information tell us whether the signal is "going-up" or "coming-down". Hysteresis in bacterial cells is a *time-lag phenomenon,* rather than a true case of flexible behaviour.

4.1.5 Gene-swapping in bacteria: a *bona fide* case of flexible behaviour

Gene swapping between individual bacteria can be regarded as an adaptive behaviour, because it informs other bacteria about what is going on, allowing them to adapt to unexpected environmental challenges like toxic mercury. Some bacteria have genes which make them resistant to mercury, as it is a naturally occurring toxin. The most widely studied and sophisticated mechanism for resistance to mercury works by bacteria *exchanging transposons, or autonomous mobile sections of their DNA.* Some transposons contain *genes* that confer resistance to mercury, by coding for specialised proteins and enzymes. Bacteria that have (or acquire) these genes can take highly toxic mercury ions into their cytoplasm (cell body) using their specialised carrier proteins, and transfer them to a specialised enzyme. The enzyme reduces ionic mercury to metallic mercury, which is relatively inert and non-toxic, and readily diffuses out of the cell (Barkay and Wagner-Dobler, 2005; Petkewich, 2002).

Certainly, there can be no doubt that the bacteria's changing response to mercury is an instance of truly *flexible behaviour.* The acquisition of new genes corresponds to a *change in the program* statement describing the organism's response to mercury, as well as the acquisition of *new functions* - i.e. new recipes for making specialised proteins and enzymes to handle the toxin. It appears that such gene swapping is a common occurrence among bacteria: for instance, genes that confer resistance to antibiotics can be passed from one species of bacteria to another. In fact, gene swapping appears to be a universal trait of organisms, as illustrated by the frequency of *lateral gene transfer* between different branches of the tree of life (see Forterre and Philippe, 1999). We can therefore regard the following conclusion as established:

<u>Conclusion 4.1:</u> All organisms exhibit flexible behaviour, to some degree.

<u>4.2 Philosophical evaluation</u>

<u>4.2.1 Do fixed patterns of behaviour warrant the ascription of mental states to organisms?</u>

Di Primio, Muller and Lengeler (2000) cite the *lac operon* system in bacteria as evidence of primitive cognition on their part:

> The coupling between stimulus and response in bacteria is *indirect* because when a sensor detects a chemical, it activates a chain of chemical reactions, each of which is reversible. The coupling is *modifiable:* if *E. coli's* sensors detect an attractant (e.g. galactose), and later sense another compound (e.g. glucose) that is more attractive than the first one, a "weighing" of the relative quality of the nutrients occurs, and the chain of reactions resulting in directed motion is amplified. The co-presence of attractants and repellents in solution generates an integration of the "run" and "tumble" responses, at the chemical level (so-called *"conflict resolution")* (2000, pp. 4, 7, italics mine).

The suggestion that *modifiable* behaviour constitutes a sign of cognition trades on an equivocation in the meaning of the word "modifiable": it may mean "responsive to external changes", or it may mean "not governed by a fixed rule or pattern". The word "modifiable", understood in the former sense, has no mentalistic connotations. Only in the latter sense does it imply the kind of flexibility that might indicate the presence of cognitive mental states. But as Kilian and Muller (2002) point out, the way in which bacteria react to a chemical is typically *inflexible,* at the molecular level:

> In unicellulars, in each of the molecules of an information transfer path, regardless out of how many elements it is composed, both a functional specificity and a goal specificity can be discerned. Each molecule contacts its goal substrates and interacts with them according to its respective function. Both specificities are pregiven … on the molecular level (2002, p. 3).

The apparently complex behaviour of bacteria in response to multiple simultaneous stimuli (positive and/or negative) is merely the resultant of two or more *inflexible* existing patterns of behaviour. The rules governing the behaviour of bacteria do not change; all that changes are the external circumstances (i.e. the presence of a new attractant or repellent). The behaviour of the bacteria in this case can be perfectly well described using the mind-neutral goal-centred version of Dennett's intentional stance. *E. coli* bacteria have a built-in preference for one *goal* (glucose) over another (galactose), which explains their response to the new *information* that glucose is nearby. There is no need to invoke mental states here.

The general point we can make here is that while fixed patterns of behaviour are sometimes accompanied by mental states in humans and other animals with minds, there is nothing about fixed patterns of behaviour *per se* which warrants a mentalistic interpretation.

> Conclusion 4.2: Behaviour by an organism which conforms to a *fixed* pattern or rule is not a sufficient warrant *per se* for ascribing cognitive mental states to that organism, even if stimulus-response coupling is indirect and modifiable (by the addition or removal of other stimuli).

As all patterns of behaviour in organisms are by definition either fixed or flexible, we may conclude:

> Conclusion 4.3: The occurrence of flexible behaviour in an organism is a necessary condition for the warranted ascription of cognitive mental states to it.

The distinction drawn between fixed and flexible patterns of behaviour might suggest that we can divide the world into mindless individuals whose patterns of behaviour are utterly rigid, and individuals that can behave flexibly and are therefore eligible candidates for having minds. However, such a dichotomy overlooks the fact that fixed patterns of behaviour are beneficial to organisms, as they provide them with a reliable way of attaining their good. Cognition is the cream on the biological cake: even in animals with minds, flexible behaviour is built upon a supporting bedrock of fixed patterns of behaviour, which the creatures rely on for their survival.

> Cognition, at least in Nature, can exist only in organisms that are able to live without it... All basic bodily functions are controlled automatically at the level of physiological regulation. Essential action patterns are innate... (Strube, 1998, pp. 2, 12).

4.2.2 Why memory matters for having a mind

Conclusion 4.2 above allows us to articulate more clearly the basis for our conclusion in chapter three, that the existence of *memory* in an organism was a necessary ground for ascribing cognitive mental states to it. An organism lacking a memory capacity could only exhibit fixed patterns of behaviour, which by themselves do not warrant the ascription of cognitive mental states. A goal-centred intentional stance is sufficient to explain the organism's behaviour.

4.2.3 Does flexible behaviour warrant the ascription of mental states to organisms?

Since flexible behaviour is often regarded as a hallmark of cognition, it would be tempting to conclude that cognition occurs in all organisms. However, I contend that it is simplistic to equate flexible behaviour with cognition, as flexible behaviour may be *externally* rather than *internally* generated. Even if organisms can acquire new information transfer paths (i.e. new cellular program instructions) that enable them to process stimuli differently, they may still lack an *in-built mechanism* for acquiring information that allows them to modify their response to a stimulus.

I suggest that gene swapping should be regarded as a *random exchange of information-processing mechanisms between organisms*, whereby they acquire new cellular program instructions. In gene swapping, the process of program modification is *externally* driven. An organism is incapable of modifying its own program: it can only acquire a new set of pre-packaged instructions from an outside source: another organism. Without this external source, the organism is incapable of changing its response to a stimulus, which remains fixed. In other words, the flexibility conferred by gene swapping is of a very limited sort: once the bacterium has acquired the instructions that alter its response to a new stimulus, its information processing pathway remains the same until another "gene-swap" occurs. One could say that after the swap, the bacterium has simply acquired a new, more adaptive kind of behavioural rigidity.

Internally driven flexible behaviour, on the other hand, arises from a self-adjusting mechanism *within* organisms, which enables them to continually "tailor their own responsive dispositions to their particular surroundings" (Beisecker, 1999, p. 298). In other words, an *internal mechanism* for generating flexible behaviour allows an organism to change its old pattern of responding to a stimulus. The change in its behaviour is *caused by the stimulus* which is sensed by the organism, but without the organism having to acquire any new *program instructions* from the stimulus. Since the change is internally driven, it seems appropriate to speak of “learning” here, but we will defer consideration of this question until the following chapter.

Another feature of gene-swapping which precludes it from being a cognitive (or belief-driven) adjustment to a stimulus is the fact that the change in response bears *no relation to the stimulus itself.* Indeed, the stimulus (mercury) is *incapable* of modifying the organism's response to it. Whenever gene-swapping alters the response of an organism (e.g. a bacterium) to a stimulus (e.g. mercury), the changes are caused by *foreign genes* from another organism, rather than by the stimulus itself. By contrast, in animal learning, a *stimulus* (the sight of a predator) can generate a new *belief* (that there is a predator in my vicinity) in a straightforward manner. Thus the causal chain between stimulus and response in gene-swapping appears to be fundamentally different from that which occurs in a cognitive mental process.

Additionally, the process of gene-swapping satisfies neither of the criteria traditionally invoked as hallmarks of mental states: it is not a *conscious* process, in any sense of the word, and it seems to lack the kind of *intentionality* we would expect of a mental state: being random, it has no conditions of satisfaction.

Finally, gene-swapping can be described using the goal-centred version of Dennett’s intentional stance: bacteria pick up *information* (transposons) from their partners, which assists them in meeting their built-in *goals* of surviving and reproducing.

I should add that if one were to envisage the acquisition of mercury resistance in mentalistic terms, one would have to view all other instances of gene-swapping between organisms in the same way. It would then follow that all organisms had minds.

Conclusion 4.4: The occurrence in an organism of flexible behaviour *per se* does not

> provide a sufficient warrant for the ascription of mental states to it.
>
> Conclusion 4.5: Internally generated flexible behaviour in an organism is a necessary condition for the warranted ascription of cognitive mental states to it. That is, the organism must be able to modify its patterns of information transfer, by means of an inbuilt mechanism, before we can justifiably impute beliefs to it.

None of the biological phenomena we have examined in the case studies in this chapter qualified as instances of internally generated flexibility. In the next chapter, I examine another class of phenomena which promises to deliver this kind of internal flexibility: the *learning behaviour* of organisms.

Chapter 5 - Learning-related Criteria for Attributing Mental States to Organisms

5.1 A scientific overview of the most common varieties of learning found in organisms, and their extent of distribution

The definitions listed in this section are technical, rather than popular; they reflect the varieties of learning distinguished by scientists who study animal cognition.

Learning is generally defined by psychologists as "a relatively permanent change in behaviour potential as a result of experience" (Abramson, 1994, p. 2). This definition excludes changes in an organism's behaviour that are due to other factors such as physical development, ageing, fatigue, sensory adaptation, or circadian rhythms (Abramson, 1994, p. 2). Despite claims that it occurs in bacteria (Di Primio, Muller and Lengeler, 2000, p. 7), learning, in the sense used by psychologists, appears to be confined to eukaryotes (organisms such as protoctista, plants, fungi and animals, whose cells have a nucleus). Claimed occurrences of learning in bacteria appear to be actually cases of sensory adaptation, in which an organism's response to a stimulus may diminish because the organism's sensory organs no longer detect the stimulus (Abramson, 1994; Illingworth, 1999).

5.1.1 Non-associative learning (habituation, sensitization and dishabituation)

The term n*on-associative learning* is used by scientists to refer to "those instances where an animal's behaviour toward a stimulus changes in the absence of any apparent associated stimulus or event (such as a reward or punishment)"(Encyclopedia Britannica, 1989). Only one kind of event (the stimulus) is involved in this kind of learning.

In the case of non-associative learning, "[t]he animal does not learn to do anything new or better; rather, the innate response to a situation or a particular stimulus is modified" (Abramson, 1994, p.

37). This type of learning is found in eukaryotes (organisms whose cells have a nucleus), but not in bacteria or archaea.

Three forms of non-associative learning are commonly distinguished in the literature: habituation, sensitization and dishabituation.

5.1.1.1 Habituation

Habituation is usually defined as the decline of a response "as a result of repeated stimulation" (Abramson, 1994, p. 106). Di Primio, Muller and Lengeler (2000, p. 7) use the same definition. For example, a loud noise nearby makes you jump when you first hear it. But if you keep hearing a loud noise at regular intervals after that, you get used to it and learn to "tune it out".

Habituation has obvious survival value: it enables an organism to *ignore* stimuli that do not transmit any biologically relevant information (e.g. about food or the presence of a predator), thereby minimizing energy that may be wasted on fleeing from "false alarms" and giving it more time to engage in behaviour that enhances reproduction. Habituation also makes it possible to engage in consummatory behaviour in unpalatable but otherwise suitable situations (Abramson, 1994, p. 107).

Habituation appears to be confined to eukaryotes. Bacteria and archaea do not display habituation proper, in which an organism's response to a stimulus may decrease because it adjusts to the repeated presentation of the stimulus. Claimed cases of habituation in bacteria are better explained as *sensory adaptation* (not to be confused with sensitization), in which an organism's response to a stimulus may diminish because the organism's sensory organs no longer detect the stimulus (Abramson, 1994). Other reasons why a response may diminish include effector fatigue (where the bodily mechanisms, or effectors, which respond to stimulation tire over the course of time), and the fact that some organismic responses diminish naturally over time, even without stimulation (decline in the base rate of responding). These distinctions are philosophically significant: *ignoring* an irrelevant stimulus is a very different thing from failing to detect it, or being too tired to respond to it, or simply slowing down over the course of time.

Di Primio, Muller and Lengeler (2000, p. 7) cite the observation that following prolonged exposure to an attractant, they change from a "run" to a "tumble" movement, as evidence for habituation in bacteria. However, if we examine the chemical basis for so-called "habituation" in bacteria, it appears to be a case of sensory adaptation, rather than habituation: at high concentrations, bacterial receptors become less sensitive (Illingworth, 1999).

Habituation proper has been identified in all major groups of eukaryotes (organisms whose cells contain a nucleus), including protoctista (especially protozoa, such as paramecia and amoebae), plants (*Mimosa;* the carnivorous plant *Drosera*, known as the Sundew; and the Passion Flower *Passionflora gracilis*) and animals (Abramson, 1994, pp. 106, 112, 116, 117; Abramson et al., 2002). According to Wood (personal email, 18 June, 2003), habituation in protozoa is biochemically distinct from adaptation in bacteria, which was rejected earlier as a possible instance of habituation.

5.1.1.2 Sensitization

Sensitization can be defined as "the opposite of habituation and refers to an increase in frequency or probability of a response" to a stimulus (Abramson, 1994, p. 105). For example, after hearing a loud crash, smaller noises, which would otherwise go unnoticed, can startle a person. Non-associative sensitization occurs when an organism is exposed to an intense stimulus that elicits a strong innate response (e.g. electric shock).

Sensitization is biologically useful, because it allows an organism to remain responsive to significant or dangerous stimuli, and to rapidly learn adaptive behaviour patterns (Abramson, 1994, p. 107).

Sensitization and habituation are "ubiquitous throughout the animal kingdom. They even appear, for instance, in experiments in which the 'animals' consist of single cells or isolated ganglia" (Abramson, 1994, p. 106).

Associative sensitization, also known as conditioned sensitization or *alpha conditioning*, is defined as an increase in the probability of an innate reaction based on the pairing of two *stimuli.* In alpha conditioning, unlike classical conditioning, the conditioned responses arise from an innate connection. There is some debate in the literature regarding the relationship between alpha and classical conditioning (Abramson, 1994, pp. 105-106).

5.1.1.3 Dishabituation

Dishabituation is defined as a "facilitation of a decremented or habituated response" (Rose and Rankin, 2001, p. 63). Dishabituation was formerly thought to be a special form of sensitization, but research on molluscs and leeches has now confirmed that it is a biologically and behaviourally distinct process (Rose and Rankin, 2001, p. 63). Dishabituation is found in most animals.

As regards the prevalence of non-associative learning in nature, we may conclude:

> Conclusion 5.1: Habituation and sensitization are confined to eukaryotes, or organisms with a nucleus in their cells.

5.1.2 Associative learning

Abramson (1994, p. 38) defines associative learning as:

> a form of behaviour modification involving the association of *two or more events*, such as between two stimuli, or between a stimulus and a response. In associative learning, an animal does learn to do something *new* or *better* (1994, p. 38, italics mine).

There are two broad categories of associative learning: classical conditioning and instrumental or operant conditioning.

All worms (with the possible exception of flatworms), and "higher" animals, such as arthropods (e.g. insects), molluscs, echinoderms and chordates (e.g. vertebrates) are capable of undergoing associative learning of some sort.

5.1.2.1 Classical conditioning

The simplest and most universal variety of associative learning is classical conditioning, which refers to the modification of behavior in which an originally neutral stimulus – known as a *conditioned stimulus* (CS) – is paired with a second stimulus that elicits a particular response -

known as the *unconditioned stimulus* (US). The response which the US elicits is known as the *unconditioned response* (UR). An organism exposed to repeated pairings of the CS and the US will often respond to the originally neutral stimulus as it did to the US (Abramson, 1994, p. 39).

Pavlov's experiment, in which he trained a dog to salivate at the sound of a bell that it associated with being fed, is an example of classical conditioning. A dog normally salivates (UR) at the smell of food (US). But if the dog hears the sound of a bell (CS) just before smelling the food, it will learn to associate the sound of the bell with being fed, and salivate when it hears the bell (CR).

It should be noted that if the CS and US occur *simultaneously*, or if the CS occurs *after* the US, virtually *no* conditioning will occur. The CS needs to *precede* the US and be predictive of it. An animal obtains no biological advantage in learning an association between a CS and a US unless the CS can be used to *predict* the US.

According to Brembs (2000, p. 2), classical conditioning is biologically beneficial because it allows organisms in the wild to associate biologically neutral stimuli with significant ones, enabling them to make better predictions about their environment.

It has been claimed that some ciliates *(paramecia)* possess a capacity for learning through classical conditioning. The original experiment was reported in a study by Hennessey, Rucker and McDiarmid (1979) and is still widely quoted (e.g. by Martin and Gordon, 2001). However, as far as I have been able to ascertain, no-one has replicated, or even attempted to replicate, this result. A study using different procedures on another ciliate *(Stentor)* failed to find evidence of *instrumental* conditioning (Wood, personal email, 18 June 2003; Hinkle and Wood, 1994).

Research into the possibility of classical conditioning in *Mimosa* plants has produced negative or conflicting results, and the methodology of studies which found conditioning has been criticised (Abramson et al., 2002, pp. 175-176). In keeping with my methodological constraints, these studies will be ignored here. A more refined plant learning experiment by Abramson et al. (2002, pp. 173 - 185) on *Philodendron cordatum* plants yielded negative results. As there are several possible explanations for this failure, follow-up research with longer intervals is recommended by the authors.

In short: although protoctista and plants are highly adaptive, multicellular, eukaryotic organisms which are capable of being habituated (unlike bacteria), there is no good evidence to date that they are capable of associative learning. Nor have I been able to uncover any evidence for associative learning in “simple” animals, such as sponges and cnidaria (coelenterates).

The attribution of associative learning to the simplest flatworms (planarians) is controversial, as different authorities use definitions of "associative learning" and some of the effects of associative learning can be mimicked by other behavioural processes in animals which do not involve learning - for example, pseudo-conditioning and sensitization. An additional complication is that most of the conditioned stimuli used in learning experiments with invertebrates are not, properly speaking, neutral. For a comprehensive discussion of technical issues, see Abramson, 1994, pp. 130-133, 157-159; for a discussion of experiments on various groups of worms, see Duane, Fortenberry and Horowitz (1998); see also Abramson and Buckbee (1995).

However, recent research on another worm, the well-studied nematode (roundworm)

Caenorhabditis elegans, whose nervous system contains a mere 302 neurons, has demonstrated that even worms with very "simple" nervous systems are capable of associative learning - in particular, *classical conditioning.* The evidence that *C.elegans* is truly capable of undergoing classical conditioning includes recent studies showing that *C. elegans* worms can actually be conditioned to radically alter their preferences: they will avoid a stimulus they had formerly been attracted to, after it has been paired with an aversive stimulus (Morrison, Wen, Runciman, van der Kooy, 1999), or after being experimentally subjected to starvation for several hours on a plate containing an attractive stimulus (Saeki, Yamamoto and Iino, 2001). This change of preference cannot be explained away as "conditioned sensitization" because the old response is not re-awakened. The worms are actually learning to do something *new:* they are changing their pattern of response to a stimulus. Additionally, Catharine Rankin, who specialises in learning mechanisms in *C. elegans,* claims (personal email, 31 May 2003) that recent work has established that it can indeed associate a CS with a US.

Classical conditioning can be observed in most or all groups of worms (with the possible exception of flatworms), as well as "higher" animals, such as arthropods (e.g. insects), molluscs, echinoderms and chordates (e.g. vertebrates).

We may conclude:

> Conclusion 5.2: Associative learning appears to be confined to organisms with central nervous systems. It is found in most but possibly not all phyla of animals with central nervous systems.

5.1.2.2 Instrumental and operant conditioning

By contrast, *instrumental conditioning* (also known as trial-and-error learning) and *operant conditioning* are "examples of associative learning in which the behavior of the animal is controlled by the *consequences* of its actions... [Whereas] classical conditioning describes how animals make *associations* between stimuli, ... instrumental and operant conditioning describe how animals associate stimuli with their own *motor actions* ... Animals learn new behaviours in order to obtain or avoid some stimulus (*reinforcement*)" (Abramson, 1994, p. 151, italics mine). For example, a rat placed in a box with a lever may press the lever by accident, or during exploratory behaviour, and receives a small pellet of food. If this happens a few times, the rat will start to press the lever more and more often, because it has learns to associate one of its bodily actions (pressing a lever) with a desired result (obtaining food).

> Operant conditioning is clearly beneficial to animals: it reinforces behaviour that satisfies animals' appetites, or enables them to avoid aversive stimuli (Brembs, 2000, p. 2).

For many psychologists, the terms "instrumental conditioning" and "operant conditioning" are *inter-changeable:* the terms are used to mean something like "behaviour controlled by its consequences" (Abramson, 1994, p. 151). However, a few researchers follow Skinner in *reserving* the term "operant learning" (also known as operant behaviour or operant conditioning) for behaviour which "appears more flexible and adaptable" (Grau, 2002, p. 85) than instrumental conditioning. Abramson (1994) also makes a distinction between "instrumental conditioning", and "operant behaviour", reserving the latter term for special cases. He remarks that if we adopt the common definition of operant behaviour, then

> ...such behavior is present in all animal groups. However, if operant behavior is defined in terms of its functional influence on the environment and the ability to use an *arbitrary response*, then... [it] is limited to vertebrates and perhaps some species of mollusks, crustaceans, and insects. A rule of thumb I have found useful... is that in operant behavior, an animal must demonstrate the ability to *operate some device* - and know *how to use it*, that is, make an *arbitrary response* to obtain reinforcement. (1994, p. 151, italics mine).

However, Abramson's distinction between instrumental and operant conditioning has been queried by other researchers - for instance, Bjorn Brembs considers Abramson's definition to be "more obfuscating than enlightening" (personal email, 11 August 2003):

> What does he mean by "device"? How does he test that the animal "knows"? (personal email, 11 August 2003).

Defined broadly as “behaviour controlled by its consequences", *instrumental* conditioning is present in all animal groups (Abramson, 1994, p. 151).

5.1.3 Other kinds of learning

Other kinds of learning distinguished in the literature on animals include *latent learning* (an exploratory phenomenon, in which an animal associates indifferent stimuli with each other without expectation of a reward); *spatial learning* (the ability to recording information about one's environment and its spatial orientation); *insight learning* (the intuitive grasp of the solution to a problem, without needing to go through the intervening series of the trial and error steps that are associated with most types of learning); *observational learning* (the ability of an animal to learn a new skill by observing another animal successfully execute the same skill); *reversal learning*, where the animal first learns to make a discrimination, such as choosing a black object in a black-white discrimination problem, and then is supposed to learn to reverse its choice – i.e. to choose the white object; and *imprinting* (a learning process, observed in young birds and mammals, in which they identify with another animal). Of the varieties listed above, insight learning is the most controversial, and there is widespread disagreement among scientists regarding its occurrence and frequency in the animal kingdom.

5.2 Philosophical evaluation

5.2.1 Which of the above kinds of learning should count as true learning?

The first thing that needs to be noted about the above classification is that the word "learning" is used by psychologists in a sense very different from that of ordinary usage. In popular parlance, the principal meaning of "learn", as defined by *The Merriam-Webster On-line Dictionary* (2006), is:

> to gain knowledge or understanding of or skill in by study, instruction, or experience.

If we follow this popular definition, we can define “true” learning as something involving either the acquisition of a *skill,* or the acquisition of (propositional) *knowledge.* If an animal acquires neither knowledge nor a skill, there has been no “true” learning. Accordingly, we need to address the question of which, if any, of the forms of "learning" described in 5.1 qualify as “true” learning.

It is not clear from the foregoing definition that learning needs to involve a *mental state*, as the acquisition of a skill by experience counts as “true” learning. In the discussion that follows, I shall therefore refrain from assuming that the term “learning” necessarily refers to a mental state.

5.2.1.1 Reproducible learning effect

Kilian and Muller (2002) argue that any organism capable of learning should not only be able to change their behaviour, but be able to reproduce the behavioural change when the occasion demands it:

> In ethology learning is defined as a change in the individual behavior which leads to a better adaptation and which is influenced by amplification and experience. Proper learning shows a *reproducible learning effect*... [E]pigenetic learning is based on the ability to form the interplay between stimulus input, information transfer, memory, and behaviour in an individual and reproducible way (2002, pp. 1 - 2, italics mine).

The authors’ point is that learning is a teleological process, which benefits its possessor. On strictly logical grounds, the benefits of learning - better adaptation - can only accrue if what is learned is *remembered:* the individual must be able to store new information and subsequently use it, in response to changing circumstances (display a "reproducible learning effect"). An organism must satisfy this criterion before scientists can justifiably attribute learning to it.

> Conclusion 5.3: Learning should not be attributed to an organism unless it displays a change in one of its patterns of behaviour, which it is able to *reproduce* on a subsequent occasion.

On the other hand, the possession of *memory* by an organism does not guarantee that it is capable of “true” learning. A creature's ability to store new information about its surroundings does not necessarily mean that it can acquire new *skills,* let alone knowledge or understanding.

5.2.2 Flexible behaviour

An additional requirement for "true" learning is *flexibility* of response patterns. According to the popular definition cited above, an individual that learns, thereby gains something - knowledge, understanding, or at least a *skill* - from its experience. But it is impossible for an organism to manifest its new knowledge, understanding or skills if its behaviour patterns are *fixed*, according to the definition given in chapter 4. We may therefore formulate the following conclusion:

> Conclusion 5.4: Learning should not be attributed to an organism unless it is capable of displaying *flexible* behaviour (as defined in chapter four).

The foregoing conclusion entails that organisms undergoing *habituation* are not learning to respond in a new way, but are simply diminishing their innate response to a stimulus, after repeated exposure to it. The waning in their response over the course of time does not represent a change in their pattern of behaviour, but can be described as a continuance of an existing fixed pattern. Putting it mathematically, we can describe an organism's response to a stimulus in terms of a function F, whose inputs include not only the intensity of the stimulus but also the number of exposures to it. During habituation, the function F does not change; all that changes over time is the value of one of the parameters (number of exposures). This is an *historical* parameter, but as we saw in chapter four, nothing in the definition of "fixed pattern" prevents us from using time-dependent variables to describe animal behaviour. The same remarks apply to *sensitization,* which is the opposite of habituation (Abramson, 1994, p. 105).

The fact that psychologists customarily refer to habituation as a kind of "learning" indicates that their definition of learning is broader than popular usage: it includes cases where an organism does not acquire a skill and therefore does not undergo “true” learning.

> Conclusion 5.5: The ability of an organism to undergo (non-associative) *habituation* and *sensitization* is not a sufficient condition for “true” learning, defined as the acquisition of a new skill or new knowledge.

It should be stressed that habituation is a phenomenon which admits of varying degrees of complexity in the animal world (see Balkenius, 2000 for a discussion), and that the distinction between associative and non-associative forms of learning is not always a clearcut one. Work by Rose and Rankin (2001) suggests that associative cues can facilitate habituation in the nematode worm *C. elegans.* The above conclusion applies only to cases where the change in behaviour is a non-associative one.

5.2.3 Why the internally generated flexibility required for true learning is unique to multicellular organisms

Kilian and Muller (2002, pp. 3-4) do not regard the *non-associative* forms of behaviour modification found in *unicellular* organisms as “true” learning. They contrast the *non-associative* forms of behaviour modification found in unicellular organisms with that of multicellular organisms, where the cells involved in chemical information transfer retain their *functional specificity*, but this specificity is de-coupled from the "goal" of the behaviour. At the beginning of the individual's life, the *goal* is unspecified or "open-ended": the synaptic connections are not fixed. The same stimulus may be linked to any one of a multitude of responses. Different paths open up, but the final selection from the range of goal cells that come into play is made by the stimulus in the individual's local external environment (e.g. in imprinting, the first object the individual happens to see), leading to an individualised, reproducible learning effect. Paths between sensors and effectors vary from one individual to another, and are not genetically determined.

Kilian and Muller argue that because unicellular organisms have "genetically fixed ... networks of information transfer" (2002, p. 3), they are unable to modify their own information transfer pathways and so cannot learn to respond to a new stimulus:

> Unicellulars are not able to synthesize or modify cellular substances not already coded genetically in a fixed way. So, *they cannot install new, individualized, ... reproducible information transfer paths within their life-time*... The synthesis of new substances for new irreversible and reproducible information transfer paths as an answer to a new, formerly not identifiable stimulus is not possible in unicellulars. *Therefore, unicellulars seem most certainly not able to learn, at least according to an ethological definition* (2002, pp. 2-3, italics mine).

I interpret Kilian and Muller's claim that in multicellular organisms, functional specificity is retained at the chemical (micro) level but that there is no goal specificity at the macro level, as equivalent to the claim that only multicellular organisms are capable of *internally generated flexible behaviour* (as I defined it in chapter 4). Kilian and Muller’s argument can be expressed mathematically, if we suppose that:

(i) each *goal-oriented* form of behaviour exhibited by a multicellular organism can be described by a *higher-level* (macro) function (or combination of functions);

(ii) higher-level functions invoke various *lower-level* (micro) functions within the organism which describe its interactions at the *chemical* level;

(iii) although the higher-level functions that describe an organism's response to various stimuli are *not* fixed at the beginning of its life, *the lower-level functions they invoke constitute a fixed, finite set (a repertoire).* Each higher-level function invokes a succession of several functions from this fixed set. A simple example: $z = F(x) = f(g(h(x)))$, where: x is an *input* variable corresponding to the intensity of a certain kind of stimulus; f, g and h are three *low-level* "chemical" functions belonging to the organism's fixed repertoire $\{f, g, h, j, k, l, \ldots\}$, which operate successively on the same stimulus; F is a *higher-level* function that describes a certain kind of behaviour; and z is the behavioural *output* corresponding to F;

(iv) the organism possesses an *internal mechanism* allowing it to *modify* its *higher-level* functions. In the above example, the definition of $F(x)$ might *change* by invoking another of the organism's fixed repertoire of *low-level* functions (e.g. $F(x) = f(g(j(x)))$ instead of $f(g(h(x)))$), or invoking the functions in a different *order* (e.g. $F(x) = f(h(g(x)))$, instead of $f(g(h(x)))$), or by changing the *program conditions* which define the output.

We can now see why bacterial *gene-swapping* (discussed in chapter 4) does not count as learning. In this case, the program governing a bacterium's response to mercury *does not modify itself:* it receives new, pre-packaged instructions from an *external* source (another bacterium).

While this qualifies as flexible behaviour, it differs in important respects from what Kilian and Muller (2002) would characterise as true learning. Lack of goal specificity is a vital feature of their account: the same stimulus may be linked to any one of a multitude of responses.

> Conclusion 5.6: The occurrence in an organism of *flexible* behaviour is not a sufficient condition for learning.

> Conclusion 5.7: True learning is confined to multicellular organisms.

Kilian and Muller (2002) suggest that the kind of information transfer functionality required for true learning is possible "only when several cells of information transfer functionality come into close contact spatially, i.e. in organisms with *central nervous systems*" (2001, p. 4, italics mine). We have already seen that *associative learning* is confined to these organisms (conclusion 5.2).

Kilian and Muller’s definition of learning is similar to but somewhat broader than that of Beisecker (1999, p. 298), who holds that organisms are capable of learning only if they have the "ability to tailor their own responsive dispositions to their particular surroundings". On this definition, even *imprinting* would not qualify as true learning, as the behavioural response, once formed, cannot be subsequently modified to fit different circumstances (e.g. the death of the "parent" imprinted on the newborn individual's memory).

5.2.4 Associative learning

5.2.4.1 Why associative learning qualifies as true learning

Associative learning certainly qualifies as *flexible* behaviour according to the definition we have given. It is not *fixed,* as the value of the output variable (i.e. the response) for the same input variable (stimulus) does not remain the same over time. There is genuine novelty here, which cannot be treated as a temporal extension of an *existing* pattern of activity within the organism by introducing extra *historical variables,* as we did with habituation. Instead, what we see here are either new *conditions* for activating an existing behaviour pattern (classical conditioning), or the emergence of a new *behaviour pattern* (instrumental or operant conditioning). In a simple case of classical conditioning, the organism learns to respond to a new stimulus (the conditioned stimulus) in the same way as it does to an existing one (the unconditioned stimulus). This is flexible behaviour, because one of the *programs* governing an organism's behaviour changes over time: there is a change in the conditions under which one of its behaviour function(s) is activated. In operant conditioning, the organism acquires a new behavioural function through "trial-and-error learning". Once again, this requires a program change.

Moreover, the new behaviour is acquired through an *internal* learning mechanism. This in-built mechanism for acquiring information allows the individual to modify its response to a stimulus.

> Conclusion 5.8: The capacity for associative learning in an organism is a sufficient condition for its being able to engage in *internally generated flexible behaviour.*

According to the dictionary definition of learning cited above, acquiring a *skill* by experience qualifies as learning. Even in the simplest cases of associative learning, this is what happens. We saw that *C. elegans* roundworms could be conditioned to avoid a stimulus that they were previously attracted to – in other words, change their preferences. Since their preference-changing behaviour is internally generated, flexible, and biologically useful to them, it surely deserves to be called a skill. It follows that some worms can learn, regardless of whether they have mental states or not. They are "educable" creatures.

> Conclusion 5.9: The ability of an organism to undergo *associative learning* (classical and/or instrumental conditioning) is a sufficient condition for its being able to *learn*, in the *true* sense of the word, as such an organism can acquire skills.

5.2.4.2 Does associative learning involve mental states?

5.2.4.2.1 Dretske: instrumental/operant conditioning as a hallmark of belief and agency

Exposition of Dretske's account

From the beginning, some pioneers of learning research (notably Pavlov) have championed a *non-mentalistic* account of conditioning (but see Brembs, 1996, for a discussion of problems with Pavolv's stimulus substitution theory). However, some philosophers have championed a mentalistic account of conditioning - especially *instrumental or operant conditioning.* In this thesis, I examine two such philosophical accounts: those of Dretske (1999) and Beisecker (1999, 2000). Both philosophers have brought valuable insights to bear on the contemporary discussion of belief and agency in non-human animals. I shall argue that the concepts they appeal to in order to provide us with a way of distinguishing animal agency from non-intentional behaviour need to be fleshed out more: Dretske needs a more robust concept of *representation* to make his account work, while Beisecker (whose proposals I evaluate in chapter seven) needs a mechanism to explain how *self-correction* can take place in animals. I shall attempt to remedy both of these problems.

Dretske (1999) considers creatures that can undergo conditioning as "educable", and credits them with beliefs that are acquired and modified by the associations they form. Dretske (1999) has argued that animals which can undergo *operant conditioning* are capable of not only memory and learning, but also *belief* and *agency.* As Dretske makes no distinction in his article between *instrumental* and *operant* conditioning, I shall assume that he uses the two terms interchangeably (unlike Abramson, 1994, who distinguishes between them). Dretske carefully differentiates agency from *behaviour:* behaviour may have a meaning, but *purposeful* acts are *governed by their meaning.* When an action occurs, events that have a meaning cause an animal to behave in a certain way, *by virtue of their meaning.* With computers and plants, on the other hand, events which have a meaning we can recognise, cause something to behave, but it is the intrinsic properties of the events, and not their meaning, that explains the behaviour.

Dretske illustrates his point with a few well-chosen examples. A speaker utters the words "Vibrate rapidly" into a microphone. The microphone's diaphragm vibrates rapidly. The sounds made by the speaker had a meaning – "Vibrate rapidly" – but it is not what the speaker says, but the physical properties of the sounds, which cause the diaphragm to vibrate rapidly. (The microphone would have vibrated rapidly even if the speaker had said, "Be still".) The microphone is sensitive to sound, not meaning.

A thermostat turns the heat on and off, keeping the room at a comfortable temperature. Is this mere behaviour or is it action? We could say that the thermostat simply behaves because it lacks beliefs and desires: a *goal-centred* intentional stance suffices to explain its behaviour. But Dretske goes beyond this obvious response and attempts to explain what would be needed to make the thermostat's behaviour a *bona fide* action.

Most thermostats have a bimetallic strip that functions as a thermometer (its degree of curvature represents room temperature) and as a switch - if the room cools to the desired temperature, the strip touches an adjustable contact, whose position corresponds to the room's desired temperature, thereby closing an electrical circuit to the furnace and turning the heat on. The thermostat senses a drop in temperature and corrects it. Although the curvature of the thermostat's bimetallic strip *means* something (it represents the ambient temperature), it is *not the meaning, but the curvature*, that governs the thermostat's behaviour. Dretske argues that if we take away the meaning but keep the curvature (e.g. by bending the strip with a pair of pliers), the thermostat will behave the same as it would if the room were cool.

The adaptive behaviour of the Scarlet Gillia illustrates why *plants* cannot act either, according to Dretske. Even though this plant can change colour from red to white in summer, thereby attracting pollinators, its adaptive behaviour is triggered not by its "meaning" or biological significance, but by the intrinsic properties of the *chemical switches* that cause its behaviour, coupled with the *historical* fact that similar behaviour by its evolutionary forebears, from whom it inherited its genes, enables it to reproduce successfully.

Now consider a foraging bird, which tries to eat a Monarch butterfly that has been reared on a toxic form of milkweed. Eating the butterfly makes the bird vomit. The next day, the bird sees a Viceroy butterfly, which looks remarkably like the poisonous Monarch. The bird flies away. Why does the bird not eat the Viceroy? Its behaviour can readily be explained in terms of *operant conditioning:* it learned to avoid the stimulus (Monarch butterflies) after a punishing experience, and its internal representation of a Monarch butterfly caused it to avoid a similar-looking butterfly. Whereas a thermostat is hard-wired to behave as it does, the bird's behaviour is triggered by its memory (stored internal representation) of an unpleasant experience, where it learned that a Monarch was poisonous.

> Learning of this sort *[operant conditioning – V.J.T.]* consists in harnessing these *internal representations* to *control circuits* so that behaviour will occur in the external conditions on which its success depends. Like the thermostat..., this internal representation ... has both a meaning and a causal role, but, unlike the instrument..., *its meaning explains its causal role* (Dretske, 1999, p. 29, italics mine).

The nub of Dretske's account is that many animals have a *learning history,* which imparts a *meaning* to their experiences. It is this meaning which explains their causal role in animal behaviour. When animals use what they have learned to achieve their goals, they can be described as *agents.* Of course, biologically adaptive hard-wired behaviour (e.g. reflexes) in animals is not agency: its *efficacy* is independent of any *meaning* (or biological purpose) it may have. Using our terminology, we might say that unlearned behaviour can be explained in terms of a mind-neutral, *goal-centred* intentional stance.

In the passage cited above, Dretske equated learning with success, as it enables an organism to achieve its goals. The organism does this by drawing upon past experiences (which it stores as internal representations). These experiences can be said to *associate* acts (e.g. eating a Monarch butterfly) with consequences (e.g. becoming ill).

This brings us to the crux of the matter: it makes no sense to ascribe agency to an individual unless it is capable of learning the consequences of at least some of its actions. The significance of associative learning with regard to *intentional agency* is that it appears to be the most fundamental, universal way in which organisms (on Earth, at least) connect *acts* with *consequences.* In the absence of such an association, it is difficult to envisage how agency (and the beliefs and desires which accompany it) could be said to occur at all. Since we argued in chapter one that the ascription of beliefs to animals is warranted only if they are capable of manifesting intentional agency, then we may conclude:

> Conclusion 5.10: An organism must be capable of associative learning before it can be said to possess mental states.

The bird's internal representation causes the avoidance behaviour precisely because it *means* something about its external environment (i.e. that a certain kind of butterfly is present - the sort of butterfly the bird, after its experience of vomiting, wants to avoid). In this case, we cannot isolate the meaning from the physical properties of the representation and say that the latter, and not the former, cause the behaviour. Here, according to Dretske, we have *bona fide* agency:

> It is the meaning of these events that explains why the internal events are producing the behaviour. In this sense, then, the behaviour is *governed* by something. We have a genuine instance of action. In avoiding the butterfly, the bird is an agent (1999, p. 30).

Dretske thus seems to regard a capacity for instrumental conditioning as not only a necessary condition for agency, but also a sufficient one. In the above case, according to Dretske, it is natural to say that the bird has *learned* the Monarch butterflies are poisonous, it *believes* that the Viceroy butterfly is a Monarch, and it flies away because it *wants* to avoid the Viceroy. Dretske uses the word "belief" rather than "knowledge" here, because the bird is in fact mistaken.

5.2.4.2.2 An equivocation in Dretske’s account?

In my opinion, Dretske does an excellent job of explaining why we say that machines and plants behave rather than act, but his defence of agency in animals is marred by an equivocation between two senses of the word "meaning" - one of which can be characterised in terms of a *goal-centred* intentional stance ("significance" or "purpose"), while the other requires an *agent-centred* stance ("aim" or "intention"). Dretske (1999) makes no attempt to differentiate between them. A thermostat's behaviour has a "meaning" for us as agents with beliefs and wants: we set it at a level that we believe will make us feel as warm as we want to feel. The thermostat has no wants of its own, but at least it possesses what Searle (1999, p. 93) refers to as *derived* intentionality. On the other hand, we can explain the “meaning” of the Scarlet Gillia’s behaviour using a *goal-centred* intentional stance, without invoking beliefs and wants at all. Information in the plant’s sensors activates the plant’s genetic program (which encodes information that benefited the plant’s ancestors), triggering behaviour that is biologically adaptive, allowing the plant to achieve its goals. Searle would regard the intentionality here as purely metaphorical, “as-if” intentionality. What about conditioned behaviour in animals?

If Dretske were trying to rigorously demonstrate that conditioned behaviour in animals could be described as agency simply because it could be characterised in *agent-centred* terminology, then he would indeed be begging the question. But in fact, Dretske seems to be making a *suasive* case (rather than a strictly logical one) for belief in animals that undergo instrumental or operant conditioning. He appears to be arguing that a belief-based account makes sense of the fact that the bird changes its behaviour because of what it remembers:

> Notice how natural it is in this case (unlike the case of the plant or the thermostat) to explain the bird's behavior in terms of what it *believes.* It is natural, I submit, because *memory* about some previously experienced object is so obviously implicated in why the bird behaves as it does... Talk of memory becomes appropriate here because behavior changes after a perceptual encounter… The thermostat and the plant come into the world *hard-wired* to behave the way they do. What happens to them is not relevant to why they behave that way... Not so with the bird (1999, pp. 28-29, italics mine).

5.2.4.2.3 A critical evaluation of Dretske's criteria for belief and agency

If we look at Dretske's writings, we can discern two criteria by which he claims to distinguish between organisms with and without beliefs. His wording above suggests the following criterion:

> Dretske Mark I:
>
> We can use the criterion of *learned, flexible behaviour* to distinguish between believing and non-believing organisms: whereas the thermostat and the plant are "hard-wired" (1999, p. 29) to behave as they do, the bird is not. It can learn.

This criterion is a little muddled: as we saw in chapter four, *flexible* behaviour is a universal feature of organisms, including plants. Dretske employs the terms "mindless", "unlearned" and "hard-wired" inter-changeably. However, *learned* flexible behaviour is restricted to certain kinds of animals: those with a central nervous system (see Conclusions 5.2, 5.8, 5.9). Can we draw the line here?

Ryder and Martin (1998) have attempted to discredit Dretske's account of belief by offering a counter-example. They have pointed out that the human autonomic nervous system (ANS) is capable of associative learning, yet few people would ascribe beliefs to it. Contrary to popular belief, the ANS does *not* simply behave according to *fixed patterns*, but can be *conditioned.* The ANS is controlled by the *central autonomic network* (CAN), which is located in the brainstem, mid-brain and fore-brain. Each individual's CAN contains an indicator which *represents* whether she is standing, and which compensates for the loss of blood to the brain (caused by gravity) when she stands up, by perfusing her brain with blood. When an astronaut stands up in zero-gravity, her blood rushes to her head, because the ANS is used to counteracting the effects of gravity on earth. However, with time, the astronaut's ANS un-learns this compensating behaviour. When the astronaut returns to earth, her ANS has to re-learn the skill of compensating for the loss of blood to the brain when she stands up. The authors argue that the above example meets Dretske's criteria for genuine learning, and that according to his criteria, the ANS is capable of having beliefs (e.g. about the astronaut's posture) and desires (e.g. to adequately perfuse her brain with blood). On the other hand, the attribution of *beliefs* to the ANS sounds peculiar, so Ryder and Martin argue that Dretske's criteria for having a belief must be insufficient.

Two other cases create similar problems for Dretske’s account. Leg withdrawal can be conditioned in headless cockroaches, or in isolated leg and thoracic ganglion preparations. The thoracic ganglion is a much more complicated cluster of nerves in the cockroach than the brain (Kentridge, 1995). Cockroaches are thus capable of "distributed" learning. Yet it would seem strange to describe their leg withdrawal behaviour as a manifestation of a belief, and there seems to be no good reason why we should resort to an agent-centred intentional stance instead of a goal-centred one, to explain this phenomenon.

The same could be said with regard to rats whose spinal cords had been severed at the second thoracic vertebra (T2), leaving them paralysed below their mid-sections. The rats were placed in an apparatus where a shock was administered to one of their hind legs, whenever it made contact with a solution of salt water beneath the rats. The rats were shown to be capable of undergoing *instrumental conditioning within their spinal cords:* they soon learned to maintain the leg in a flexed (up) position, thereby avoiding shock (Grau, 2002).

These three examples undermine the notion that internally generated flexible behaviour patterns suffice to distinguish organisms with beliefs from those that lack them. Learning organisms have an *internal mechanism* that allows them to vary their patterns of responding to their surroundings by acquiring new survival *skills.* But the acquisition of a skill need not be envisaged in mentalistic terms: a mind-neutral, goal-centred stance might be adequate to characterise it, as appears to be the case in these three examples.

> <u>Conclusion 5.11:</u> A capacity for associative learning in an organism does *not* provide a sufficient warrant for our being able to ascribe cognitive mental states to it.

If flexible behaviour cannot serve to identify instances of belief and/or agency in animals, then it might be more profitable to search for cognition by examining what it is that newly acquired behaviour patterns enable their owners to do. Dretske's second criterion for distinguishing organisms with beliefs from those without (which I shall call "Dretske Mark II"), addresses this very question by invoking the notions of *representation* and *control:*

Dretske Mark II

> Learning of this sort (operant conditioning) consists in harnessing these *internal representations* to *control circuits* so that behaviour will occur in the external conditions on which its success depends (Dretske, 1999, p. 29, italics mine).

How are we to understand Dretske's claim that operant conditioning can be understood in terms of internal representations linked to a controlled behavioural response? Two readings are possible. On a "maximalist" reading (which I shall defend in my model of operant agency in section C of this thesis), an organism with beliefs is one that can use its internal representations of its environment, acquired through learning, to control its surroundings. The verb "control" is here meant to describe an action, performed by an agent. (Of course, the concept of agency needs to be further developed within this account.) An alternative "minimalist" reading (which appears to be Dretske's own view) equates operant agency with behaviour that is controlled by an organism's internal representations of its environment. Here, "controlled by" simply means "caused by".

Dretske's account of representation is a thorough-going naturalistic, causal account, which eschews appeal to "interpreters" as the arbiters of what counts as a representation. Representations, on his account, are indicators that carry information about law-like connections (say, between As and Bs), but they are something more. Representations, unlike indicators, can be mistaken, because they have a *function* which they can fail to perform. More precisely, representations are indicators whose *natural function* is to indicate as they do, because doing so confers a selective advantage on the organism possessing them. Representations, unlike other natural indicators, are *not* hard-wired: they acquire a function or *meaning* for an animal only when the animal *learns* what they indicate. An animal's learning history imparts a meaning to its experiences. Belief-type representations are recruited as causes of bodily movements in an animal because the animal *learns* what they indicate. Thus beliefs are both *reasons* and *causes* of actions (Ryder and Martin, 1999, pp. 5-7; Pitt, 2002). This explains Dretske's (1999) contention that while behaviour may have a meaning, purposeful acts are governed by their meaning, insofar as events cause an animal to behave in a certain way, by virtue of their meaning.

Thus in Dretske's case of the bird that shuns the Viceroy after tasting a similar-looking noxious Monarch butterfly, the bird's internal *representation* causes the avoidance behaviour precisely because it *means* something about its external environment (i.e. that a certain kind of butterfly is present - the sort of butterfly the bird, after its unpleasant experience, wants to avoid).

But does Dretske provide us with an adequate account of representation? Once again, it appears that his account fails to handle Ryder and Martin’s (1998) counter-example. Dretske might claim that in this case, the system involved is not forming a representation. But in fact, the astronaut's ANS is representing its external environment, and its *function* is undeniably a natural, biological one. Moreover, some kind of learning is going on: in space, the astronaut's ANS learns not to compensate for loss of blood to the brain when she stands up, and once the astronaut returns to earth, her ANS has to re-learn this skill. Moreover, the astronaut’s internal representation is controlling her behaviour. Why does this not qualify as operant conditioning?

Recently, Procyshyn (2001) has exposed the flaws in contemporary naturalistic concepts of representation, focusing in particular on Dretske's and Millikan's (1984, 1993) accounts. The thrust of his criticism is that neither account adequately characterises the concept of a representation. The notion of representation makes no sense without the possibility of misrepresentation. However, neither account, argues Procyshyn, manages to successfully delineate cases of misrepresentation in the natural world from cases where the system in question is not representing anything at all.

The test case Procyshyn examines is that of the magnetosome bacterium, a marine microbe that contains a magnetotactic sensor, which aligns itself to magnetic north under normal circumstances. The sensor has a biological *function:* it enables the bacterium to avoid oxygen-rich surface water that is toxic to it. Since in the northern hemisphere magnetic north tends downward, the bacterium is thus carried toward deeper water and sediment, and away from toxic, oxygen-rich surface water.

Procyshyn (2001) shows that both accounts do a poor job of explaining why the bacterium is not *misrepresenting* magnetic north when it aligns itself to a bar magnet that we hold over it. The key point he makes here is that a representational account has to do some extra explanatory work that a purely causal account cannot; otherwise it is redundant. The familiar example of the patellar reflex illustrates why: when someone taps my bent knee with a mallet, a neural pathway fires, causing my leg to jerk and straighten - a reflex process over which I have no control. Now, we could say that the external stimulus (mallet) is represented by my neural pathway, and that the *function* of this representation of to straighten my leg. Or we could explain the reflex in purely *causal* terms. Given that both accounts describe what goes on equally well, we should prefer the simpler, causal account. A causal account of the magnetosome explains its behaviour perfectly well, so there is no need to postulate an internal representation.

Is there a better account of representation? Procyshyn thinks he can explain why magnetosomes and sunflowers do not represent: “they do not have computational, cognitive systems” (2001, p. 142). In order for an internal state to represent, it must have a special function: “to supplement a causal chain of events by encoding, using or delivering information indicated … [about the external world] … to the cognitive architecture of that organism” (2001, p. 142). Misrepresentation, on this account, is a “malfunction of the organism’s cognitive architecture” (2001, p. 142).

I find this account unsatisfactory, on two grounds. First, it leaves the term “cognitive” undefined. We could define it in terms of learning – but that would leave us unable to explain away Ryder and Martin’s (1998) example of the astronaut’s ANS, which is also capable of undergoing learning. “Truth-valuation” is another concept invoked by Procyshyn to shed light on the divide between representational and non-representational systems: “truth is implicit to misrepresentation and must be graspable by an organism that represents” (2001, p. 142). Very well; but what kind of behaviour counts as grasping truth?

Second, it is unclear how Procyshyn’s truth-functional account of representation supplements the causal chain of events. Does it make any predictions about the system’s behaviour that a causal account does not? It seems that all it does is describe what is going on inside the organism in more normative terminology, and at a higher level.

We have arrived at an impasse. The central problem is that Dretske's notion of representation is too weak to sustain mentalistic inferences. For him, the decisive features of representations are that they carry information which indicates something about the world, they confer a selective advantage on their possessor and they are learned. As we have seen in the preceding chapters, these properties are necessary for having mental states but not sufficient to justify the ascription of mental states to an organism.

5.2.5 A richer notion of representation

In my model of operant agency which I develop in section C, I argue that *mental representations have a distinctive internal structure,* which also allows them to serve as beliefs through which agents control their actions.

What I am proposing here is that the representational notion of a **minimal map** *is what warrants a mentalistic account of operant conditioning.* Specifically, I claim that the existence of a minimal map (i.e. a map-like representation of an animal's current state, goal and pathway to its goal) is what differentiates operant from merely instrumental conditioning. The value of a minimal map is that it allows an animal to control its motor movements.

The map metaphor for belief is by no means new. Its clearest articulation can be found in Ramsey (1990):

> A belief of the primary sort is a map of neighbouring space by which we steer. It remains such a map however much we complicate it or fill in details (1990, p. 146).

Before I explain what I mean by a *minimal* map, I would like to make it clear that I am *not* claiming that all or even most beliefs are map-like representations. What I am proposing here is that Ramsey's account provides us with a useful way of understanding the beliefs which underlie operant behaviour in animals, as well as three other kinds of agency (spatial learning of visual landmarks, refined tool making and social learning, which I discuss in Section C).

What do I mean by a minimal map? Stripped to its bare bones, a map must be able to do three things: it must be capable of showing you where you are now, where you want to go, and how to get to where you want to go. The phrase "where" need not be taken literally as referring to a place, but it has to refer to a specific *state* - for instance, a specific color, temperature, size, angle, speed or intensity - or at least a relatively narrow range of values.

> Definition - "minimal map"
>
> A *minimal map* is a representation which is capable of showing:
>
> > (i) an individual's *current state*,
> >
> > (ii) the individual's *goal* and
> >
> > (iii) a suitable *means* for getting to the goal.
>
> A minimal map need not be spatial, but it must represent specific states.

A *minimal map* can also be described as an *action schema*. The term "action schema" is used rather loosely in the literature, but Perner (2003, p. 223, italics mine) offers a good working definition: "[action] schemata (motor representations) not only represent the *external conditions* but the *goal* of the action and the *bodily movements* to achieve that goal."

A minimal map is more than a mere *association.* An association, by itself, does not qualify as a minimal map, because it does not include information about the individual's *current state.*

On the other hand, a minimal map is not as sophisticated as a *cognitive map.* In a cognitive map, "the geometrical relationships between defined points in space are preserved" (Giurfa and Capaldi, 1999, p. 237).

My description of a minimal map draws heavily upon the conceptual framework for operant conditioning developed by Wolf and Heisenberg (1991, pp. 699-705). In their model, which they developed for the fruit fly *Drosophila melanogaster,* the animal *compares* its *motor output* with its *sensory stimuli*, which indicate how far it is from its goal. When a temporal coincidence is found, a *motor program* is selected to modify the sensory input so the animal can move towards its goal. If the animal *consistently* controls a sensory stimulus by selecting the same motor program, then we can speak of *operant conditioning.*

5.2.5.1 How a minimal map encodes current position

The first component of a minimal map is an *internal representation* of the present value of its motor plan. How might this be realised? Wolf and Heisenberg (1991, p. 699), in their model for *Drosophila melanogaster,* suggest that an insect could maintain an **"efference copy"** of its motor program, in which "the nervous system informs itself about its own activity and about its own motor production" (Legrand, 2001). The concept of an *efference copy* was first mooted in 1950, when it was suggested that "motor commands must leave an image of themselves (efference copy) somewhere in the central nervous system" (Merfeld, 2001, p. 189). However, efference copy cannot simply be compared with the sensory afference elicited by the animal's movement, since one is a motor command and the other is a sensory cue (Merfeld, 2001, p. 189).

Merfeld (2001) has developed a model of the process whereby the nervous system interprets sensorimotor information. Interestingly, Merfeld’s model resembles one developed by Gray (1995) to explain *consciousness:*

> Analogous to Gray's description of his model, ... this model [like mine] (1) takes in sensory information; (2) interprets this information based on motor actions; (3) makes use of learned correlations between sensory stimuli; (4) makes use of learned correlations between motor actions and sensory stimuli; (5) from these sources predicts the expected state of the world; and (6) compares the predicted sensory signals with the actual sensory signals (Merfeld, 2001, p. 190).

One important difference between Merfeld's model and Gray's is that in the event of a mismatch between the expected and actual sensory signals, the *mismatch* is used as an *error signal* to guide the estimated state back toward the actual state. I will discuss the notion of an error signal further in chapter seven.

Recently, Barbara Webb (2004) has reviewed proposals that invertebrates such as insects make use of "forward models", as vertebrates do:

> The essential idea [of forward models] is that an important function implemented by nervous systems is prediction of the sensory consequences of action... [M]any of the purposes forward models are thought to serve have analogues in insect behaviour; and the concept is closely connected to those of 'efference copy' and 'corollary discharge' (2004, p. 278).

Webb discusses a proposal that insects may make use of some sort of "look-up table" in which motor commands are paired up with their predicted sensory consequences. The table would not need to be a complete one in order to be adequate for the insect's purposes. The contents of this table (predicted sensory consequences of actions) would be acquired by learning on the insect's part.

5.2.5.2 Actions and their consequences: how associations can represent goals and pathways on the motor map

The internal representation of an animal's motor commands has to be coupled with the ability to form and remember associations between different possible *actions* and their *consequences*. Heisenberg explains why these associations matter:

> A representation of a motor program in the brain makes little sense if it does not contain, in some sense, the possible consequences of this motor program (personal email, 15 October 2003).

An animal could steer itself by an *internal motor map* which allowed it to directly associate its bodily movements with their consequences. The goal could be stored as a motor memory of the motor pattern associated with the animal's attainment of the reward, while the pathway could be encoded as an ordered sequence of stored motor memories (resulting from the animal's previous exploratory behaviour) of the movements which allow the animal to attain its reward.

An alternative representation by which the animal could steer itself would be an *internal* *sensorimotor map.* Such a map would associate motor movements with consequences indirectly: each move would be associated with a previously remembered sensation (e.g. a colour, shape, sound or smell), which would in turn be associated either with the animal’s reward or the pathway to its reward. For instance, the goal could be encoded as a sensation that the animal associates with the reward, while the pathway would be the sequence of sensations that the animal would experience on the way to its reward, which would allow it to steer its way towards its reward.

The second kind of map would require an animal to possess a multimodal memory. Multimodal memory is surprisingly common in the animal kingdom:

> [A]nimals behaving in a complex, three-dimensional environment receive a large amount of information from external and internal receptor arrays. Clearly, the integration of sensory afference arising from different modalities into a coherent 'gestalt' of the world is essential to the behaviors of most animals (New, 2002, p. 177).

The animal must also possess a **correlation mechanism**, allowing it to find a *temporal coincidence* between its motor behaviour and the attainment of its goal. Once it finds a temporal correlation between its behaviour and its proximity to the goal, "the respective motor program is used to modify the sensory input in the direction toward the goal" (Wolf and Heisenberg, 1991, p. 699; Brembs, 1996, p. 3).

Having fleshed out the notion of a minimal map, we may now formulate the following conclusions:

Conclusion 5.12: An animal must be capable of undergoing *operant conditioning* before it can be said to have mental states.

Conclusion 5.13: Possession of a *minimal map* (as defined above) is a requirement for operant conditioning; hence no animal can be said to have mental states without such a map.

Conclusion 5.14: An animal must possess a *temporal correlation mechanism* before it can be said to have mental states.

5.2.5.3 How do animals follow their minimal maps?

I wish to make it clear that I do **not** regard the animal's map as something *separate from the animal, which it can consult: rather, it is instantiated or realised within the animal's body*.

Nor is the map merely some *internal program which tells it how to navigate.* There would be no room for agency or control in such a picture.

Rather, the map consists of a set of *associations* between motor patterns, sensory inputs and consequences which are formed in the animal's brain. The animal uses these associations to *steer* its way out of trouble. Although we can speak of the animal as updating its internal map, we should think of the animal as observing its *environment,* rather than the map itself.

Although we can say that an animal controls its movements by following its internal map, this should not be taken to mean that map-following is a lower level act. It simply means that the animal uses a map when exercising control over its movements.

5.2.5.4 Areas remaining to be developed in my account of minimal maps

I have proposed that the value of minimal maps is that they allow an animal to steer its way around its environment, using either an internal motor map or a multimodal sensorimotor map. But that leaves us with two outstanding questions: first, how does the steering movement of an animal with a minimal map differ from the physical movements of an animal undergoing instrumental conditioning; second, why are mental states required to explain these movements? I shall develop my answers to these questions in the following two chapters, where I examine the notions of control and self-correction, and their role in intentional agency.

5.2.6 Associative learning: too innate to qualify as cognitive?

In section C of this thesis, I defend the view that one form of associative learning – namely, *operant conditioning,* defined according to Abramson's strong usage of the term – manifests the occurrence of *mental states.* However, Gould (2002) has argued that any process whereby animals learn to form associations is too *innate* to qualify as genuine cognition:

> To most minds ... cognition implies an ability to step outside the bounds of the innate, including the innate wiring that enables animals to learn through classical and operant conditioning. It means, instead, a capacity to perform mental operations or transformations and thus to plan or make decisions (Gould, 2002, p. 41, italics mine).

Gould's argument appears to confuse learning (which is *not* innate) with its mechanism, which is necessarily innate, as Cosmides and Tooby (1997) argue:

> To learn, there must be some mechanism that causes this to occur. Since learning cannot occur in the absence of a mechanism that causes it, *the mechanism that causes it must itself be unlearned* - must be "innate" (1997, italics mine).

Gould's case is further weakened by his damaging admission (2002, p. 44) that even human cognition may not be flexible enough to meet his exacting criteria. All learning, including human language acquisition, is to some extent innately driven, leading Gould to pessimistically conclude that "by the strictest standards, perhaps there is no genuine cognition in any species, our own included" (2002, p. 44) – which is surely a *reduction ad absurdum* for Gould’s argument.

The possibility that some forms of associative learning involve mental processes therefore remains open.

Chapter 6- Mind and Movement: the Significance of Control in the Identification of Intentional Agency

The relevance of *intentional agency* to the question of how we can identify mental acts in non-human organisms is obvious. But what constitutes “agency”, and what kinds of agency qualify as intentional? As most of our actions are bodily movements, the natural starting point for our investigation of agency is to look at the different kinds of movement performed by living things, before we attempt to elucidate the conditions for intentional agency.

6.1 A short summary of the different grades of agency found in nature, from viruses to vertebrates

Case study 1: the lysis-lysogeny “decision” in viruses

The phenomenon of phenotypic plasticity is perhaps the best *prima facie* candidate for agency in viruses. Phenotypic plasticity can be defined as the ability of organisms with the same genotype to vary their developmental pattern, phenotype or behaviour in response to varying environmental conditions (Ancel Meyers, personal email, 18 May 2003).

A well-known case of phenotypic plasticity in viruses is the *lysis-lysogeny decision*, in which parasitic lambda-phage viruses adopt a bet-hedging strategy when they invade a host bacterium: they may either "decide" to kill the host immediately by multiplying until the host's cell walls burst (lysis) or to remain quiescent and confer immunity to infection upon its host (lysogeny). This is a “decision” in a purely metaphorical sense: actually, it is *random* thermal background noise that determines whether the viral DNA is expressed or remains quiescent (Preuss P., 2000). A decision, according to the ordinary English usage of the word, requires rational justification. Moreover, as the article by Preuss makes clear, the behaviour displayed by the viruses is in no way *self-initiated*: it is driven entirely by *external* environmental changes (thermal fluctuations).

Case study 2: directed movement in bacteria

Common bacteria like *E. coli* are capable of directed movement towards or away from the objects they sense. For instance, they swim in chemical gradients towards attractants (e.g. glucose) or away from repellents (e.g. benzoate) - a phenomenon known as *chemotaxis* (Di Primio, Muller and Lengeler, 2000, pp. 4 - 5). Other bacteria display *phototaxis* and *magnetotaxis*, or directed movement in response to light and magnetic fields, respectively (Martin and Gordon, 2001, p. 219).

As we saw in chapter two, there are in fact two kinds of bacterial motion: (a) the random tumbling movements which bacteria initiate in order to probe their surroundings, and (b) the directed "runs" which they make along chemical gradients towards attractants. The mechanisms underlying these behaviours are well-known (Aegerter, 1997; Cotterill, 2001). According to the criteria proposed in chapter four, the directed “runs” of bacteria towards attractants are *fixed* patterns of behaviour.

The occurrence of directed bodily movement in bacteria (and, as we shall see, protoctista and even plants), suggests the following conclusion:

Conclusion 6.1: All cellular organisms are capable of directed movement.

Case study 3: directed movement in protoctista

Various kinds of directed movement, such as *chemotaxis*, *thermotaxis* (movement in response to heat), *phototaxis* and *geotaxis* (movement in response to gravity) are well-attested for protoctista (Martin and Gordon, 2001, p. 409). There does not seem to be any reason to treat these behaviours any differently from the directed movement of bacteria, as they do not appear to be flexible.

Case study 4: directed movement in plants

The case of plants is philosophically interesting, because plants, unlike bacteria and protoctista, are incapable of *locomotion:* they stay where they are put. However, the possibility of cognition among plants should not be ruled out *a priori.* As Di Primio, Muller and Lengeler (2000) put it:

> ...[T]he fundamental thing [for an organism to behave] is not the ability to move to a new location, but the ability to modify itself (by developing effectors as needed), i.e. to respond appropriately to changing conditions (2000, p. 10).

Plants can move in a variety of ways. Whereas a taxis is defined as a movement of a cell in response to a stimulus, a tropism is the directional growth of a plant organ in response to a stimulus such as light *(phototropism),* water *(hydrotropism),* touch *(thigmotropism)* or gravity *(geotropism),* while a nastic movement *(nasty)* is a movement of a plant organ in response to stimuli, that is independent of the direction of the stimuli (e.g. the opening of flowers in response to changes in temperature or light, or the folding up of the leaves of the *Mimosa* plant when touched) (Isaacs, Daintith and Martin, 1999).

While conceding that the movement of plants is "either reduced to the cellular level or rather slow", Di Primio, Muller and Lengeler compare it favourably to that of animals:

> Their abilities to perform undirected (so-called nasties) and directed movements (taxes, tropisms), however, is almost as complex and diverse, and certainly as purposeful, as those of animals. Plants have to solve the same problems as other organisms (2000, p. 10).

There is nothing in the literature, as far as I am aware, to suggest that plant movement is *flexible*, according to the definition used in chapter 4 of this thesis, and as we have seen, research to date suggests that plants are incapable of internally generated flexible behaviour (associative learning).

Case study 5: navigation by animal cells

Albrecht-Buehler (2003a) believes that animal cells possess a kind of intelligence, and criticises the commonly held view that these cells are nothing but "rigidly operating chemical machines that derive their operating instructions internally from their genes and externally from chemicals and electrical signals emitted rigidly by other cells." He has made some intriguing claims regarding the *centrosome*, a spherical area near the nucleus of a cell, which (in animal cells but *not* in most plant cells) contains a pair of cylindrical structures called *centrioles.* Albrecht-Buehler claims that the centrosome is actually the control centre (or "brain") of an animal cell, while the centrioles function as the cell's "eyes". These "eyes" can detect objects and other cells by pulsating near-infrared signals, and steer the cell towards their source. The movement of animal cells differs from the phototactic behaviour of bacterial cells in several significant ways: although bacteria are sensitive to light, they cannot see objects. This means that cells can order and integrate a large amount of visual data.

According to the *narrow* definition of “sense” proposed in chapter 2 (following Cotterill, 2001), the word “sense”, properly speaking, should be restricted to organisms whose sensors are dedicated *receptor cells* which trigger a *distinctive, built-in, rapid-response motor pattern* which is *specific* to the signal and *independent* of the organism's internal state. The centrioles described by Albrecht-Buehler (2003a) do not meet these criteria; nevertheless, their ability to detect objects at a distance places them in a different category from bacteria.

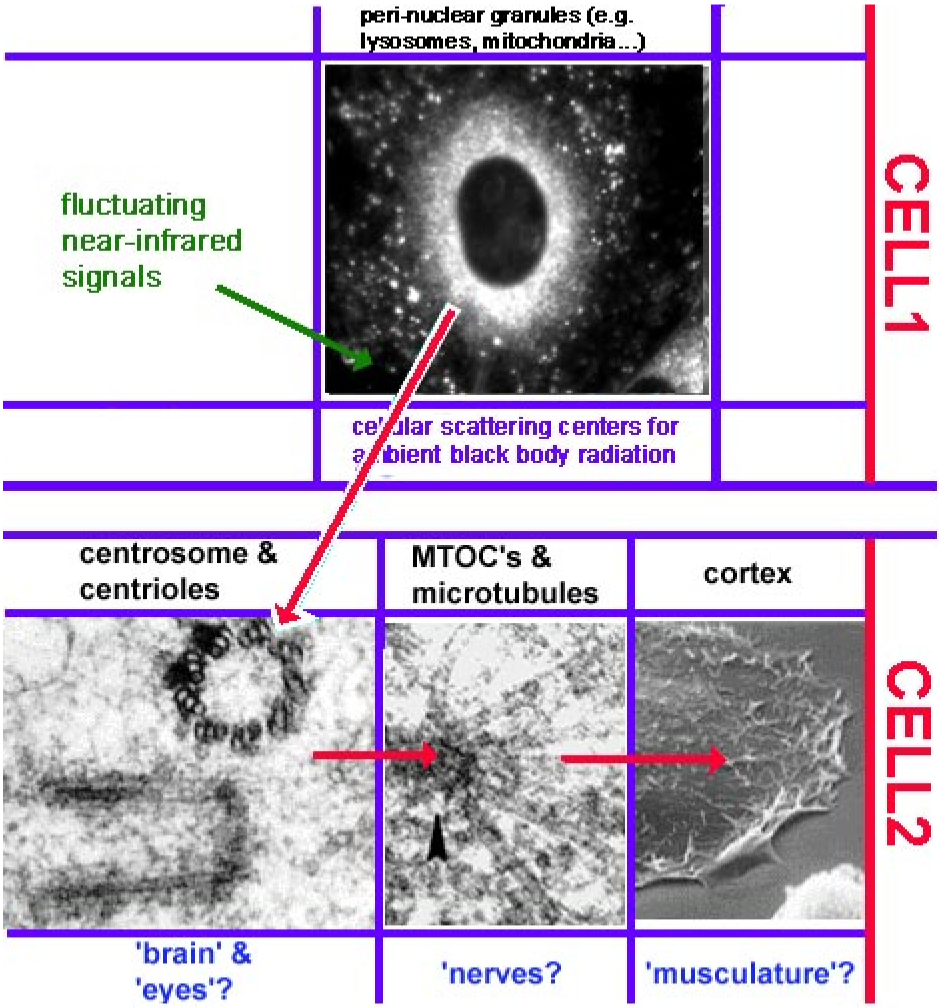


Diagram 6.1 Dr. Guenther Albrecht-Buehler believes that animal cells have “eyes” in the form of centrioles, which are able to detect near-infrared signals and steer the cell towards their source. Illustration taken from Dr. Albrecht-Buehler’s online book, “Cell Intelligence”, at http://www.basic.northwestern.edu/g-buehler/cellint0.htm.

Albrecht-Buehler believes that animal cells can navigate (see diagram), and that this ability is a manifestation of cell intelligence. He suggests that "the best place to start searching [for cell intelligence is] the field of cell movement." As he puts it:

> A moving cell has to operate its own body in sophisticated ways and, in addition, may have to navigate in space and time while dealing with numerous unforeseeable events, such as encounters with other cells and other objects that its genome could not possibly have anticipated. I think that cell motility, indeed, revealed cell intelligence (Albrecht-Buehler, 2003a).

In addition to navigational capacities, animal cells also possess internal movement programs, and are programmed to turn at certain times in their lives at certain angles. In order to do this, they have to be able to measure times and angles. According to Albrecht-Buehler, cells' internal movement programs are not fixed: cells can over-ride them when circumstances warrant it. For instance, when travelling along a "road" made of tiny ridges, they sometimes leave the road, to investigate something "interesting". To do this, they use the distant "clue" to derive a new heading, and follow it. They are even programmed to seek information about their surroundings if they encounter more than one path they can follow. At intersections, the cells extend projections called pseudopodia in all directions, removing obstacles and often changing directions. According to Albrecht-Buehler, cells' internal movement programs are *not fixed:* cells can over-ride them when circumstances warrant it. They are also programmed to seek information about their surroundings if they encounter more than one path they can follow. Individual cells can even co-ordinate their movements: their motility control systems appear to be able communicate with one another about shape changes, direction and timing (Albrecht-Buehler, 2003b).

In chapter four, I defined *flexible behaviour* as the ability to acquire a *new pattern of responding to a sensory stimulus.* Albrecht-Buehler's claim that the cells' internal movement programs are not fixed might seem to suggest that they are flexible. However, another more likely interpretation, hinted at by Albrecht-Buehler, is that there are different *levels of control* within each cell. We could think of each cell as having a master program that governs its movements. The specific functions that happen to be activated by the program may vary as environmental conditions change, but the program coding for the functions themselves, which describe the cells' patterns of responding to sensory stimuli, need not vary over time, nor do the program instructions need to change. If this is so, then the cells' behaviour would *not* qualify as flexible, according to the definition I am using.

There is no evidence to date that the cells have a *memory* - although this is partly due to the difficulty of designing an experiment that would prove it. Nor is there any evidence that the cells have a "cognitive map" of their surroundings (Albrecht-Buehler, personal email, 30 September 2003).

Case study 6: action selection in cnidaria

The simplest phylum of "true" animals *(Eumetazoa)* is the *Cnidaria* (commonly known as coelenterates). Cnidaria have no brain, but they have a rudimentary nervous system, with neurons positioned regularly over the surface of the animal. Each neuron is in contact with its neighbours (Abramson, 1994, p. 176).

Although cnidaria do not possess a central nervous system or a brain, their nerve net permits rapid communication between cells (in some cases taking only milliseconds), over relatively long distances.

Prescott (2007) considers the behaviour of cnidaria to be an important advance over that of bacteria, plants and “simple” animals such as sponges. Unlike bacteria, cnidaria are multicellular creatures, which face the task of co-ordinating their entire bodies in response to sudden changes in their environment. For instance, swimming in jellyfish requires the "synchronous, simultaneous contraction of the entire perimeter of the bell" (Prescott, 2007, p. 7). And while plants, like cnidaria, are multicellular, plants are not motile. Finally, cnidaria possess biologically significant features that “simple” animals such as sponges do not - in particular, a nervous system which (in a few cases) allows very fast signal conduction as well as (possibly) reflexes, and enables fast attack, escape, or self-defence reactions. Cnidaria exhibit "internally generated, rhythmic behavior, and co-ordinated patterns of motor response to complex sensory stimuli", allowing them to display an "integrated global response" to their environment (Prescott, 2007, pp. 5, 7). In some cnidaria, such as the hydrozoan jellyfish, the nerve net is arranged in a longitudinal circuit which supports fast attack, escape and defense reactions (Prescott, 2007, pp. 6-7). By contrast, sponges respond only to direct stimulation, at a very slow rate (about twenty minutes). As sponges lack neurons, internal communication can only occur between neighbouring cells.

Prescott considers the nervous system found in cnidaria to be a fundamental advance in the evolution of what he calls "action selection" or the problem of "resolving conflicts between competing behavioural alternatives" (2007, p. 2). We might define an *action selection mechanism* in an organism as a repertoire of actions, combined with the ability to select the most appropriate one for the present circumstances. By itself, this definition does not take us very far, as action selection is part of a problem faced by *all* living creatures: behavioural integration, or the task of co-ordinating the activities of their parts and sub-systems. Even bacteria and plants possess this ability (Godfrey-Smith, 2001, pp. 6-7).

Nevertheless, Prescott (2007) contends that the speed and co-ordination of the way in which some cnidarians respond to stimuli represents a milestone in the history of action selection. In some jellyfish, the nerve net is functionally divided into two relatively independent systems - one for feeding and the other for movement - which interact in neuron clusters. Others possess a single nerve net, which can carry two different types of action potentials, enabling either rapid escape swimming (to avoid predators), or slow rhythmic swimming for feeding (Prescott, 2007, pp. 5-7).

Prescott likens this decentralised neural arrangement to the subsumption architecture described by Brooks (1986). A Brooksian "agent" has no central control: it is hierarchically organised from the bottom up. Control is distributed between different components, making the "agent" better able to withstand damage (i.e. more robust). Behaviour patterns are hard-wired, and sensors and actuators (which produce movement) are closely coupled, to allow rapid response times. Co-ordination between the different components is ensured by built-in timers and by having behaviour modules that can inhibit one another. Simple behaviours combine to produce more complex patterns of behaviour (Laird, 1994). According to Prescott, the functional subdivision of the nerve net into two distinct circuits for feeding and movement, which is found in some jellyfish and sea anemones, resembles the Brooksian architecture proposed for some behaviour-based robots (2007, p. 6).

A Brooksian "agent" has a very "low-tech" design. It has no internal model of the outside world, and does not engage in planning or learning of any kind. All of its behaviour is hard-wired and built-in, to ensure co-ordination and cope with unforeseen contingencies. In other words, its patterns of behaviour are fixed. If cnidaria do indeed behave like Brooksian "agents", then they cannot learn new ways of responding to unforeseen events. In other words, they *lack* an internal mechanism enabling them to modify their behaviour patterns and learn to do something new or different, as occurs in *associative learning* (Abramson, 1994, p. 38). Habituation has been documented in cnidaria (Encyclopedia Britannica, 1989), but I have not found any evidence for associative learning in cnidaria in the course of my thesis research.

Case study 7: centralised action selection in flatworms

Flatworms are considered to be the most "primitive" phylum of worms. The simplest central nervous systems are found in flatworms (platyhelminthes). Flatworms also have an important evolutionary significance, as one particular sub-group – the acoel flatworms - is thought to resemble the common ancestor of all animals with bilateral symmetry (Prescott, 2007, p. 12).

Prescott considers the appearance of the platyhelminthes in the fossil record (565 to 544 million years ago) to be the next major breakthrough in the evolution of action selection, after the evolution of cnidaria. He cites research by Raup and Seilacher (1969, cited in Prescott, 2007, pp. 9 - 12) showing that trace fossils of meandering foraging trails left by the earliest animals possessing bilateral symmetry can be generated by combining four simple behaviour mechanisms, one of which functions as a centralised conflict-preventing mechanism, of vital importance to an organism with a primitive brain and bilateral symmetry. However, robotics researchers such as Mataric (1990, cited in Prescott, 2007, p. 9) have pointed out that the other three mechanisms imputed to these flatworms are identical to those used by a robotic wall-follower, while the conflict-preventing mechanism can be generated simply by transferring control of staying close to the wall from one sensor arm to the other.

Although flatworms' action selection mechanisms (unlike those of cnidaria) are centralised, this does not imply that they are flexible in the sense we defined in chapter four. As we saw in chapter five, there is no good evidence to date that flatworms (planarians) are capable of associative learning.

> Conclusion 6.2: The appearance of multicellular organisms, a primitive nervous system and a centralised nervous system represent important milestones in the history of action selection. However, none of these milestones entails a capacity for internally generated flexible behaviour, which seems to have arisen later in evolutionary history.

On the other hand, some worms (such as the well-studied nematode *C. elegans*) are clearly capable of undergoing classical conditioning. It was argued in chapter five that their behaviour qualifies as flexible and internally generated.

Prescott (2007) regards the *vertebrate brainstem* as the next great leap forward in the history of action selection, in the evolutionary lineage leading to vertebrates. So-called “advanced” invertebrates such as insects and cephalopods belong to a separate lineage from the one leading to vertebrates; Prescott ignores them in his account, as he is concerned with the evolution of action selection in vertebrates, whose nearest evolutionary relatives have relatively simple nervous systems. Prescott marshals evidence that the basal ganglia in the vertebrate brainstem play a crucial role in action selection. Additionally, neurobehavioural data from rats suggests that the *reticular formation* (a complex neural network in the central core of the brainstem) provides a brainstem substrate for action selection in the *vertebrate* central nervous system. Humphreys, Guerney and Prescott (2005) propose that the reticular formation evolved to provide effective arbitration between innate behaviors. I shall not attempt to chronicle the evolutionary history of action selection any further at this point, as one of the central concerns of this thesis is to identify the criteria for a *minimal mind* in nature; I also wish to examine whether the complex behaviour exhibited by some of the "higher" *invertebrates* (such as fruit flies, honeybees and octopuses) qualifies as *intentional agency.*

6.2 Different grades of agency in nature: their relevance for the possession of mental states

6.2.1 What are the conditions for intentional agency?

The foregoing account of the lysis-lysogeny "decision" in viruses points to two conditions that must be met before we can identify intentional agency (and hence mental states) in organisms:

> Conclusion 6.3: Behaviour by an organism must vary in a *non-random* manner before it can be regarded as a manifestation of a mental state.

> Conclusion 6.4: Behaviour by an organism must vary in response to its *internal* states, as well as external conditions, before it can be regarded as a manifestation of a cognitive mental state.

It was argued in chapter one that our ability to describe an entity's behaviour according to Dennett's *intentional stance* is a necessary condition for our being able to ascribe cognitive mental states to it. The intentional stance characterises the behaviour of an organism as *movement* towards a goal, which it has information about. For this reason, directed movement is a requirement of intentional agency:

> Conclusion 6.5: An organism must be capable of directed bodily movements before these movements can be regarded as a manifestation of a cognitive mental state.

The term "directed movement" is not meant to suggest whole-body locomotion, however. There is no inherent reason why the absence of locomotion in plants, fungi and certain animals should necessarily preclude the possibility of cognition on their part, as Di Primio, Muller and Lengeler have cogently argued (2000, p. 10):

> Conclusion 6.6: A capacity for local movement (locomotion) in an organism is not a requirement for its possession of mental states.

In case study 5, we looked at navigation by animal cells, but the term “navigation” was left undefined. I would like to propose the following definition:

> Definition - "navigation"
>
> Any organism that can use its *senses* to steer itself *or a part of its body* around its environment is capable of navigation.

This definition is meant to lend greater precision to the term “directed bodily movement” used above. I have added the phrase "or a part of its body" to the definition of navigation, to allow for the possibility of navigation in organisms such as plants, which can move their body parts but are incapable of *locomotion.* The definition used here can be construed broadly: it could include, for instance, the probings of bacteria, whereby they find their food. However, just as we can distinguish between broad and narrow meanings of the word “sense”, we can do the same for navigation. In chapter two, it was argued that the possession of *senses* – according to the *narrow* meaning of the term defined by Cotterill (2001) – was a necessary condition for the possession of mental states. Combining this with conclusion 6.5 and the definition above, we may conclude:

> Conclusion 6.7: An organism must be capable of navigation in the *narrow* sense (i.e. where its movements are guided by “true” senses) before its movements can be regarded as a manifestation of a cognitive mental state.

As we saw above, the advantage of having a *centralised* action selection mechanism is that it serves as a conflict-preventing mechanism, of vital importance to an organism with a primitive brain and bilateral symmetry. From a purely *a priori* standpoint, there is no obvious reason why we should regard this feature as a necessary condition for possessing mental states: we can certainly imagine a creature that has a mind, but lacks such a mechanism. However, we rejected such “thought experiments” as a legitimate tool for investigating mental states in the Introduction. It was argued in chapter five that a capacity for *associative learning* was a requirement for being able to justifiably impute mental states to creatures, and it is an *empirical* fact that only organisms with a brain and bilateral symmetry are capable of associative learning. *These* organisms require a central mechanism to prevent conflicts between the left and right sides of their brains.

> Conclusion 6.8: An organism must possess a *centralised action selection mechanism* before its movements can be regarded as a manifestation of a cognitive mental state.

6.2.2 Which of the various kinds of agency found in nature warrants the description of "intentional"?

Having identified some of the necessary conditions for intentional agency, we need to address the issue of whether they are *sufficient*, either singly or collectively, to warrant the ascription of mental states to other creatures*.*

It might be argued that any organism with the ability to select the most appropriate action for the circumstances is an agent of sorts, as it is choosing the best means of achieving its ends: "This action, not that one, will get me what I want". However, the selection need not involve the organism having its own internal *means-end schema.* A selection can be triggered by incoming sensory *information* (a stimulus). The *goal* achieved is what the action is for, but the organism making the selection does not need to know that, if its behavioural programs are working properly. Nor do these programs need to be flexible; as we saw above, the existence of an action selection mechanism is compatible with *fixed* patterns of behaviour. As flexibility is not required for action selection, I conclude that it can be adequately described using a *goal-centred* intentional stance.

We saw above that the ability of bilaterally symmetric animals (such as worms) to engage in *centralised* action selection provides no guarantee that they are capable of undergoing associative learning and displaying internally generated, *flexible* response patterns: flatworms seem to lack this capacity. Since it has been argued in chapter five that there can be no "true" learning (and hence no mental states) in an organism unless its behavioural response patterns are flexible, we may conclude:

> Conclusion 6.9: In the absence of a demonstrated capacity for associative learning,

> even the combination of centralised action selection, a central nervous system, navigation and “true” senses in an organism does not provide a sufficient warrant for the ascription of mental states to it.

It was argued in chapter five that even a capacity for associative learning does not entail the presence of mental states. Nevertheless, some philosophers (Dretske, 1999) have proposed that a particular kind of associative learning – *operant conditioning* – presupposes the occurrence of mental states. The definition of “operant conditioning” remains controversial, as we saw in chapter five. I suggest that the definition could be refined, by invoking the concept of *fine-tuning.*

6.3 The significance of fine-tuning for intentional agency

Many animals are capable of learning to perform certain actions for the sake of the consequences they learn to associate with them (e.g. the attainment of some reinforcement). Most psychologists refer to such behaviour as operant behaviour, but Abramson calls this kind of behaviour "instrumental conditioning", and reserves the term "operant behaviour" for a special sub-set, which he defines as "the ability to operate some device - and know how to use it, that is, make an arbitrary response to obtain reinforcement" (1994, p. 151, italics mine). As we saw in chapter five, some scientists find the distinction between instrumental and operant conditioning unacceptably vague. In an email exchange, Abramson provided me with more specific examples of behaviour conforming to his more restrictive definition:

> For example, we know that rats can be taught to press a lever in various directions and *with various degrees of force.* They can also be trained to run down an alley *with speeds selected by the experimenter*... (personal email, 2 February 2003, italics mine).

I would suggest that Abramson's description of operant behaviour could be re-expressed using the concept of *fine-tuning.* What the rats in the above examples are doing is confining some parameter that describes their movement (speed or force exerted) *within a narrow range*, in order to achieve some goal that they have learned to associate with staying within that range*.* Because it arises through a learnt association between the performance of an action and the obtaining of a goal (e.g. a reward, or the avoidance of punishment), fine-tuning is a *flexible* and *internally generated* behaviour.

More precisely, fine-tuning can be regarded as a (learned) r*efinement* of *action selection*, in which the animal stabilises a basic motor pattern in its repertoire *at a particular value* or confines it *within a narrow range of values*, in order to achieve a *goal* that it has learned to *associate* with the action.

If the animal consistently confines the same motor pattern within the same range in order to obtain a goal, then we can say it has undergone a form of learning: *operant conditioning.*

On the account I am defending here, *operant conditioning* can be regarded as a process of instrumental conditioning in which an animal learns to stabilise (fine-tune) an existing motor pattern from its repertoire by confining it within a narrow range in order to obtain some goal.

The notion of "fine-tuning" can be tied to another related concept: that of *control.* If an animal can adjust and fine-tune its responses to a variable stimulus, then we can say that it has *control* over its responses, and is hence able to engage in *operant behaviour.* In ordinary discourse, the notion of acting intentionally is closely tied to the notions of trying and control.

In chapter five, I referred to a conceptual framework developed by Wolf and Heisenberg (1991, pp. 699-705), which I used to elucidate the concept of a minimal map. Since the framework also has relevance for the notion of control, I shall quote Wolf and Heisenberg's description directly:

> On the basis of these results a conceptual framework of operant behavior is proposed: (1) It requires a goal (desired state) of which the actual state deviates. (2) To attain the goal a range of motor programs is activated (initiating activity…) (3) Efference copies of the motor programs are compared to the sensory input referring to the deviation from the desired state (e.g. by cross-correlation). (4) In case of a significant coincidence the respective motor program is used to modify the sensory input in the direction towards the goal. (5) Consistent control of a sensory stimulus by a behavior may lead to a more permanent behavioral change (conditioning). In this scheme operant activity (1-4) and operant conditioning (1-5) are distinguished (Wolf and Heisenberg, 1991, online abstract).

In their model, the animal *compares* its *motor output* with its *sensory stimuli*, which indicate how far it is from its goal. When a temporal coincidence is found, a *motor program* is selected to modify the sensory input so the animal can move towards its goal. If the animal *consistently* controls a sensory stimulus by selecting the same motor program, then we can speak of *operant conditioning.*

In the above account, the animal selects one motor program from the range available to it, and uses its selection to steer itself towards its goal. It should be quite clear that the selection of one option from a limited range of choices – just four in the case of a tethered fruit fly (Heisenberg, Wolf and Brembs, 2001, p. 2) – cannot be considered “fine tuning”, as the element of fine-grained motor control is absent here. Fine-tuning must occur subsequent to the selection of an appropriate motor program, when the fly refines its selection and thereby exercises **control over its bodily movements.**

I hypothesise that the animal uses a *minimal map* of some sort to accomplish this – i.e. an *internal motor map* or *sensorimotor map*, as described in chapter five. Also, because the animal does not use this map when undergoing *instrumental* conditioning, I predict that scientists should be able to find a *clearcut distinction* (which should be detectable on a neurological level) between operant conditioning and merely instrumental conditioning.

In the fine-tuning process described by Wolf and Heisenberg, there is a continual inter-play between an animal’s "feed-back" and "feed-forward" mechanisms. According to Webb, “the essential idea [of forward models] is that an important function implemented by nervous systems is *prediction of the sensory consequences of action”* (2004, p. 278, abstract). Additionally, Wolf and Heisenberg (1991, p. 699) suggest that the fly maintains an **"efference copy"** of its motor program. An efference copy is one in which "motor commands must leave an image of themselves …. somewhere in the central nervous system" (Merfeld, 2001, p. 189). In Merfeld’s (2001) model of efference copy, the body “(1) takes in sensory information; (2) interprets this information based on motor actions; (3) makes use of learned correlations between sensory stimuli; (4) makes use of learned correlations between motor actions and sensory stimuli; (5) from these sources predicts the expected state of the world; and (6) compares the predicted sensory signals with the actual sensory signals” (Merfeld, 2001, p. 190). In the event of a mismatch between the expected and actual sensory signals, the *mismatch* is used as an *error signal* to guide the estimated state back toward the actual state. I shall return to this point in chapter seven, when I discuss self-correction.

I propose that the notion of intentional agency presupposes that of *control*, which is manifested in *fine-tuning* behaviour as described above. (It will be argued in Section C, chapter nine, that there are at least four distinct kinds of behavioural manifestations of "fine-tuning".) Specifically, I claim that *an organism must be capable of fine-tuning its bodily movements before it can be identified as an agent, capable of having mental states.*

To see why, it will be helpful to examine Carruthers' (2004a) argument against the possibility of attributing minds to animals, solely on the basis of what they have learned through conditioning:

> ...engaging in a suite of innately coded action patterns isn't enough to count as having a mind, even if the detailed performance of those patterns is guided by perceptual information. And nor, surely, is the situation any different if the action patterns aren't innate ones, but are, rather, acquired habits, learned through some form of conditioning (2004a, pp. 4-5, online PDF version).

An organism with an action selection mechanism has a pre-existing repertoire of actions, from which it is able to select the most appropriate one for the present circumstances. In *instrumental conditioning,* the organism learns that performing one of the actions in its repertoire, in a given set of circumstances, will enable it to achieve one of its goals. We can account for this behaviour parsimoniously by supposing that the associations it has formed (encoded and stored as *information* in its brain) bias its internal action selection mechanism, causing it to perform the action that will in fact obtain for it something it seeks (its *goal*). The association between action and goal is a *direct* one: there is no need to invoke a *means-end* schema to account for it, hence there is no extra explanatory work for an agent-centred explanation to do. Although the behaviour of the animal is flexible, a *goal-centred* intentional stance appears adequate to explain the animal's behaviour.

By contrast, in operant conditioning, the animal not only selects an action mechanism, but also refines it by stabilising it within a range. The association between the basic motor pattern and the goal is an *indirect* one (select -> refine -> goal). There is room here for a *means-end* schema: we might postulate that the animal *believes* that it can attain what it *desires by controlling* its motor behaviour (an agent-centred stance). The act of control is viewed as a *means* whereby the animal achieves one of its desired ends. It should be noted that I am not arguing here that fine-tuning is a *sufficient* condition for intentional agency – merely that it is a *necessary* one:

> <u>Conclusion 6.10:</u> The ability of an animal with a centralised action selection mechanism to learn from *instrumental conditioning* does not, of itself, warrant the ascription of cognitive mental states to it.

> <u>Conclusion 6.11:</u> An organism must be capable of *fine-tuning* its bodily movements before it can be identified as having cognitive mental states.

6.4 Which organisms are capable of fine-tuning?

Abramson (1994) describes the nervous system of cnidaria:

> Neurons are located regularly over the surface of the animal... The propagation of a nerve impulse is not transmitted along a linear chain of neurons, but radiates from its point of origin... The effect of such an arrangement is that a stimulus applied to any part of the animal will be directed to all parts, much like sticking your finger in a cup of jello will make the whole mass move... *Such a system is not conducive to fine control of motor movements* (1994, p. 176, italics mine).

The foregoing account provides us with an additional reason for thinking that mental states are confined to animals with central nervous systems.

> Conclusion 6.12: Only organisms with central nervous systems are capable of fine-tuning their bodily movements for the performance of intentional acts.

A corollary of this result is that we can only ascribe cognitive mental states to organisms with central nervous systems.

Some questions still need to be addressed. Exactly how are we supposed to identify fine-tuning? And how do we know if an organism is exercising "control" over its patterns of behaviour? I will discuss these matters further in the next chapter, where I discuss self-correction.

Chapter 7 - Getting it Wrong: The Centrality of Self-Correction to Belief

The focus of this chapter is a naturalistic, non-biological account of intentionality recently put forward by Beisecker (1999, 2000), who has proposed empirical criteria that would enable us to identify which creatures are capable of mental acts. Although I take issue with Beisecker’s criteria, I believe he has done philosophy a valuable service, both in exposing the inability of a purely biological account of intentionality to account for mental acts, and in attempting to provide a scientifically rigorous account of how we could discern mental activity to creatures on empirical grounds, without having to first establish whether they are phenomenally conscious.

7.1 The relevance of non-biological self-correction for having mental states

Whereas for Dretske (1999) the salient feature of beliefs is that they are internal representations, acquired through learning, which bring about changes in organisms' behaviour by virtue of their meaning, for Beisecker the defining quality of beliefs is that they can be correct or mistaken. Beisecker (2000) adopts an avowedly Searlian account of intentionality at the outset:

> *Normativity* must lie at the heart of any satisfactory account of intentionality. Beliefs are ground-level intentional states, and believers are *essentially* things that can be *correct or mistaken* with respect to the way things are. Thus you don't have anything that warrants being called an *account* of belief (or doxastic states more generally) without an account of error (and correlatively, of correctness). Similarly, any intelligible account of desire (or any other conative state or "pro-attitude") requires a story about how to identify conditions of *satisfaction* - understood as possibly non-actual states of affairs that might not ever obtain, which a subject is *in some sense* disposed or inclined to bring about "all things being equal." Such conditions of correctness and satisfaction are,

> of course, usually identified as the *content* of an intentional state. Hence the intuition that intentional states are inherently contentful (as well as the attendant puzzles behind such contentfulness) can be understood as expressions of this commitment to the essential normativity of the intentional. While other kinds of intentional states (e.g., curiosity, envy, or pride) might not possess evident "directions of fit" – their intentionality is likely to be explained in terms of the intentionality of beliefs and desires that they presuppose (Beisecker, 2000, online).

Contemporary *teleo-biological* accounts of intentionality construe an organism's "getting it wrong" as some sort of organ or system *malfunction* on the creature's part. Animals ought to seek food, attract mates and flee predators; when they fail to do these things, they do something which harms their prospects of being able to pass on their genes to the next generation. Such malfunctions can be considered as errors.

Beisecker argues that this form of intentionality, while genuine, is the *wrong sort* of intentionality to invoke, in order to account for mental activity in animals. The intentionality captured by biological accounts is not intrinsic but *derived*, like that of artifacts. Beisecker contends that these norms are not those of the creature itself but those of Mother Nature, or the interests of the entire species. Also, a biological account of intentionality is too generous to account for mental activity as such; indeed, as Beisecker observes, biological norms can be applied to organisms as simple as bacteria.

While I concur with Beisecker's rejection of biological norms as a criterion for mental activity, I believe his criticism is slightly off the mark here. If a creature is alive, it can be said to have a good of its own (Cameron, 2000, pp. 331-335). Insofar as a set of biological norms describe that creature's good, they can be considered *intrinsic*, rather than derived. However, Beisecker is quite right in pointing out that these norms are *generic*, as they pertain to the good of the species to which the creature belongs, rather than what is good for this *particular* creature, here and now. In other words, the crucial feature which makes them the wrong sort of norms to characterise the mental states of a particular creature is simply the fact that they fail to refer to any particular individual.

Another vital difference between biological intentionality and that which characterises mental activity is that in the case of biological intentionality, there is no *acknowledgement of errors* on the individual's part:

> [C]onsider the poster child of the biological approach, the frog of philosophical legend. The thing that is so striking about the frog is that it isn't impressively responsive to its biologically discerned mistakes. Other than by simply breaking down, it doesn't *acknowledge* them *as mistakes*, say, by *correcting* itself in the face of persistent error. Like slot machines, the frog doesn't compellingly *acknowledge* the norms by which it can be evaluated, as one might reasonably require, if it were to be capable of getting things right or wrong "by its own lights" (Beisecker, 2000, online).

The key intuition underlying Beisecker's argument at this point is that human beings (and other creatures with minds) have beliefs and desires *of their own* and can therefore exhibit a normativity that is *independent of* and even contrary to their *biological ends.* This normative autonomy, he asserts, is what accounts for the *original* intentionality of their beliefs and desires.

Biological accounts, according to Beisecker, describe an intentionality of sorts, but they fail to explain how it can be said to be good for this particular creature, or why it should be considered the only kind of intentionality. Accordingly, Beisecker has attempted to describe a *non-biological* form of intentionality, which has survival value to its possessor (allowing it to have arisen naturally through the process of evolution), but whose standards do *not* invoke biological norms.

But how can we discern this kind of intentionality in other creatures? First, Beisecker suggests that we search for *self-corrective* behaviour in creatures which is directed towards some *goal:*

> Our discussion of the frog of philosophical legend suggests that in order to capture anything that deserves to be regarded as behavior appropriately governed by an *acknowledgement* of a norm, we will need to have some account of self-correction. But in order to sustain the claim that such activity is appropriately regarded as involving the *correction* of *errors*, it would seem that we need to have some account of a subject's aims as well, for how could we recognize mistakes *as mistakes* unless they are somehow liable to prevent a subject from attaining its desired ends? That is, it would be difficult to tell a story with the requisite normative punch without including some account of goals. It is thus reasonable to suppose that discernibly rational activity requires elements of both *critical* (self-corrective) and *practical* (means-end) reasoning. So to a first approximation, *I propose we regard a pattern of activity as discernibly rational if it exhibits self-corrective behavior that is directed towards some end or goal* (Beisecker, 2000, online, italics mine in last sentence).

Beisecker's reasoning allows us to formulate a necessary condition for having mental states:

<u>Conclusion 7.1:</u> An organism must be capable of self-correcting goal-directed behaviour before it can be said to have cognitive mental states.

The behaviour of animals that undergo associative learning would meet this requirement; they are “educable”, insofar as they can “tailor their own responsive dispositions to their particular surroundings” (Beisecker, 2000, online). By contrast, bacteria containing magnetsosomes, which we discussed in the previous chapter, do not tailor their responses in this way; hence they should not be described as being "in error" when they move toward the bar magnet instead of the bottom of the water. Thus they do not qualify as holders of expectations or beliefs.

While the behaviour of creatures that undergo *associative learning* is clearly more sophisticated than that of the bacteria, it nevertheless fails to exhibit the kind of *intentionality* Beisecker is looking for: intentionality that can be understood, *independently of* the purposes for which the creatures have been naturally selected or designed. The creatures learn from their mistakes, but we can only recognise them as such by appealing to the creatures’ *built-in* biological ends.

To resolve this impasse, Beisecker outlines an account of how creatures can exhibit a *non-biological* form of intentionality which explains their mental capacities in a way that is compatible with evolution by natural selection. Some learning theorists have argued that the educability of certain animals is best explained by attributing to them *expectations* that their responses to certain kinds of events will bring about certain outcomes. These expectations can be regarded as *structures* that mediate between the creatures’ sensory input and their behavioural output. Insofar as a creature engages in behaviour expected to bring about a certain outcome, we may regard that outcome as one of *its* goals. Beisecker calls these goals *non-biological*, because we can identify them without having to know anything about the creature's biological evolution, which has determined its built-in ends through natural selection. Of course, expectations may be disappointed: a creature may make errors of *commission* (when an animal's expectation - say, of obtaining food - is activated and it responds, but the expected consequence does not eventuate) and errors of *omission* (when the animal fails to respond because its expectation is not activated, but in fact, the response would bring about a desired consequence). However, because such a creature continually revises its expectations while leaving correct expectations as they are, the creature is taking the realisation of its expectations to be a *regulative ideal* or norm. Thus it can be said to possess a kind of *critical rationality:*

> [U]nder certain circumstances, an expectation will be activated and the creature will then anticipate that a certain response will yield a particular outcome. Should that turn out not to be the case, its dispositions to form such anticipations will change. *Expectation-mongering creatures* can now be defined as those whose overall behavior is most systematically described as governed in part by the consequence conditions of currently activated expectations. For example, a creature might be disposed to engage in any responses associated with the outcome of its acquiring cookies, or it might be disposed not to engage in any response associated with electric shocks (Beisecker, 2000, online).

Beisecker concludes that “[s]ince they can be evaluated as having gotten things right or wrong, we are justified in crediting these creatures with some sort of intentional capacity” (1999, p. 303).

Beisecker believes that his expectations-based account of error has three appealing features. First, unlike teleo-biological accounts, it *doesn't require us to identify the creature's biological ends* before we can recognise its mistakes. A certain outcome can be regarded as one of a creature's *goals*, "to the extent that the creature is disposed to engage in responses *expected* to bring about that outcome" (Beisecker, 2000, online). Second, the standards for correctness are not relative but *categorical,* insofar as they apply, irrespective of the particular goals a creature might possess. Finally, creatures can get things right or wrong *in a variety of respects* (feature selectivity), as they can be correct with respect to some features of their environment but mistaken with respect to others.

Beisecker's expectation-generated account of intentionality even includes a feature that might be construed as *sensitivity to intensional context*: an animal's expectation of goal A and its expectation of goal B might happen to have the same conditions of satisfaction, yet the two goals are quite distinct.

Although Beisecker does not explicitly treat of the question of whether animals with expectations are *phenomenally conscious*, the fact that Beisecker is willing to describe animals whose expectations are disappointed as "surprised" (see section 7.2) suggests that he thinks they are. Beisecker does not discuss the possibility that a creature could possess internal representations that warranted the appellation of "expectations", even though the animal in question lacked the neurological wherewithal for phenomenal consciousness.

We have yet to discuss the criteria that would allow us to ascertain which creatures have expectations. However, we may formulate the following negative conclusion:

Conclusion 7.2: Purely biological forms of intentionality cannot serve as a sufficient warrant for the attribution of mental states to other creatures.

7.2 Does the phenomenon of blocking enable us to identify creatures with expectations?

Beisecker not only develops an interesting *non-biological* account of intentionality which could explain how non-human animals can be said to have mental states, but he also suggests an empirical criterion that would allow us to identify creatures with expectations: the phenomenon of *blocking,* which often arises in the context of operant conditioning.

As we saw in chapter five, there are differing schools of thought as to whether *operant conditioning* should be envisaged in mentalistic terms or purely causal terms. However, one phenomenon which is commonly exhibited by creatures undergoing operant conditioning – namely, *blocking* – has been proposed by Beisecker as an example of behaviour that *expectation-generating* animals would engage in. Blocking is a highly unusual prediction of one model of associative learning - the *Rescorla-Wagner model* - whose central ideas are usually expounded using mentalistic terminology. While the model is not able to explain all phenomena connected with classical conditioning, it is still regarded as "the 'best' theory of classical conditioning" (Jackson, 2002, online), although it has undergone several refinements. For a discussion of various models of conditioning, see Schmajuk, 2007; for an outline of a radical alternative model, see Gallistel and Gibbon, 2001.

The basic principle of the Rescorla-Wagner model is that "the amount of conditioning depends on how *surprising* the association between the CS and US is. The more unexpected or surprising the US, the more conditioning will occur" (Lipp, 1998, online, italics mine). The word “surprise” appears to suggest that we are dealing with a mental state here, and a phenomenally conscious one at that. Beisecker describes how the Rescorla-Wagner model accounts for blocking:

> The model readily explains the phenomenon of *blocking*, in which animals that have been trained to associate a conditioned stimulus (e.g. a bell tone) with an unconditioned stimulus (e.g. an electric shock) will *fail* to associate a second conditioned stimulus (e.g. a red light) with the unconditioned stimulus, if the latter is subsequently presented along with the original conditioned stimulus (the bell tone). The model explains this curious fact as follows: initially, the animals associate the bell tone with the electric shock because the co-occurrence of the two events *surprises* them. The addition of a second conditioned stimulus generates *no new unpleasant surprises* for the animals, so they do not make any new associations. What the model seems to suggest is that animals undergoing operant conditioning form new expectations of what will happen, and hence have beliefs. *If this is correct, the phenomenon of blocking could be used to distinguish those animals which are capable of forming expectations from those which are not* (Beisecker, 1999, pp. 298-299, italics mine).

There are two comments that I wish to make here. First, most of the various kinds of animals that are capable of operant conditioning have *not* yet been shown to exhibit blocking. Operant conditioning is, as we have seen, well-documented for many kinds of insects, and there is even tentative evidence that worms such as *C. elegans* may be capable of it. By contrast, the occurrence of blocking even in so-called "higher" invertebrates (honeybees) remains controversial.

In response to a query as to whether blocking has been confirmed yet in any invertebrates, Dr Bjorn Brembs (personal email, 22 December 2002) wrote:

> As far as I have heard, the jury is still out, whether there is blocking, although those that have found it still claim that there is no dispute about their data. There are a few finds, but alternative explanations have not been ruled out, yet. So far, blocking, if it is there, is definitely not as universal and general as in vertebrates, at the least.

More recently, Guerrieri *et al.* (2005) acknowledged that “[t]he question of whether blocking exists in olfactory conditioning of proboscis extension reflex (PER) in honeybees is under debate” (2005, online abstract). The authors claimed to find evidence of blocking, but concluded that “blocking is not a consistent phenomenon, nor does it depend on odor similarity” (2005, online abstract).

If Beisecker wishes to propose *blocking* as a litmus test of cognitive mental states, then he will have to limit these states to vertebrates for the time being.

Second, blocking may turn out to be explicable in *non-mentalistic* terms, in any case. Some scientists use the term "expectation" in a mind-neutral sense, to denote a memory retrieval function that results from former learning. Some even attempt to explain blocking as the result of peripheral sensory integration (Menzel, personal email communication, 21 July 2003). We have already seen that sensory capacities and memory can be explained by adopting a *goal-centred* intentional stance. If many scientists believe that these capacities are sufficient to explain blocking, then *(pace* Beisecker) it would be unwise to invoke blocking as evidence of mental states.

> Conclusion 7.3: The occurrence of blocking in an organism does not provide a sufficient warrant for our ascription of cognitive mental states to it.

7.3 Could other "higher-order" forms of associative learning serve as evidence of mental states?

As part of my research for this thesis, I examined evidence for so-called "higher-order" forms of associative learning (not only blocking, but also overshadowing, sensory pre-conditioning (SPC) and second-order conditioning (SOC)) in fruit flies and other animals. However, there were no features of these forms of learning which could not be accounted for satisfactorily in terms of Dennett's goal-centred intentional stance, and I could not locate any author who was prepared to argue that overshadowing, SPC or SOC constituted evidence of mental states. These negative findings point to the following conclusion:

> Conclusion 7.4: The occurrence of higher-order forms of associative learning in an organism do not, taken by themselves, warrant the conclusion that it has cognitive mental states.

7.4 Could motor fine-tuning serve as evidence of mental states?

One might also ask whether *motor fine-tuning*, which we discussed in chapter six, could count as self-correcting behaviour of the kind required by Beisecker. Merfeld (2001) has developed a detailed model of how fine-tuning would work, on both the internal and external levels. He describes how the nervous system forms an *internal representation* of the dynamics of its sensory and motor systems:

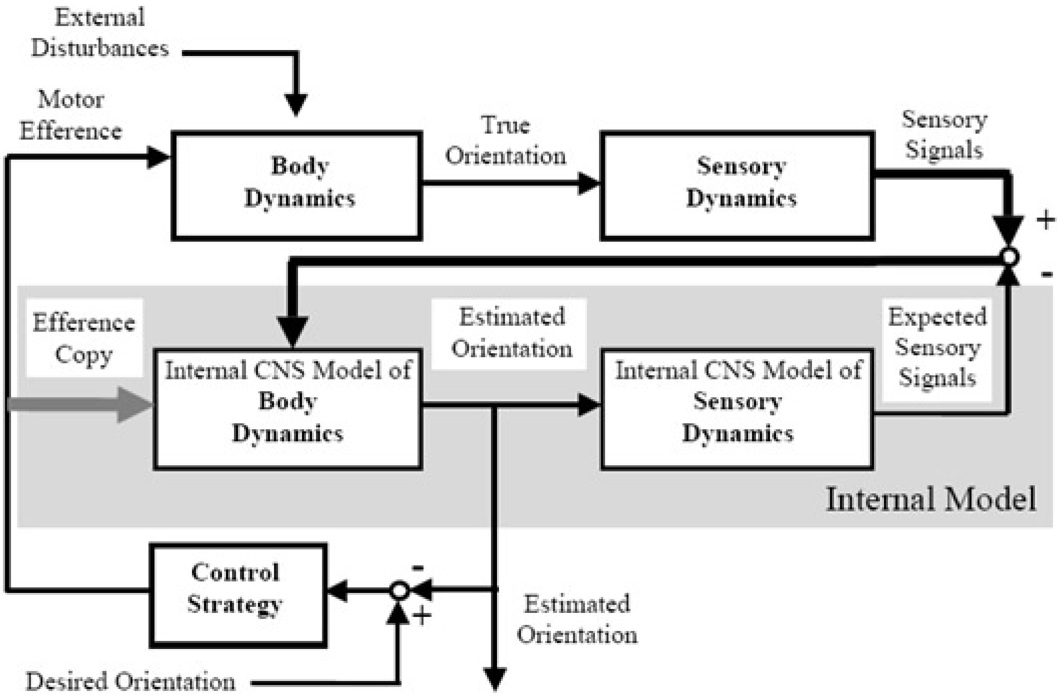


Diagram 7.1 Block diagram of Merfeld's (2001) internal representation model. The thick black arrows (*sensory signals*) represent the feedback pathway by which the sensory systems influence the estimates of the current states, (2) the thick gray arrow (*efference copy*) represents the feedforward paths which help predict what the sensors will measure based on any planned action, and (3) the highlighted boxes (*internal CNS models*) represent neural processes that help the nervous system interpret structural sensorimotor cues based on previous experience. The primary input to this model is *desired orientation*, which when compared to the *estimated orientation* yields *motor efference* via a *control strategy*. These motor commands are filtered by the *body dynamics* (e.g., muscle dynamics, limb inertia, etc.) to yield the *true orientation*, which is measured by the sensory systems with their associated *sensory dynamics* to yield *sensory signals*. In parallel with the real-world body dynamics and sensory dynamics, a second neural pathway exists that includes an *internal representation of the body dynamics* and an *internal representation of the sensory dynamics*. Copies of the efferent commands (*efference copy*) are processed by these internal representations to yield the *expected sensory signals*, which when compared to the sensory signals yield an error (mismatch). This error is fed back to the internal representation of body dynamics to help minimize the difference between the *estimated orientation* and true orientation (diagram and citation are from Merfeld, 2001, pp. 189-190).

The internal representations posited here accord very well with the notion of a *minimal map* which was proposed in chapter five. Although they contain no explicit representation of the organism's goal – the second component of a minimal map – they presuppose it, as the planned action associated with the organism's efference copy has to be directed towards some goal.

Error correction plays a vital role in Merfeld's (2001) model: the expected sensory signals generated by the animal's efferent commands are compared with the actual sensory feedback it receives, in order to ascertain the difference between the animal's *estimated orientation* and its *true orientation*. This invites the question of whether the occurrence of *fine-tuning* can be used as an observational *litmus test* for deciding which animals are capable of having cognitive mental states. I would answer in the negative, for two reasons.

First, Merfeld's description makes it clear that fine-tuning is an *internal* process as well as an external one. In Merfeld's diagram there are two internal models: an internal CNS (central nervous system) model of body dynamics and an internal CNS model of sensory dynamics. Although we might postulate the occurrence of these models within an organism's CNS on the basis of its observed motor and sensory behaviour, we also need to ascertain whether the organism's CNS actually has the right sort of complex structure to instantiate these models, at some level.

Second, the mere ability to continually adjust one's *bodily movements* while pursuing a target does not count as *error acknowledgement:* such adjustment could occur even in an organism that lacked memory of any kind. As Beisecker (2000) argues, an organism cannot be said to have acknowledged an error on its part unless it also updates its *beliefs*, so as to avoid making the same mistake in the *future.* This means adjusting one's *internal* representations – i.e. one's *minimal map.* Self-correction cannot be cashed out in purely external terms; it is an internal process as well.

> Conclusion 7.5: Motor fine-tuning cannot be invoked as a behavioural litmus test for an organism's having cognitive mental states.

> Conclusion 7.6: In addition to motor fine-tuning, an animal must possess some way of *updating* its *minimal map* before it can be said to have cognitive mental states.

The chief merit of Beisecker's account of intentionality is that it provides us with an intelligible way of attributing mental states to animals that lack language. Animals that try to correct behaviour arising from expectations that turned out to be mistaken can be said to be engaging in mental acts. However, we cannot decisively identify an organism's motor fine-tuning as self-correction of mistaken expectations unless we can show that it is also correcting its internal map of its surroundings. Only then can we be sure that the organism in question has beliefs and desires. In section C (chapter nine), I propose a detailed answer, in the context of operant agency.

In the next chapter, I examine one more ingredient of animal belief: *concepts*. I propose some appropriate criteria for deciding whether non-human animals have concepts or not.

Chapter 8 – Do Other Animals Have Concepts?

The main theme that emerges from this chapter is that we have barely even begun to ask the right kinds of questions about what kinds of concepts other animals are capable of having. I make no new proposals here. The points I wish to make are as follows: first, that the term "concept" is a multivocal one, with a variety of different usages, and there is very little that all concepts can be said to have in common; second, that concepts have an inherent normativity about them, insofar as they can be appropriately used or mis-applied; and third, that Ramsey's map metaphor seems to capture the normativity required, while at the same time being flexible enough to accommodate the fact that human and animal concepts serve a variety of different purposes, and that our own concepts are likely to be much richer and "deeper" than those of creatures with minimal minds. I then suggest that insects probably possess concepts in the minimal sense required for the map metaphor, but that there is evidence for further conceptual refinement among a few invertebrates: honeybees intuit *abstract rules* from a variety of stimuli which do not resemble each other physically, and apply them in a rule-like fashion to novel stimuli.

8.1 Methodological issues

The first major limitation which philosophers face is that there is a great deal that scientists do not yet know about animals' concept-forming abilities:

> Studies of concepts such as same-different, and perceptual concepts such as the concept of a tree or a person, have only been carried out on a very narrow range of species (pigeons, monkeys and recently horses) – far too few to be able to draw any conclusions about the distribution of these kinds of ability (Wynne, 2001, p. 189).

I should point out that since Wynne wrote, extensive research has been carried out on honeybees (Menzel and Giurfa 2001; Giurfa 2003; Giurfa, 2005); still, the list of animals whose conceptual abilities have been rigorously investigated remains disappointingly small.

The second methodological problem is that concepts differ from each other in a great many ways – so much so that Machery (2005) has recently argued cogently that *concepts are not a natural kind.* Machery questions the widely held assumption that concepts, despite their differences, share a large number of properties simply by virtue of their being mental representations. Cognitive psychologists continue to treat concepts as a homogeneous class about which certain generalisations can be formulated. Insofar as there is a debate about concepts, the debate tends to centre on which view of concepts is correct – the exemplar view, the prototype view or the theory view. Rejecting the premises of this debate, Machery argues for a *heterogeneity hypothesis:* “the class of concepts is divided into several kinds of considerations that have little in common” (2005, p. 450). Moreover,

> [M]ost categories are represented by several concepts that belong to kinds that have little in common. For example, we have several concepts of dog, each of which belongs to a different kind of concepts. There are few properties common to these concepts because in general, the corresponding kinds of concepts have little in common (Machery, 2005, p. 450).

Although we have several concepts of “dog”, each of these qualifies as concepts “because they are poised to be used in our higher cognitive processes” (Machery, 2005, p. 451). Each of these concepts is a specific body of knowledge in our mind, which resides in long-term memory and is used by default in our higher cognitive processes.

Machery concludes with some practical advice for psychologists:

> [T]he controversy between the main psychological theories of concepts is deeply misguided. Concepts are neither prototypes, nor exemplars, nor theories. Some concepts are prototypes, some concepts are sets of exemplars, some concepts are theories (2005, p. 465).

Machery's iconoclastic approach has much to commend it: indeed, one wonders why the received view that concepts are a natural kind went unquestioned for so long in psychological circles. However, this approach is insufficient by itself to resolve the question of whether other kinds of animals have concepts. It will not do to say that concepts are simply representations that are "poised to be used in our higher cognitive processes". What counts as "higher" in another species of animal, and more importantly, what counts as "cognitive"? For instance, the fact that we can describe a process using the richly cognitive terminology of "beliefs" does not mean that we should, if a more parsimonious account is available – which is why I have differentiated between two versions of Dennett's intentional stance in this thesis.

More fundamentally, Machery's account of concepts, when applied to other species of animals, overlooks their essentially *normative* character – a point well brought out by Davidson (1999):

> [B]eing able to discriminate cats is not the same thing as having the concept of a cat.

> You have the concept of a cat only if you can make sense of the idea of misapplying the concept, of believing or judging that something is a cat which is not a cat (1999, p. 8).

Something more is needed – but what? This brings us to our final methodological problem when discussing animal concepts: there is currently no philosophical consensus even on the question of what set of necessary conditions an animal's internal representation must meet in order to qualify as a *bona fide* concept. There is, however, widespread agreement that discriminatory ability is not enough.

8.2 A Proposal: Concepts as maps

G. G. Brittan Jr. builds upon Allen's (1999) theory of animal concepts (discussed in chapter one), arguing that Allen's account captures the insight that concepts serve as standards of comparison, before proposing that Ramsey's map metaphor can be extended to cover their usage:

> On Allen's account, to put it very briefly, an animal has a concept of X just in case (1) the animal systematically discriminates some instances of X from non-Xs, (2) the animal is capable of detecting some of its own discrimination errors between Xs and non-Xs, and (3) the animal is capable of better learning to better discriminate Xs from non-Xs as a consequence of this capacity. What is crucial here is that for an animal to have this capacity would seem to entail that it has some sort of "internal standard-of-comparison", a representation of the world that is independent of my present perceptual representation. Such representations, "maps by which we steer" David Armstrong calls them, bid fair to be called "concepts"; at least they do much of the same sort of work in explaining behavior… The antelope doe, for instance, is

> capable not simply of distinguishing between predators and non-predators in some general way, but also of gradually refining this distinction, learning over time which animals in its environment are rightly to be feared, and gradually correcting any initial errors she might have made. But again, such correction presupposes some sort of standard, a concept of predator, against which comparisons can be made and behavior regulated (1999, pp. 71-72).

It seems that Brittan's account, with its map metaphor, does justice to the *normativity* of concepts, as well as the fact that they have an *internal structure* of their own, whose richness may vary considerably across species. My concept of a cat (or, as Machery would say, my concepts of cat) will contain a far greater number of links to other concepts (animal, mammal, carnivore; pet; killer of native birds; team mascot; and so on) than a dog's concept of a cat. Nevertheless, the map metaphor has to do some explanatory work. All concepts, no matter how simple, have an internal structure or schema of some sort. How does one verify whether the animal is using that schema when applying the concept? To answer this question we need to ask what kinds of *misapplications* the animal would make in the absence of the schema.

The kinds of mistakes animals make under laboratory conditions and in the wild can tell us what kinds of concepts they do and do not have. Budiansky (1998, pp. 126-127) provides examples of how capuchin monkeys, despite their extensive use of tools in the wild, make the most elementary mistakes when presented with spatial problem-solving puzzles in laboratory tasks. Although they can use tools, they have no idea how they work. Does this mean that they lack the concept of a tool?

I would argue otherwise. In the wild, capuchin monkeys “have been reported to use sticks to kill a snake, as weapons against other monkeys, and as probes for food, much like chimpanzees” (Budiansky, 1998, p. 127). Although they do not know how the tools achieve their purpose, they obviously know how to hold them when using them to get what they want. They can pick up the stick and hold it in the most appropriate way to accomplish the task they are engaging in. In other words, the concept they possess of a stick is a very primitive one: a purely *instrumental* concept - which can be represented as a kinesthetic space of “right moves” for the *correct use* of the tool in order to obtain a desired object - rather than a *mereological* concept of the tool’s internal structure, which would enable its possessor to devise new uses for it.

Can we speak of a minimal map here? I believe so. The minimal map simply has to include an internal representation of the animal’s current state (its bodily orientation, especially the way it is using its hands right now), as well as a sequence (or pathway) of finely tuned “right moves” which the animal needs to execute manually, in order to obtain its goal.

Certainly, Ramsey’s “map” metaphor can be over-used, but its usage here would seem perfectly appropriate if experimental evidence also shows that animals are capable of modifying different components of their instrumental concepts – e.g. learning to use a stick in a new way, which requires them to update their minimal maps.

8.3 Which animals have concepts?

8.3.1 Discrimination and stimulus categorisation

In chapter two, I cited evidence that insects, fish and crustaceans, were able to recognise and discriminate successfully between objects such as food items, nest-mates, prey, predators, potential mates, and so on (Franklin *et al.*, 2005, online). Individual recognition has also been verified in wasps (Tibbetts, 2002). On the other hand, the evidence for discriminatory abilities among worms and simple molluscs is largely limited to odors. It would therefore seem unwise to attribute concepts to these creatures. However, if insects and crustaceans prove to be capable of *refining* their object recognition as they mature and/or undergo training, then we could properly attribute what I call *instrumental concepts* to these animals, in accordance with G. G. Brittan's (1999) proposal discussed above. These concepts would be truly *normative*, insofar as they could be mis-applied.

Stimulus categorisation is a conceptual capacity that can be applied not only to familiar stimuli but also to *novel* stimuli. To date, stimulus categorisation has only been verified in mammals, birds and recently honeybees. Giurfa (2005) defines stimulus categorisation as follows:

> Categorization refers to the classification of perceptual input into defined functional groups... It can be defined as the ability to group distinguishable objects or events on the basis of a common attribute or set of attributes, and therefore to respond similarly to them ... Categorization deals, therefore, with the extraction of these defining attributes from objects of the animal's environment. Our use of the term categorization will be restricted to those cases in which *animals transfer their choice to novel stimuli that they have never met before* on the basis of common features shared with known stimuli (Giurfa, 2005, online, italics mine).

It is not known how widespread this capacity is among the animal kingdom. However, Giurfa (2005) argues that that this capacity might prove to be explicable in a straightforward sense, within the terms of an *associative* learning paradigm, which is very common in the animal world:

> In my opinion, categorization does *not* reflect retrospective analysis of events but results from simple *associative learning.* To explain this view, the neural mechanisms underlying categorization could be considered, in particular with respect to the organization of the bee brain… If we admit that visual stimuli are categorized on the basis of specific features such as *orientation* or *symmetry*, the neural implementation of category recognition could be relatively simple. The feature(s) allowing stimulus classification would activate specific neuronal detectors in the optic lobes, the visual areas of the bee brain. Examples of such feature detectors are the orientation detectors whose tuning and orientation have been already characterized by means of electrophysiological recordings in the honeybee optic lobes … In the case of category acquisition, the activation of an additional neural element is needed. Such [an] element would be necessary and sufficient to represent the *reward* (sucrose solution) and should contact and modulate the activity of the visual feature detectors in order to assign *value* to appropriate firing. This kind of neuron has been found in the honeybee brain as related to the olfactory circuit (Giurfa, 2005, online, italics mine).

If Giurfa is right here, then given that many other insects have sophisticated visual systems, it seems reasonable to expect that they should also be capable of performing stimulus categorisation tasks. If the ability to transfer to *novel* stimuli were verified in these insects, it would suggest that the category is indeed serving as a kind of *norm* or standard for the insects. Because we can speak of *conditions of satisfaction* applying in such cases, I would argue that we can attribute rudimentary concepts to any animal capable of this kind of stimulus categorization.

8.3.2 Learning of abstract rules in honeybees

Honeybees appear able to go a step further than other insects. There is experimental evidence that they can acquire *abstract concepts* such as “same” and “different”, possession of which presupposes the grasp of a *rule.* This is a remarkable claim, so I shall quote Giurfa directly:

> An example of rule learning is the learning of the so-called principles of sameness and of difference. These rules are usually uncovered through the delayed matching to sample (DMS) and the delayed non-matching to sample (DNMS) experiments, respectively. In DMS, animals are presented with a sample and then with a set of stimuli, one of which is identical to the sample and which is reinforced. As the sample is being changed regularly, they have to learn the sameness rule ‘*choose always what is shown to you (the sample), independently of what is shown to you*’. In DNMS, the animal has to learn the opposite, i.e. ‘*choose always the opposite to what is shown to you (the sample)*’. The interesting point concerning these protocols is that predictive analysis based on stimulus or feature generalization does not necessarily hold as the rule is ideally independent of the physical nature of the stimuli used. To discover the rule, the animal has to operate on the set of examples known such that retrospection and different forms of heuristics can be applied to solve the problem. Neural accounts based on simple associative networks such as that proposed for visual categorization (see above) may not be valid in this case. Although reinforcement can still be represented by a specific neural pathway or element … the novel, differing sample (e.g.; a color) will not activate the same network components responding to a previous sample (e.g., an odor). Extracting the rule in a changing learning set means therefore

> going beyond stimulus modality and performing a form of retrospective or diagnostic analysis of the problem faced (Giurfa, 2005, online).

Honeybees foraging in a Y-maze learned to solve both DMS and DNMS rules (Giurfa et al., 2001). Given that the rule involved here is *independent of the physical nature of the stimulus*, it is hard to see on what grounds one could refuse to grant that mastery of the rule requires possession of an *abstract concept*, which takes us to a level of refinement not seen in other insects to date.

The word “abstract” is important here: the fact that only honeybees, among insects, appear to be capable of following an *abstract* rule, should not blind us to the fact that *other insects* which can categorize novel stimuli are also following rules. For these insects, however, the rules pertain to *perceptual categories.* What I am proposing here is that insects that are capable of following *perceptual* rules can also be described as possessors of simple concepts, even if they cannot grasp these rules, as honeybees appear to be able to do.

The emerging picture appears to be that many insects and crustaceans possess primitive *instrumental concepts* that are acquired through a process of recognizing *perceptual categories,* which they learn to associate with some *reward* or *punishment.* These concepts have a rudimentary internal structure, which is simply a representation of how they may be used to obtain their goal. These concepts are, I would suggest, the most primitive kinds of concepts: their significance is exhausted by the fact that they serve as indicators of something else the animal wants, which can only be obtained by fine-tuning its behaviour. Honeybees, on the other hand, appear to operate on a more *abstract* level of cognition.

> Conclusion 8.1: An animal has to possess *instrumental concepts*, which not only allow it to associate *means* with *ends*, but also contain a representation of the *manner* in which the means can be used to obtain the ends, before we can justifiably ascribe mental states to it.

I have not discussed here whether invertebrates possess the *concept of a physical object.* Proust (1999) makes a good case that *calibration* of different *perceptual systems* (e.g. in barn owl chicks calibrating their hearing to their vision – see Brainard and Knudsen, 1998) is required before a creature can form the concept of a physical object. Although a variety of animals are capable of multimodal integration of their sensory input (New and Kang, 2000; New, 2002; Chittka, 2004) it seems that the calibration of different perceptual systems required by Proust has only been identified in mammals and birds to date (see Proust, 1999; Bryson and Hauser, 2002). Interestingly, these creatures are also the only animals capable of solving displacement tasks and acquiring the concept of Piagetian object constancy, as we saw in chapter one. Further research is needed to show whether there is any connection between the neurological requirements for possessing the concept of a physical object and the neurological requirements for phenomenal consciousness, which, as we saw in chapter one, is probably confined to mammals and birds.

8.4 Reversal learning

The ability to show rapid improvements in *reversal learning tests* has been proposed as a higher-level cognitive capacity, the possession of which might enable us to distinguish creatures with minds from those lacking them (Varner, 1998). In chapter nine, I shall criticise the view that reversal learning is a *sine qua non* for having mental states. Here, I simply examine the question of whether the capacity for reversal learning is a reliable indicator of mental states, and if so, which creatures possess this capacity.

8.4.1 Is rapid reversal learning evidence for the acquisition of beliefs and concepts?

Although “learning” in the broadest sense of the word (i.e. the acquisition of new skills) may not necessarily require mental states, no-one would deny that *meta-learning* ("learning to learn") is a critical capacity, which presupposes an ability to evaluate and correct one's actions. A creature with this ability would qualify as what Dennett (1997, p. 112ff.) calls a "Popperian creature". Such a creature is one level above a Skinnerian creature, which is capable of learning from its trial-and-error mistakes and successes, and can associate information about one kind of event with information about another kind. A Skinnerian creature may stumble upon a "smart move", but it cannot *predict* what works and what does not. Its first move may be a fatal one, if it is unlucky.

A "Popperian creature" can avoid such an outcome, because it can foresee the consequences of its actions in the "inner environment" of its imagination, which lets the creature manipulate information in its memory, about its external environment. In this inner environment, try-outs or simulations can be executed without harming the animal, allowing it to select the best course of action and make a smart first move in its external ("real") environment. The advantage of foresight is that it "permits our hypotheses to die in our stead", as Popper put it (Dennett, 1997, p. 116). The creature can make a smart first move, because it can think about smart moves.

Varner (1998), following Bitterman (1965), has suggested an experimental way to identify meta-learning in animals and resolve the question of whether they have mental states. He has argued that if they are genuinely learning, (and not merely mechanically associating), they should be forming *hypotheses* about the changes in their environment. He has proposed that *reversal tests* offer a good way to test animals' abilities. *Multiple reversal tests* involve repeatedly reversing the reward pattern in simple learning experiments. For instance, a rat is first presented with two levers and rewarded for pressing the left lever instead of the right. When the rat has learned to press the left lever all the time, the reward pattern is reversed. Once the rat has learned the new reward pattern, it is reversed again, and so on. Varner suggests that if an animal shows no improvements in the time it takes to adjust to subsequent reversals, that suggests an inflexible, non-cognitive mechanism is governing its behaviour. By contrast, Bitterman predicted that an animal that can form hypotheses should take *longer* to learn the new pattern the first time it is reversed, but should adjust more and more rapidly to subsequent reversals, as it learns to quickly revise its expectations.

Varner's proposal invites two questions. First, is rapid reversal learning a sign of intelligence? Second, does progressive improvement in multiple reversal tests indicate the presence of mental states?

The ability to adapt rapidly to changes sounds like a mind-like feature. However, the consensus from animal behaviourists is that it need not be so. According to Ben-Shahar (personal email communication, 19 August 2003), the rapid reversal learning of honey bees surpasses even that of pigeons and rats. However, Ben-Shahar cautions against the use of reversal learning *per se* as a measure of intelligence in animals, as the rapid reversal learning appears to be an adaptive trait for some animals, and adaptive behaviour is not necessarily intelligent:

> I'm not convinced that reversal learning is necessarily directly related to intelligence. It is possible that for some species, reversal is highly adaptive, and hence the good performance. In bees one could speculate that reversal is very important to an animal that forages on unstable resources. In bees and other social species this is even more critical since they use communal foraging strategies. Bees will follow other bees to resources previously identified. If these have dried out the new forager has to look for new resources fast or she will come back empty - a big waste of time (personal email communication, 19 August 2003).

We can formulate the following negative conclusion:

> Conclusion 8.2: The capacity for rapid reversal learning in an animal does not, by itself, warrant the ascription of mental states to it.

8.4.2 Is progressive adjustment in multiple reversal learning trials evidence for the acquisition of beliefs?

The second and more interesting question is whether the existence of *progressive adjustment* in multiple reversal learning trials indicates intelligence.

The ability to improve in multiple reversal learning trials is readily explained by the hypothesis that the animal is forming a *hypothesis* about changes in its environment. I have not been able to find a non-cognitive explanation as to why such improvement might occur. Certainly, the fact that the cognitive explanation makes a highly specific prediction (that the animal should take longer to learn the new pattern the first time it is reversed), which has been experimentally confirmed, tends to bear out a mentalistic interpretation. It should be borne in mind, however, that even if the behaviour cannot be accounted for in terms of associative learning, that does not necessarily make it cognitive.

Even if progressive adjustment shows that an animal has mental states, that does not necessarily make it what Dennett (1997, pp. 112 ff.) calls a *Popperian creature.* The ability to formulate primitive hypotheses need not imply the ability to foresee the consequences of one's actions in the "inner environment" of one's imagination.

> Conclusion 8.3: Progressive adjustments in serial reversal tests constitute good *prima facie* evidence that an animal is trying to adjust to sudden changes in its environment, by rapidly revising its expectations.

Basing his arguments on research by Morton Bitterman (1965), Varner has claimed (1998, p. 32) that progressive adjustment in multiple reversal learning trials is found only in reptiles, birds and mammals. Since then, it has become apparent that fish (Wakelin, 2003) and even honeybees (Komischke, Giurfa, Lachnit and Malun, 2002), are also capable of this kind of learning. Komischke, Giurfa, Lachnit and Malun (2002) compared the responses of bees that had experienced reversals with those of bees that had not experienced such reversals when both were confronted with a new reversal situation. They found that bees that had experienced three previous reversals were better in solving the final reversal task than bees with no previous reversal experience. They also showed that one reversal learning trial was enough for bees to perform better in the final reversal task.

The evidence to date from serial reversal learning suggests that honeybees, at least, are capable of *learning to learn.* This ability may turn out to be widespread among insects, but very little research has been done with most groups of insects. Brembs claims that serial reversal learning in insects is not confined to honeybees:

> *Drosophila* can reversal learn and if the pattern-heat contingency is reversed, learning is faster (personal email communication, 11 August 2003).

However, neither Brembs nor *Drosophila* researcher Josh Dubnau was able to supply a reference to serial reversal learning by *Drosophila melanogaster* in the published literature.

The evidence from *serial reversal learning* is thus of limited value. At most, it suggests that honeybees are capable of meta-learning, while saying nothing about other insects.

Section C: The Conditions for Intentional Agency – a Synthesis

Chapter 9 - Four Models of a Minimal Mind

9.1 Synthesis: The Sufficient Conditions for Intentional Agency

In section B (chapters two to eight), I examined various aspects of animal behaviour in my quest to identify the conditions which a creature would have to satisfy in order to warrant being called an "intentional agent". Although I identified many necessary conditions for intentional agency, I made no attempt to integrate them, to see if some combination of these criteria might constitute a set of sufficient conditions for intentional agency. This is what I shall now endeavour to do.

On the account I am developing here, an animal's satisfaction of the following set of conditions suffices to warrant our referring to it as an agent, with *bona fide* beliefs, desires and intentions of its own – even if these turn out to be entirely devoid of phenomenal consciousness. The conditions relate to the categories of neuroanatomy; innate behaviour; sensory capacities; memory; associative learning; representation; action selection and fine tuning; self-correction; and primitive concepts.

Neuroanatomy

Possession of a *central nervous system and a primitive brain* was a critical requirement for intentional agency that emerged during the course of our investigation (see for instance Conclusion 6.12). Without these features, a creature would be incapable of meeting most of the other requirements for intentional agency: possession of true senses; procedural memory; a capacity for associative learning; a "minimal map" representation in the creature's brain and central nervous system; the feedback and feed-forward capacities presupposed in fine-tuning; the ability to revise incorrect beliefs; and the ability to form primitive, instrumental concepts.

Innate behaviour

An animal could never manifest intentional agency unless it had *innate preferences* that served to render intelligible the creature's pursuit of its current goal. As I use the term, “innate preferences” denotes a general category (e.g. the various kinds of foods that an animal is attracted to); whereas an animal’s current goal is something *specific* – e.g. *this* food, available here and now. Additionally, the animal needs to possess a suite of *innate motor programs,* stored in its brain, from which it can make a selection. *Fine-tuning* (discussed in chapter 6) presupposes the existence of these programs, as it involves refining one of them. Finally, without an innate tendency on the animal's part to engage in *exploratory behaviour,* the animal would never be able to discover pathways that would allow it to reach its current goal, or stabilise its motor pattern by a trial-and-error process.

Sensory capacities

An animal exhibiting intentional agency requires ongoing *sensory input*, to inform it about whether it has attained its goal, and if not, whether it is getting closer to achieving it. Since the animal needs to respond rapidly if is moving away from its goal (or towards danger), only "true" senses (defined in section 2.3.4) will suffice here (Conclusion 2.3).

Memory

Any creature capable of agency needs to possess a *procedural memory*, so that it can acquire new skills (Conclusion 3.6). Without this capacity, the ascription of mental states to the animal would be superfluous.

Associative learning

We have already seen that agency is impossible without the capacity to *associate actions with their consequences* (Conclusion 5.10). As we saw in chapter five, this kind of learning is critically dependent on the timing of the events being associated. A *correlation mechanism,* allowing an animal to find a *temporal coincidence* between its behaviour and the attainment of its goal, is therefore vital (Conclusion 5.14).

Representation

As existing definitions of representation in the contemporary literature were found to be either inadequate or question-begging, the concept of a *minimal map* was proposed as a simple representation of an individual's *current state*, its *goal* and its *path* for "steering itself" towards its goal. The notion of belief I invoked here was Ramsey's account of belief as a "map … whereby we steer" (1990, p. 146). Not all beliefs are map-like, but I argued that the map metaphor is indispensable for understanding intentional agency (Conclusion 5.13). The "steering" need not be spatial. In this chapter, I shall argue that behaviour belonging to four different categories (fine motor behaviour; navigation; tool use; and social behaviour) can satisfy the "steering" metaphor in a robust sense of the word. If this is correct, then we can say that there are four different kinds of minimal minds.

Action selection and fine tuning

In order to manifest intentional agency, an animal not only needs to possess an *action selection mechanism,* which allows it to select the best motor pattern in its built-in repertoire for obtaining its goal (see Conclusion 6.8), but it also needs to *control* that pattern by *stabilising (fine-tuning) it within a narrow range of values,* with the help of mechanisms that provide it with rapid feedback and the ability to predict the sensory consequences of its actions (Conclusion 6.11).

Self-correction of behaviour and beliefs

Intentionality is marked by its own unique brand of normativity, which can be described in terms of *conditions of satisfaction.* Of course, all biological functions possess some degree of normativity, but this proved to be the wrong sort for characterising the intentionality of mental acts. It was argued that educable creatures who demonstrated a capacity for self-correction were good *prima facie* candidates for mental acts, but that left us with the problem of nominating criteria for identifying self-correction. Merfeld's (2001) model of efferent copy suggested a solution to this problem, on the *behavioural* level: in the event of a mismatch between expected and actual sensory signals, the mismatch is used as an *error signal* to guide the estimated state back toward the actual state. But as we saw in chapter seven (Conclusion 7.5), this capacity for motor fine-tuning is not enough to warrant the ascription of error acknowledgement to an animal. An animal cannot be meaningfully said to have acknowledged an error on its part unless it also updates its *beliefs,* so as to avoid making the same mistake in future. If we suppose that these beliefs are recorded on its internal minimal map, then the animal additionally requires a way of *updating* its minimal map when circumstances demand (Conclusion 7.6).

Primitive Concepts

In chapter eight, we proposed that animals that are capable of operant agency could be said to possess crude *instrumental concepts,* which represent the *manner* in which they should make use of certain *means,* in order to obtain their *ends*, and that possession of these instrumental concepts was a necessary condition for the warranted attribution of mental states to an animal (Conclusion 8.1). It was argued that a variety of invertebrates (especially insects and cephalopods) were capable of representing structures according to this *means-end schema.*

At a more refined level, we can describe having a concept in terms of *abstracting (and following) a rule* (Giurfa, 2005), where the rule is *independent* of the physical nature of the stimulus. Honeybees, birds and mammals are the only creatures to which we can ascribe concepts of this sort, at present.

9.2 A model of Operant Agency

9.2.1 Definition of Operant Agency

I would now like to propose a set of sufficient conditions for *operant conditioning* and what I call *operant agency:*

Definition - "operant conditioning"

An animal can be described as undergoing operant conditioning if the following features can be identified:

Innate behaviour:

(i) *innate preferences* or drives;

(ii) *innate motor programs*, which are stored in the brain, and generate the suite of the animal's motor output;

(iii) a tendency on the animal's part to engage in *exploratory behaviour;*

Sensory and discriminatory capacities:

sensory inputs (from “true” senses, as defined in section 2.3.4 of chapter two) which inform the animal whether it has attained its goal, and if not, whether it is getting closer to achieving it;

Memory:

an ability to remember motor movements that enable an animal to obtain what it seeks, and store these movements in its procedural memory;

Learning-related requirements: associations and a correlation mechanism:

(i) direct or indirect *associations* between (a) different motor commands; (b) sensory inputs (if applicable); and (c) consequences of motor commands, which are stored in the animal's memory and updated when circumstances change;

(ii) a *correlation mechanism*, allowing it to find a *temporal* coincidence between its motor behaviour and the attainment of its goal;

Representation:

an internal representation *(minimal map)* which includes the following features:

(i) the animal's *current motor output* (represented as its *efference copy*);

(ii) the animal's *current goal* (represented as a stored memory of the motor pattern or sensory stimulus that the animal associates with the attainment of the goal, which is a "reward" or the avoidance of a "punishment"); and

(iii) the animal's *pathway* to its current goal (represented as a stored memory of the sequence of motor movements or sensory stimuli which enable the animal to steer itself towards its goal);

Action selection and fine tuning:

(i) an *action selection mechanism*, which allows the animal to make a selection from its suite of possible motor response patterns and pick the one that is the most appropriate to its current circumstances;

(ii) *fine-tuning behaviour:* efferent motor commands which are capable of stabilising a motor pattern at a particular value or within a narrow range of values, in order to achieve a goal;

(iii) the ability to store and compare internal representations of its current motor output (i.e. its *efferent copy*, which represents its current "position" on its internal map) and its *afferent* sensory inputs;

Self-correction of behaviour and beliefs:

(i) an ability to *rectify any deviations* in motor output from the range which is appropriate for attaining the goal;

(ii) *abandonment* of behaviour that increases, and *continuation* of behaviour that reduces, the animal's "distance" (or deviation) from its current goal; and

(iii) an ability to *form new associations* and alter its internal representations (i.e. *update its minimal map)* in line with variations in surrounding circumstances that are relevant to the animal's attainment of its goal.

Primitive Concepts:

a primitive *instrumental concept* of a bodily movement, which represents the *manner* in which it can be used to obtain the animal’s current goal.

If the above conditions are all met, then we can legitimately speak of the animal as an *intentional agent* which *believes* that it will get what it *wants*, by doing what its internal map tells it to do.

Definition - "Operant Agency"

Operant conditioning is a form of learning which presupposes an *agent-centred* intentional stance.

Animals that are capable of undergoing operant conditioning can thus be said to exhibit a form of agency called *operant agency.*

9.2.2 Case study: Operant Agency in fruitflies

In this section, I argue that fruit flies satisfy the set of sufficient conditions defined above for the occurrence of intentional agency, in the context of operant conditioning. My proposals are based on Prescott (2007), Abramson (1994, 2003), Dretske (1999), Wolf and Heisenberg (1991), Heisenberg, Wolf and Brembs (2001), Brembs (1996, 2003), Cotterill (1997, 2001, 2002), Grau (2002), Beisecker (1999) and Carruthers (2004a). Of particular relevance are the experiments described by Heisenberg (1991), Heisenberg, Wolf and Brembs (2001), and Brembs (1996, 2003), with the fruit fly *Drosophila melanogaster* at the flight simulator.

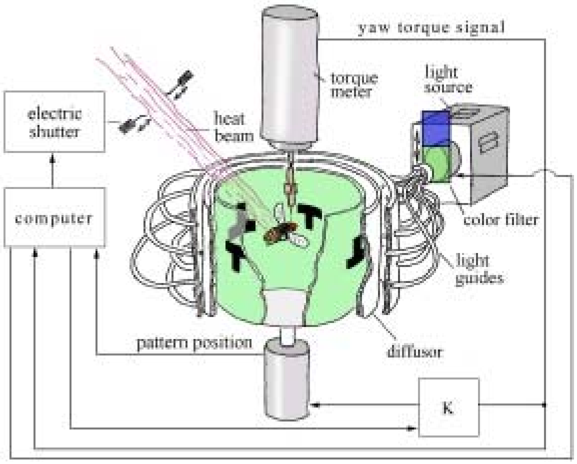


__Figure 9.1__ **Flight simulator set-up. Taken from "An Analysis of Associative Learning in Drosophila at the Flight Simulator", Ph.D. thesis by Bjorn Brembs. In the experiments, a fruit fly is tethered to a computer. The fly is flying stationarily in a cylindrical arena homogeneously illuminated from behind. The fly has only one degree of freedom in its movements: its yaw torque, or tendency to perform left or right turns. The fly's tendency to perform left or right turns (yaw torque) is measured continuously and fed into the computer. The fly is then subjected to simple operant conditioning, classical conditioning, or a combination of the above (either flight-simulator mode or switch-mode). The computer controls pattern position (via the motor control unit K), shutter closure and color of illumination according to the conditioning rules.**

__Innate preferences__

The experimental set-up depicted above for monitoring the operant behaviour of the fruit-fly *(Drosophila melanogaster)* is constructed on the assumption that fruit-flies have an **innate aversion to heat**, and will therefore try to avoid an infra-red heat-beam. The flies in the experiment face a formidable challenge: *they have to "figure out" what to do in order to shut off a heat beam which can fry them in 30 seconds.* The flies therefore satisfy a crucial condition for our being able to ascribe mental states to animals: they display selfish behaviour, which is directed at satisfying their own built-in biological needs.

Innate Motor Programs and Exploratory Behaviour

In the experiment, the tethered fruit fly is placed in a cylindrical arena which is capable of rotating in such a way as to simulate flight, even though the fly is stationary. The fly has **four** basic **motor patterns** that it can activate - in other words, four degrees of freedom. It can adjust its yaw torque (tendency to perform left or right turns), lift/thrust, abdominal position or leg posture (Heisenberg, Wolf and Brembs, 2001, p. 2).

The fly selects an appropriate motor pattern by a trial-and-error process of *exploratory behaviour.* Eventually, it manages to stabilise the rotating arena and prevent itself from being fried by the heat beam:

> As the fly initially has no clue as to which behavior the experimenter chooses for control of the arena movements, the animal has no choice but to activate its repertoire of motor outputs and to compare this sequence of activations to the dynamics of arena rotation until it finds a correlation (Heisenberg, Wolf and Brembs, 2001, p. 2).

Action selection

Prescott (2007, p. 1) defines **action selection** as the problem of "resolving conflicts between competing behavioural alternatives". The behavioural alternative (or motor pattern) selected by the fly is the one that enables it to avoid the heat. The fly engages in action selection when undergoing operant conditioning, and also when it is in "flight-simulator mode" and "switch-mode" (see below for a definition of these terms).

Fine-tuning

A tethered fly has four basic motor programs that it can activate. Each motor program can be implemented at different strengths or *values.* A fly's current yaw torque is always at a particular angle; its thrust is always at a certain intensity, and so on. In other words, each of the fly's four motor patterns can be *fine-tuned.*

In the fruit-fly experiments described above, flies were subjected to four kinds of conditioning, the simplest of which is referred to by Brembs (2000) as *pure operant conditioning* and by Heisenberg, Wolf and Brembs (2001) as *yaw-torque learning.* However, if we follow the naming convention proposed by Abramson (1994, p. 151) in chapter five, the flies' behaviour might be better described as *instrumental conditioning*, as the essential ingredient of *fine-tuning appears to be absent.* As Heisenberg (personal email, 6 October 2003) points out, all that the flies had to learn in this case was: "Don't turn right."

The range of permitted behaviour in yaw-torque learning (i.e. flying anywhere in the left domain) is too broad for us to describe this as fine-tuning. Only if we could show that *Drosophila* was able to fine-tune one of its motor patterns (e.g. its thrust) while undergoing yaw torque learning could we then justifiably conclude that it was a case of true operant conditioning.

In *flight-simulator* mode (or fs-mode), the flies faced a more interesting challenge: they had to stabilise a rotating arena by modulating their yaw torque (tendency to turn left or right), and they also had to stay within a safe zone to avoid the heat.

In *switch mode* (or sw-mode), the fly received two forms of feedback: heat (the unconditioned stimulus) and a change in colour (from blue to green) or pattern (from an upright to an inverted T) of the panorama (the conditioned stimulus). For instance, if the fly's yaw torque range fell into the left domain, heat was switched on and the upright T pattern was placed in front; flying in the right domain switched the heat off and the arena was quickly rotated by 90 degrees, shifting the inverted T to the front. Thus the fly underwent parallel operant conditioning (a combination of operant and classical conditioning). Unlike flight simulator mode, the fly had to respond in a very artificial way to avoid heat in switch mode: it had to keep flying around in counterclockwise circles in order to avoid being incinerated (Brembs, personal email, 11 August 2003).

In other experiments (Brembs, 2003), flies were able to adjust their thrust to an arbitrary level that stopped their arena from rotating. I would argue that the ability of the flies to narrow their yaw torque range or their thrust to a specified range, in order to avoid heat, fulfils the requirements for fine-tuning as defined in chapter six: namely, stabilising a basic motor pattern at a particular value or confining it *within a narrow range of values*, in order to achieve a goal that the individual had *learned* to associate with the action. **We can conclude that *Drosophila* is capable of true operant behaviour.**

Recent research by Frye and Gray (2005) lends strong support to this view. The authors discuss the neural mechanisms of aerodynamic control in insects, and detail the ways in which sensory feedback enables insects to control their flight trajectory. It appears that mechanosensory organs play a critical role here. Scientists are still investigating the precise manner in which the nervous system co-ordinates subtle changes in wing kinematics and aerodynamics for stability and control of hovering, forward flight and steering.

Other requirements for conditioning: a current goal, sensory inputs and associations

As well as having innate goals, the fly also has a **current goal:** to avoid being incinerated by the heat beam.

Sensory input can also play a key role in operant conditioning: it informs the animal whether it has attained its goal, and if not, whether it is getting closer to achieving it. A fly undergoing operant conditioning in sw-mode or fs-mode needs to continually *monitor* its sensory input (the background pattern on the cylindrical arena), so as to minimise its deviation from its goal (Wolf and Heisenberg, 1991; Brembs, 1996, p. 3).

By contrast, a fly undergoing yaw torque learning has no sensory inputs that tell it if it is getting closer to its goal: it is flying blind, as it were. The only sensory input it has is a "punishment" (the heat beam is turned on) if it turns right.

Finally, an animal undergoing conditioning needs to be able to form **associations.** In this case, the fly needs to be able to either associate motor commands *directly* with their consequences (yaw torque learning) or associate them *indirectly*, by forming direct associations between motor commands and sensory inputs (changing patterns on the fly's background arena), and between these sensory inputs and the consequences of motor commands.

Correlation mechanism

The animal clearly also possesses a **correlation mechanism**; otherwise it could not find a temporal coincidence between its motor behaviour and the attainment of its goal (avoiding the heat). Once it finds a temporal correlation between its behaviour and its proximity to the goal, "the respective motor program is used to modify the sensory input in the direction toward the goal" (Wolf and Heisenberg, 1991, p. 699; Brembs, 1996, p. 3). For instance, in the case of the fly undergoing flight-simulator mode training, finding a correlation would allow it to change its field of vision, so that it could keep the inverted T in its line of sight.

Internal representations and minimal maps

First, *Drosophila* can form *internal representations* of its own bodily movements, for each of its four degrees of freedom, within its brain and nervous system.

Second, it can either (a) *directly* associate these bodily movements with good or bad consequences, or (b) associate its bodily movements with sensory stimuli (used for steering), which are in turn associated with good or bad consequences (making the association of movements with consequences *indirect*). In case (a), *Drosophila* uses an *internal motor map*; in case (b), it uses an *internal sensorimotor map*. In neither case need we suppose that it has a spatial grid map.

A minimal map, or action schema, is what allows the fly to *fine-tune* the motor program it has selected. In other words, the existence of a *minimal map* (i.e. a map-like representation of an animal's current state, goal and pathway to its goal) is what differentiates *operant* from merely *instrumental* conditioning.

For an *internal motor map*, the *current state* is simply the present value of the motor plan the fly has selected (e.g. the fly's present yaw torque), the *goal* is the value of the motor plan that enables it to escape the heat (e.g. the safe range of yaw torque values), while the *means* for getting there is the appropriate movement for bringing the current state closer to the goal.

For an *internal sensorimotor map*, the *current state* is the present value of its motor plan, coupled with the present value of the sensory stimulus (color or pattern) that the fly is using to navigate; the *goal* is the color or pattern that is associated with "no-heat" (e.g. an inverted T); and the *means* for getting there is the manner in which it has to fly to keep the "no-heat" color or pattern in front of it.

I propose that *Drosophila* employs an internal *sensorimotor map* when it is undergoing flight-simulator mode (fs-mode) learning. I suggest that *Drosophila might* use an internal *motor map* when it is undergoing pure operant conditioning (yaw torque learning). (I am more tentative about the second proposal, because as we have seen, in the case of yaw torque learning, *Drosophila* may not be engaging in fine-tuning at all, and hence may not need a map to steer by.) *Drosophila* may make use of *both* kinds of maps while flying in switch mode (sw-mode), as it undergoes parallel operant conditioning.

An internal motor map, *if* it existed, would be the simplest kind of minimal map, but if (as I suggest) what Brembs (2000) calls "pure operant conditioning" (yaw torque learning) turns out to be merely instrumental learning, then we can explain it without positing a map at all: the fly may be simply forming an *association* between a kind of movement (turning right) and heat (Heisenberg, personal email, 6 October 2003).

In the fine-tuning process I describe, there is a continual inter-play between *Drosophila*'s "feed-back" and "feed-forward" mechanisms. *Drosophila* at the torque meter can adjust its yaw torque, lift/thrust, abdominal position or leg posture. I propose that the fly has an internal motor map or sensorimotor map corresponding to *each* of its four degrees of freedom, and that once it has selected a motor program, it can use the relevant map to steer itself away from the heat.

Actions and their consequences: how associations can represent goals and pathways on the motor map

The internal representation of the fly's motor commands has to be coupled with the ability to form and remember associations between different possible *actions* (yaw torque movements) and their *consequences*.

On the hypothesis which I am defending here, the internal *motor map* (used in sw-mode and possibly in yaw torque learning) directly associates different yaw torque values with heat and comfort. The fly's *goal* (to escape from the heat) could be represented on this map as a stored motor memory of the *motor pattern* (flying clockwise) which allows the fly to stay out of the heat, and the *pathway* as a stored motor memory (based on the fly's previous exploratory behaviour) of the *movement* (flying into the left domain) which allows the fly to get out of the heat.

The internal *sensorimotor map*, which the fly uses in fs-mode and sw-mode, indirectly associates different yaw torque values with good and bad consequences. For instance, different yaw torque values may be associated with the upright T-pattern on the rotating arena (the conditioned stimulus) or with the inverted T-pattern. These patterns are associated with heat (the unconditioned stimulus) and “comfort” respectively. On this map, the fly's *goal* could be encoded as a stored memory of a *sensory stimulus* (e.g. the inverted T) that the fly associates with “comfort” (i.e. the absence of heat), while the *pathway* would be the stored memory of a *sequence of sensory stimuli* which allows the animal to steer itself towards its goal.

The underlying assumption here, that the fly can form associations between things as diverse as motor patterns, shapes and heat, is supported by the proposal of Heisenberg, Wolf and Brembs (2001, p. 6) that *Drosophila* possesses a *multimodal memory*, in which "colors and patterns are stored and combined with the good and bad of temperature values, noxious odors, or exafferent motion".

Self-correction

One of the conditions that we identified for **self-correcting behaviour** in chapter 7 was that the animal had to be able to rectify motor patterns which *deviate* outside the desired range. Legrand (2001) has proposed that the "efference copy" of an animal's motor program not only gives it a sense of *trying* to do something, but also indicates the necessity of a *correction*.

However, as Beisecker (1999) points out, self-correction involves modifying one's *beliefs* as well as one's actions, so that one can avoid making the same mistake in future. This means that animals with a capacity for self-correction have to be capable of updating their *internal representations.* One way the animal could do this is to continually update its multimodal associative memory as new information comes to light and as circumstances change. For example, in the fly's case, it needs to update its memory if the inverted T design on its background arena comes to be associated with heat rather than the absence of it.

Instrumental concepts

In chapter eight, we defined the notion of an **instrumental concept**, which contains a rudimentary representation of the *manner* in which the *means* can be used to obtain the *end* pursued by the animal. The proposals outlined above regarding the representations in the fly's multimodal memory and on its minimal map would imply that the fly does indeed possess instrumental concepts. If these proposals are correct, then these concepts could be either sensorimotor concepts or purely kinesthetic concepts.

9.3 Is operant agency a manifestation of underlying beliefs?

In this section, I argue that the existence of map-like representations which underlie the two-way interaction between the animal's self-generated motor output and its sensory inputs during operant conditioning, require us to adopt an *agent-centred* intentional stance. By definition, this presupposes the occurrence of *beliefs* and *desires*. In the case of operant conditioning, the *content* of the agent's beliefs is that by following the pathway, it will attain its goal. The goal is the *object* of its desire.

In the operant conditioning experiments performed on *Drosophila*, it is appropriate to say that the fly *desires* to attain its goal of avoiding the heat. The content of the fly's *belief* is that it will attain its goal *by* adjusting its motor output and/or steering towards the sensory stimulus it associates with the goal. For instance, the fly may believe that *by* staying in a certain zone, it can avoid the heat.

9.3.1 Why use an agent-centred stance to account for operant agency?

This *means-end* schema allows us to explain why an *agent-centred* mentalistic account of operant conditioning is to be preferred to a *goal-centred* intentional stance. A *goal-centred* stance has only *two* components: an animal's *goal* and the *information* it has which helps it attain its goal. The animal's goal-seeking behaviour is triggered by the information it receives from its environment. Even in cases where the animal performs a sequence of activities to attain some goal (as occurs in the nest-building activities of some wasps), each step in the process acts as a “releaser” for the performance of the next one. Thus a two-step schema is adequate.

By contrast, our account of operant conditioning involves a three-step schema, as it includes not only *information* (about the animal's present state and end state) and a *goal* (or end state), but also an *internal representation* of the *means* or *pathway* by which the animal can steer itself from where it is now towards its goal - that is, the sequence of movements and/or sensory stimuli that guides it to its goal. In this account of operant conditioning, the animal uses its memory of this "pathway" to continually fine-tune its motor patterns and correct any "overshooting".

A sceptic might object: "Why should a fine-tuned movement be called an *action*, and not a *reaction?"* The reason is that fine-tuned movement is *self-generated:* it originates from *within* the animal's nervous system, instead of being triggered from without. The fly's *efference copy* enables it to monitor its own bodily movements whereby the animal's nervous system sends out impulses to a bodily organ (Legrand, 2001), and it receives sensory feedback (via the visual display and the heat beam) when it varies its bodily movements. The animal also *takes the initiative* when it subsequently compares the fine motor output from the nervous system with its sensory input, until it finds a positive correlation (Wolf and Heisenberg, 1991, p. 699; Brembs, 1996, p. 3). Talk of *action* is appropriate here, because of the two-way interplay between the agent adjusting its motor output and the new sensory information it receives from its environment. Wolf and Heisenberg (1991, quoted in Brembs, 1996, p. 3, italics mine) define operant behaviour as "the *active choice* of one out of several output channels in order to minimize the deviations of the current situation from a desired situation", and operant conditioning as a more permanent behavioural change arising from *"consistent control* of a sensory stimulus."

These points should go some of the way towards answering the objections of Varner (1998), Gould (2002, p. 41) and Carruthers (2004a), who regard association as too mechanical a process to indicate the presence of mental states.

But the sceptic might still object: “Why should we call the fly's internal representation a *belief?"* To begin with, there are three strong parallels between the fly’s internal representation and that used by an intentional agent. First, the fly's internal representation *tracks the truth* in a robust sense: it not only *mirrors* a state of affairs in the real world, but *changes* whenever the situation it represents varies. The fly's internal representation changes if it has to suddenly learn a new pathway to attain its goal. Indeed, the fly's *self-correction* can be regarded as a kind of *rule-following* activity. Heisenberg, Wolf and Brembs (2001, p. 3) contend that operant behaviour can be explained by the following rule: "Continue a behaviour that reduces, and abandon a behaviour that increases, the deviation from the desired state."

Second, the way the internal representation *functions* in explaining the fly's behaviour is similar in important respects to the behavioural role played by human beliefs. *If* Ramsey's account of belief is correct, then the primary function of our beliefs is to serve as *maps* whereby we steer ourselves. We have argued that the fly's internal representations serve as *minimal maps*, allowing it to reach its goals and avoid harm.

Third, the animal’s *means-end* representation (or *minimal map)* by which it steers itself, is formed by a process under its *control.* The *fine-tuning* process whereby an animal controls its behaviour was described in chapter six.

Now, if we accept Ramsey's account of beliefs as typically being "maps … by which we steer" (Ramsey, 1990, p. 146), then it follows that the fly's minimal maps belong in the same category as beliefs. Given these robust parallels between the fly's behaviour and that of an intentional agent, we have two options: we can either ascribe beliefs to flies on the grounds that their behaviour conforms to the Ramseyan paradigm, or reject the Ramseyan account of belief. The sceptic is bound to choose the latter course. In that case, the sceptic owes us an explanation of why the Ramseyan account of belief is an inadequate one.

Another argument for regarding the fly's behaviour as a manifestation of an underlying *belief* is that it is governed by *conditions of satisfaction,* in the *same* way that a belief is. We examined the phenomenon of *intentionality* in chapter one. This is one of the hallmarks by which some philosophers have endeavoured to unify mental states. According to Searle's (1983, 1999) influential account, intentional states can be characterised by their *conditions of satisfaction.* We demonstrated in chapter seven that operant behaviour also has conditions of satisfaction, and we also showed in this chapter that the conditions of satisfaction are grounded in the way the animal updates its *minimal map*, thereby refining its internal representations which govern its *future* as well as its *present* motor behaviour, in a manner akin to the way we revise the beliefs which guide our actions. Moreover, the intentionality we are talking about here is *intrinsic, not derived:* we can identify the animal's own goals from its rule-following behaviour, even if we do not know exactly what its biological ends are. We can now formulate the following argument for a mentalistic interpretation of the fly's behaviour:

(1) the intentionality that characterises mental states can be defined in terms of its unique kind of

normativity;

(2) the behaviour of the fly exhibits the same kind of normativity as mental states, as the fly's minimal map is subject to *conditions of satisfaction;* therefore

(3) the fly's behaviour qualifies as the manifestation of a mental state.

We are not done yet. To establish that flies actually have beliefs, we need to show that a belief-based, *agent-centred* account is scientifically superior to a *goal-centred* intentional stance in some way. In other words, we need to show that it makes *predictions* that the latter account does not.

9.3.2 The scientific advantage of a mentalistic account of operant conditioning

Consider the following scenario from the 1980s. It is twilight, and you are walking back to the parking lot where you parked your car. You walk up to the front door of your car, take your key ring out of your pocket, select your car key, insert it into the lock, and turn the key, confidently expecting the lock to open. Nothing happens. Annoyed, you jiggle the key around in the lock, a little more vigorously this time. Still nothing. You try a third time. Nothing. Finally, you realise what has happened. Your key ring has several keys, one of which is similar in shape to your car key – similar enough to fit in the lock, but not a good enough fit to open the car door. In the poor lighting conditions, you selected the wrong key by mistake. You now choose the right one and unlock the front door of your car.

Your realisation that you selected the wrong key by mistake the first time you tried to open the front door of your car required an insight on your part. However, the behaviour you engaged in when the lock didn't open at first – jiggling the key around a little more vigorously, and trying again, before giving up – required no insight or even conscious thinking. All it required was persistence. It was natural behaviour for someone who *believed* that the key would unlock the front door. You did not simply give up and try another key; you persisted for a while, because you believed you had selected the right one for the job. Your belief can be characterised as a kind of *commitment* to following a certain procedure when you unlock a car door, because this procedure forms part of your mental motor map of how to unlock a door.

Now let us imagine a fly, seeking a goal that it has learned to procure by refining one of its motor patterns. Each basic pattern is like one of the keys on your ring. The fly fine-tunes its movements, in accordance with its minimal map. But nothing happens, because the circumstances in the fly's environment have changed. Its map is out of date. What does the fly do next?

If we think of the fly as an intentional agent that continually probes its environment, modifies its beliefs and fine-tunes its movements in order to obtain what it wants, we can formulate predictions about what it will do if unexpected changes occur. Because it *believes* that fine-tuning is the key to getting what it wants, and expects to obtain the goal in the same way it previously did, an unsuccessful attempt will result in the animal's "trying out" the same move again, and if this does not work, trying out minor variations on the same move (equivalent to jiggling the key around in the lock a little more vigorously than usual). In other words, initial repetition of the same move (persistence), followed by *minor variations* in the original fine-tuned movement are two behaviours that an animal with beliefs would be expected to engage in, before trying out something radically different (selecting a new motor pattern, just as you eventually selected a different key). This pattern of persistence followed by minor variation was not consciously driven: it arose naturally from your c*ommitment* to acting in that way, because of your underlying *belief* that it would work.

Flies are not phenomenally conscious, but I would expect them to exhibit the same pattern of *commitment*, simply from having a minimal map. For this reason, I would argue that we can legitimately speak of a belief here.

By contrast, an animal which learns to get what it wants by instrumental conditioning does not fine-tune its movements, so it should do one of two things if the move suddenly fails: persist irrationally with *no variation* at all in the conditioned behaviour; or (eventually) give up and try something different (radical change). One would not expect it to make minor variations in its movements, because its goal-seeking behaviour is grounded in a simple association between one of the basic motor patterns in its repertoire and the attainment of the goal, whereas the fly's behaviour is grounded in a belief about the appropriate means it should use to obtain its ends.

Here, then, is an empirical test whereby one could differentiate between the belief-governed behaviour of an agent and the conditioned behaviour of an animal that has learned to make a simple association. How does the fly vary its motor behaviour in the first fraction of a second after it fails to achieve a goal that it has learned to associate with a fine-tuned bodily movement? If it conforms to the pattern of persistence followed by minor variations, then this would be evidence that it has beliefs, in the sense in which I have defined them.

We should always keep in mind that beliefs come in varying degrees of sophistication: whereas in the scenario described above, all it required on your part to abandon a futile course of action was a sudden flash of insight ("Aha! Wrong key!"), the fly is incapable of making such a leap. (As we saw in chapter eight, honey bees are the only insects that appear to possess any capacity for insight.) The fly's abandonment of an unsuccessful course of action is presumably governed by its internal programs. A fly, in other words, is capable of revising its beliefs, but lacks any capacity to critically evaluate them.

9.3.3 Carruthers' cognitive architecture for beliefs

Carruthers (2004a) argues that the presence of a mind is determined by the animal's cognitive architecture - in particular, whether it has beliefs and desires. Carruthers is prepared to regard any insect that can find its way about on a mental map as acting on a belief (2004a). The mental maps Carruthers has in mind are spatial ones; nevertheless, if the account of operant agency I am defending here is correct, then an insect undergoing operant conditioning is guided by its own mental map. In section 9.4.2, I discuss Carruthers' proposed architecture and caution against his assumption that there is a single *core cognitive architecture* underlying all kinds of minds.

9.3.4 Varner's arguments against inferring mental states from conditioning

Varner (1998) maintains that animals that are genuinely learning should be able to form primitive *hypotheses* about changes in their environment.

I would argue that Varner has set the bar too high here. Forming a *hypothesis* is a more sophisticated cognitive task than forming a *belief*, as: (i) it demands a certain degree of creativity, insofar as a hypothesis is an attempt to explain the facts; (ii) for any hypothesis, there are alternatives, which are also consistent with the facts.

I would also like to point out that some animals that are capable of operant agency engage in a very sophisticated form of trial-and-error learning which is strongly reminiscent of hypothesis formation. In section 9.7, I describe a particularly impressive case: the behaviour of the jumping spider *Portia* (Wilcox, 2002), whose flexible trial-and-error learning processes, apparent ability to plan ahead, and continual updating of its own cognitive map place it among the foremost of invertebrate groups in the cognitive arena.

Varner (1998) proposes (following Bitterman, 1965) that tests of *reversal learning* offer a good way to test animals' abilities to generate hypotheses. I discussed this proposal in chapter eight, and concluded that the capacity for *rapid reversal learning* in an animal did not, by itself, warrant the ascription of mental states to it (Conclusion 8.2). I also examined arguments that creatures which show improvements in *serial reversal learning* are capable of *meta-learning*, insofar as they have to develop primitive *hypotheses* about changes in their surroundings, and tentatively concluded (Conclusion 8.3) that in these cases, the animal does indeed appear to be adjusting to sudden changes in its environment, by rapidly revising its expectations. However, I should add that only one species of insect – the honeybee – has actually been shown to be capable of improvements in serial reversal learning.

Finally, it should be borne in mind that even if a capacity for improvements in serial reversal learning proves to be a *sufficient* condition for intentional agency, that does not make it a *necessary* one.

9.3.5 Criticising the belief-desire account of agency

Bittner (2001) has argued that neither belief nor desire can explain why we act. A belief may convince me that something is true, but then how can it also steer me into action? A desire can set a goal for me, but this by itself cannot move me to take action. (And if it did, surely it would also steer me.) Even the combination of belief and desire does not constitute a reason for action. Bittner does not deny that we act for reasons, which he envisages as historical explanations, but he denies that internal states can serve as reasons for action.

If my account is correct, the notion of an *internal map* answers Bittner's argument that belief cannot convince me and steer me at the same time. As I fine-tune my bodily movements in pursuit of my object, the sensory feedback I receive from my probing actions shapes my beliefs (strengthening my *conviction* that I am on the right track) and at the same time *steers* me towards my object.

A striking feature of my account is that it makes *agency, or control, prior to the acquisition of belief:* the agent manages to control its own body movements, and in so doing, *acquires* the belief that moving in a particular way will get it what it wants.

Nevertheless, Bittner does have a valid point: the impulse to act cannot come from belief. In the account of agency proposed above, the existence of innate goals, basic motor patterns, exploratory behaviour and an action selection mechanism - all of which can be explained in terms of a goal-centred intentional stance - were simply assumed. This suggests that operant agency is built upon a scaffolding of innate preferences, behaviours and motor patterns. These are what initially moves us towards our object.

Bittner's argument against the efficacy of *desire* fails to distinguish between desire for the *end* (which is typically an innate drive, and may be automatically triggered whenever the end is sensed) and desire for the *means* to it (which presupposes the existence of certain beliefs about how to achieve the end). The former not only includes the goal (or end), but also moves the animal, through innate drives. In a similar vein, Aristotle characterised locomotion as "movement started by the object of desire" (*De Anima* 3.10, 433a16). However, desire of the the latter kind presupposes the occurrence of certain *beliefs* in the animal. An object X, when sensed, may give rise to X-seeking behaviour in an organism with a *drive* to pursue X. This account does not exclude desire, for there is no reason why an innate preference for X cannot also be a desire for X, if it is accompanied by an *internal map.* Desire, then, may move an animal. However, the existence of an internal map can only be recognised when an animal has to fine-tune its motor patterns to attain its goal (X) - in other words, when the attainment of the goal is not straightforward.

9.3.6 Is my account falsifiable?

The proposal that *fruit flies* are capable of undergoing operant conditioning would be refuted if a simpler mechanism (e.g. instrumental conditioning) were shown to be able to account for their observed behaviour in flight simulator experiments.

Likewise, the *theoretical* basis of my account of operant conditioning would be severely weakened by the discovery that there is no hard-and-fast distinction, at the neurological level, between *instrumental* and *operant* conditioning in animals undergoing conditioning.

Finally, my account would also be falsified by the discovery that flies’ pattern of motor behaviour when circumstances suddenly change is no different from that of animals trained by merely instrumental conditioning. In section 9.3.2, I predicted that when the conditions for obtaining a desired end are suddenly altered, flies should persist in following the old motor pattern, with *minor variations* (persistence in a belief), before they abandon it and select a completely different motor pattern. If this turns out not to be the case, I would then have to conclude that operant agency is *not* an indicator of belief, after all. The next logical place to look for belief would be in the *insight-governed* behaviour of bees, which as we saw in chapter eight are capable of abstracting rules.

9.4. Navigational Agency

In this section, I argue that the navigational behaviour of some insects is essentially similar to operant agency. Accordingly, I have proposed a set of conditions for identifying the specific features of an animal's navigational behaviour which manifest an underlying belief on its part.

9.4.1 Definition of Navigational Agency

The following definition has been adapted somewhat from the definition of operant agency above.

Definition - "Navigational Agency"

We are justified in ascribing *agency* to a navigating animal if the following features can be identified:

Innate behaviour

(i) *innate preferences*;

(ii) *innate motor programs*, which are stored in the brain, and generate the suite of the animal's motor output;

(iii) a tendency on the animal's part to engage in *exploratory behaviour*, in order to locate food sites;

Sensory and discriminatory capacities

visual *sensory inputs* (from "true" senses) that inform the animal about its current position, in relation to its long-term goal, and enable it to correct its movements if the need arises;

Memory

an ability to learn new spatial routes to obtain goals such as food, and store these routes in its *procedural memory*;

Learning-related requirements: associations

direct or indirect *associations*, of two kinds: (a) between visual landmarks and local vectors; and (b) between the animal's short term goals (landmarks) and long term goals (food sites or the nest). These associations are stored in the animal's memory and updated when circumstances change. Unlike the case of operant conditioning, *no* temporal correlation mechanism is required;

Representation: minimal map

an internal representation *(minimal map)* which includes the following features:

> (i) the animal's *current motor output* (represented as its *efference copy*);
>
> (ii) the animal's *current goal* (represented as a stored memory of a visual stimulus that the animal associates with the attainment of a "reward" - usually a distant food source); and *sub-goals* (represented as stored memories of visual landmarks, which the animal uses to steer itself towards its goal); and
>
> (iii) the animal's *pathway* to its current goal, via its *sub-goals* (represented as a stored memory of the sequence of visual landmarks which enable the animal to steer itself towards its goal, as well as a sequence of vectors that help the animal to steer itself from one landmark to the next);

Action selection, fine tuning and ability to compare input vs. output

(i) an *action selection mechanism*, which allows the animal to make a selection from its suite of possible motor response patterns and pick the one that is the most appropriate to its current circumstances;

(ii) *fine-tuning behaviour:* efferent motor commands which are capable of steering the animal in a particular direction - i.e. towards food or towards a visual landmark that may help it locate food;

(iii) the ability to store and compare *internal representations* of its current motor output (i.e. its efferent copy, which represents its current "position" on its internal map) and its afferent sensory inputs. Motor output and sensory inputs are linked by a two-way interaction;

Self-correction of behaviour and beliefs

self-correction, that is:

> (i) an ability to *rectify any deviations* (or mismatches) between its view and its internally stored image of its goal or sub-goal - first, in order to *approach* its goal or sub-goal, and second, in order to keep *track* of it;
>
> (ii) *abandonment* of behaviour that increases, and *continuation* of behaviour that reduces, the animal's "distance" (or deviation) from its current goal; and
>
> (iii) an ability to *form new associations* and alter its internal representations (i.e. *update its minimal map)* in line with variations in surrounding circumstances that are relevant to the animal's attainment of its goal.

Primitive Concepts

a primitive *instrumental concept* of a landmark, represented as a useful guide to the animal’s current goal.

If the above conditions are all met, then the animal can be said to exhibit what I will call *navigational agency.* Such an animal qualifies as an *intentional agent* which *believes* that it will get what it *wants*, by doing what its internal map tells it to do.

9.4.2 Case study: Navigational Agency in insects

The current state of research into spatial learning in insects remains fluid. What is not disputed is that some insects (especially social insects, such as the ants, bees and wasps) employ a highly sophisticated system of navigation, and that they employ at least two mechanisms to find their way back to their nests: *path integration* (also known as dead reckoning) and memories of *visual landmarks*. A third mechanism - *global* (or *allocentric) maps* - has been proposed for honey bees, but its status remains controversial, so I will not discuss it here (see Bennett (1996); Menzel, Brandt, Gumbert, Komischke and Kunze (2000); Giurfa and Capaldi (1999); Giurfa and Menzel (2003); Collett and Collett (2002); Harrison and Schunn (2003) for a discussion of the issues).

I argue in section (a) that path integration does not require the use of *minimal maps* and hence does not qualify as a case of intentional agency. In section (b), I examine the way insects use *visual landmarks* to steer by. The evidence suggests that this kind of navigation does indeed require "minimal maps". Additionally, the continual self-monitoring behaviour of insects navigating by landmarks suggests that they are indeed in *control* of their bodily movements, and that they construct their own visual maps using a flexible learning process. I conclude that insects that navigate using visual landmarks are *bona fide* intentional agents.

(a) Path integration (dead reckoning)

The best studied mechanism is *path integration*, which allows insects to navigate on bare terrain, in the absence of visual landmarks. Collett and Collett (2002, p. 546) describe it as follows:

> When a honeybee or desert ant leaves its nest, it continually monitors its path, using a sun and polarized light compass to assess its direction of travel, and a measure of retinal image motion (bees) or motor output (ants) to estimate the distance that it covers. This information is used to perform path integration, updating an accumulator that keep a record of the insect's net distance and direction from the nest (2002, p. 546).

The insect's *working memory* of its "global" position - its distance and direction from the nest - is continually updated as it moves. An insect's global vector allows it to return home in a straight line from any point in its path. Tests have shown that if a desert ant returning to its nest is moved to unfamiliar terrain, it continues on the same course for a distance roughly equal to the distance it was from the nest. Finding no nest, it then starts a spiral search for the nest. If obstacles are placed in its path, the ant goes around them and adjusts its course appropriately (Collett and Collett, 2002, p. 546; Gallistel, 1998, p. 24; Corry, 2003). When ants have to navigate around obstacles, they memorise the sequence of motor movements corresponding to movement around the obstacle, thereby cutting their information processing costs (Schatz, Chameron, Beugnon and Collett, 1999).

Can we describe the navigation systems of insects as representational? If so, is it a map-like representation, like the one I proposed for operant conditioning, and does it qualify an insect to be a belief-holder?

Ramsey proposed that a "belief of the primary sort is a map of neighbouring space by which we steer" (1990, p. 146). However, if we look at the path integration system alone, the representation fails to meet even the minimal conditions which I suggested in the section above for a *minimal map*, as neither the *goal* nor the *current position* is represented: only the directional displacement from the goal is encoded. Without an internal representation of one's goal, we cannot speak of agency, or of control. A naive foraging insect, relying solely on path integration, is merely following its internal compass, which is continually updated by events beyond its control.

The same remarks apply to the ability of *monarch butterflies* to navigate from as far north as Massachusetts to central Mexico, in autumn. Scientists have recently discovered that the monarch uses an in-built sun-compass and a biological clock to find its way (Mouritsen and Frost, 2002). Once again, the butterflies appear to be lack any internal representation of their goal (a place with warm, tropical weather where they can spend the winter, as they do not hibernate) or their destination (Mexico). Thus we do not need to characterise these butterflies as intentional agents when describing how they migrate.

(b) Landmark navigation

The other way by which insects navigate is the use of *landmarks* which they observe en route and at their home base. Naive foragers initially rely on path integration to find their way home, but when they repeat their journey, they *learn* the appearance of new landmarks and *associate* local vectors with them, which indicate the distance to the next landmark. With experience, these local vectors take precedence over the global vectors used in dead reckoning.

> [T]he basic structure of each route segment is a landmark and an associated local vector or some other stereotyped movement (for example, turning left)... [T]he primary role of a landmark is to serve as a signpost that tells the insect what to do next, rather than as a positional marker... (Collett and Collett, 2002, p. 547).

It is still not known how insects encode landmarks, and what features of the image are stored in its memory. However, it is agreed that ants, bees and wasps not only memorise landmarks, but guide their final approach to their goal by matching their visual image of the landmark with their *stored image* (snapshot?) of how it looks from the goal. It appears that when comparing their current view to a stored image, they use the retinal positions of the edges, spots of light, the image's centre of gravity and colour. They also learn the appearance of an object from more than one distance, as their path home is divided into separate segments, each guided by a separate view of the object. Additionally, *consistency of view* is guaranteed because the insect, following the sun or some other cue, always faces the object in the same direction. Finally, the insect's view of the distant panorama from a landmark can help to identify it (Collett and Collett, 2002).

Can landmarks serve as maps to steer by?

Path integration on its own works perfectly well even for an untrained insect: it does not require the insect to associate a motor pattern or sensory stimulus with attaining its goal. Navigation by landmarks, on the other hand, requires extensive *learning*. The location and features (colour, size, edge orientation and centre of gravity) of each landmark have to be memorised. Multiple views of each landmark have to be stored in the insect's brain. Additionally, some insects use panoramic cues to recognise local landmarks. Finally, a local vector has to be associated with each landmark (Collett and Collett, 1998, 2002). The fact that insects are capable of learning new goals and new patterns of means-end behaviour means that they satisfy a necessary condition for the ascription of mental states - though by itself not a sufficient one, as we have seen that associative learning can take place in the absence of mental states.

While navigating by landmarks, insects such as ants and bees can learn to associate their final goal (the nest) with views of nearby landmarks, which guide them home, even when their final goal is out of sight. These insects often follow "a fixed route that they divide into segments, using prominent objects as sub-goals" (Collett and Collett, 2002, p. 547).

It is a matter of controversy whether insects possess a *global* (or *allocentric) map* of their terrain, which combines multiple views and movements in a common frame of reference (see Giurfa and Menzel, 2003; Collett and Collett, 2002; Harrison and Schunn, 2003; Giurfa and Capaldi, 1999; Gould, 1986, 2002). However, even if an insect has "only a piecemeal and fragmented spatial memory of its environment" (as suggested by Collett and Collett, 2002, p. 549), it clearly meets the requirements for a *minimal map*. Its *current position* is represented by the way its nervous system encodes its view of the external world (either as a visual snapshot or as a set of parameters), its short-term *goal* is the landmark it is heading for, and the *path* is its local vector, which the insect recalls when the landmark comes into view (Collett and Collett, 2002, pp. 546, 547, 549). The map-like representation employed here is a *sensory map*, which uses the visual modality.

What are the goals of navigating insects?

For a navigating insect, its long-term goals are food, warmth and the safety of the nest, all of which trigger innate responses. Short term goals (e.g. landmarks) are desired insofar as they are associated with a long term goal. Even long term goals change over time, as new food sources supplant old ones.

Additionally, insects have to integrate *multiple goals*, relating to the different needs of their community (Seeley, 1995; Hawes, 1995). In other words, they require an action selection mechanism. For example, a bee hive requires a reliable supply of pollen, nectar, and water. Worker field bees can assess which commodity seems to be in short supply within the hive, and search for it.

Are navigating insects agents?

The statement by Collett and Collett that "the primary role of a landmark is to serve as a signpost that tells an insect what to do next" (2002, p. 547), recalls Ramsey's claim that a belief is a map by which we steer. Should we then attribute agency, beliefs and desires to an insect navigating by visual landmarks?

Corry (2003) thinks not, since the insect is not consciously manipulating the symbols that encode its way home. It does not calculate where to go; its nervous system does. Corry has a point: representations are not mentalistic *per se*. As we have seen, the autonomic nervous system represents, but we do not say it has a mind of its own. Nevertheless, Corry's "consciousness" requirement is unconvincing, as he seems to tacitly equate mind with consciousness – a controversial philosophical position, to say the least.

I would suggest that Ramsey's *steering metaphor* can help us resolve the question of whether insects navigate mindfully. The word "steer", when used as a transitive verb, means "to control the course of" (Merriam-Webster Online, 2006). Thus steering oneself suggests *control* of one's bodily movements - a mindful activity, on the account I am developing in this chapter. The autonomic nervous system, as its name suggests, works perfectly well without us having to control it by fine-tuning. It is presumably mindless. On the other hand, *Drosophila* at the torque meter needed to control its motor movements in order to stabilise the arena and escape the heat. Which side of the divide does insect navigation fall on?

Earlier, I proposed that an animal is *controlling* its movements if it can compare the *efferent copy* from its motor output with its incoming *sensory inputs*, and make suitable adjustments. The reason why this behaviour merits the description of "agency" is that fine-tuned adjustment is *self-generated:* it originates from within the animal's nervous system, instead of being triggered from without. This is an internal, neurophysiological measure of control. The occurrence of internal representations required for such control (e.g. an instantiation of the model postulated by Merfeld, 2001) could easily be confirmed empirically for navigating insects.

In the meantime, one could use external criteria for the existence of control: *self-correcting patterns of movement*. The continual self-monitoring behaviour of navigating insects suggests that they are indeed in control of their bodily movements. For instance, wood ants subdivide their path towards a landmark into a sequence of segments, each guided by a different view of the same object (Collett and Collett, 2002, p. 543). Von Frisch observed that honeybees tend to head for isolated trees along their route, even if it takes them off course (Collett and Collett, 2002,

p. 547). Insects also correct for changes in their environmental cues.

Two facts may be urged against the idea that a navigating insect is exercising control over its movements, when it steers itself towards its goal. First, insects appear to follow *fixed routines* when selecting landmarks to serve as their sub-goals (Collett and Collett, 2002, p. 543). Second, *cues* in an insect's environment (e.g. the panorama it is viewing) may determine what it remembers when pursuing its goal (e.g. which "snapshot" it recalls - see Collett and Collett, 2002, p. 545). In fact, it turns out that no two insects' maps are the same. Each insect's map is the combined outcome of:

(i) its exploratory behaviour, as it forages for food;
(ii) its ability to learn about its environment;
(iii) the position and types of objects in its path;
(iv) the insect's innate response to these objects; and
(v) certain fundamental constraints on the kinds of objects that can serve as landmarks (Collett and Collett, 2002, pp. 543, 548-549).

In reply to the first objection: the fact that certain aspects of insect navigation are determined by fixed routines in no way implies that the entire ensuite of such behaviour is hard-wired. (In a similair vein, Carruthers (2004a, pp. 10-12, online PDF version) argues that even though many insects, notably the Sphex wasp and the Australian digger wasp, exhibit what he calls triggered fixed action sequences, it does not follow that all insect behaviour can be accounted for in this way.) The second objection also misses the mark. It is simplistic to claim that an insect's *environment* determines its map, since in fact, each insect has its own *learning history* and foraging behaviour.

We argued above that control requires explanation in terms of an *agent-centred* intentional stance, as a goal-centred intentional stance is incapable of encoding the *two-way interplay* between the agent adjusting its motor output and the new sensory information it is receiving from its environment, which allows it to correct itself. In a *goal-centred* stance, the animal's goal-seeking behaviour is triggered by a one-way process: the animal receives information from its environment.

Finally, we cannot speak of agency unless there is *trying* on the part of the insect. The existence of *exploratory behaviour*, coupled with self-correcting patterns of movement, allows us to speak of the insect as trying to find food. Studies have also shown that when the landmarks that mark an insect's feeding site are moved, insects try to find the place where their view of the landmarks matches the view they see from their goal.

The above findings lead me to conclude that visual navigation using landmarks does indeed constitute a form of intentional agency, subject to the list of conditions laid down in section 9.4.1 above.

The basic cognitive architecture that I have proposed for navigational and other forms of agency differs in four important respects from that of Carruthers (2004a), who proposes that a navigating animal’s perceptual states inform its belief states, which interact with its desire states, to select from an array of action schemata, which determine its motor behaviour. On my proposal, the initial selection is not mediated by beliefs, which only emerge when the animal fine-tunes its selected action schema in an effort to obtain what it wants. Second, I propose a two-way interaction between motor output and sensory input. Third, the mental maps discussed by Carruthers are spatial ones, whereas I also allow for motor or sensorimotor maps. Fourth, Carruthers' model applies only to navigation; I recognise four equally basic kinds of primitive agency: operant agency, navigational agency, tool agency and social agency.

9.5 Tool Agency

Beck’s (1980) definition of tool use in connection with his research on non-human primates is still widely cited in the literature, so I shall use it in my definition of tool agency below:

> [T]ool use is the external employment of an unattached environmental object to alter more efficiently the form, position or condition of another object, another organism, or the user itself when the user holds or carries the tool during or just prior to use and is responsible for the proper and effective orientation of the tool (1980, p. 10).

The key features of this definition are that in order to qualify as a tool user, an animal must be able to modify, carry or manipulate an item external to itself, before using it to effect some change in the environment (Mather and Anderson, 1998). I would like to note in passing that Beck's yardstick, taken by itself, cannot tell us whether the tool-using behaviour it describes is *fixed* or *flexible*, and therefore is an insufficient criterion for intentional agency.

9.5.1 Definition of Tool Agency

Definition - Tool Agency

An animal can be described as using a tool intentionally if the following features can be identified:

Innate behaviour:

(i) *innate preferences* or drives;

(ii) *innate motor programs*, which are stored in the brain, and generate the suite of the animal's motor output;

(iii) a tendency on the animal's part to engage in *exploratory behaviour*, by using its tools to probe its environment;

Sensory and discriminatory capacities:

sensory inputs (from “true” senses), of a visual and/or tactile kind, which inform the animal whether it has attained its goal with its tool, and if not, whether it is getting closer to achieving it;

Memory:

an ability to acquire and store new *motor skills* relating to tool manipulation, in its *procedural memory;*

Learning-related requirements: associations and a correlation mechanism:

(i) *associations* between different tool-using motor commands and their consequences, which are stored in the animal's memory;

(ii) the ability to store and compare internal representations of its *current motor output* while using the tool (i.e. its efferent copy, which represents its current "position" on its internal map) and its *afferent sensory inputs*;

(iii) a *correlation mechanism*, allowing it to find a temporal coincidence between its motor behaviour and the attainment of its goal;

Representation: minimal map:

an internal representation *(minimal map)* which includes the following features:

> (i) the animal's *current motor output* (represented as its *efference copy*) – i.e. the animal knows what move it is making with its tool;

(ii) the animal's *current goal* or *end-state* (represented as a stored visual and/or tactile memory involving the goal sought by the animal and its procurement with a tool that the animal has learned to associate with attaining its goal) – i.e. the animal remembers previously obtaining the goal with the tool; and

(iii) the animal's *pathway* to its current goal (represented as a stored memory of a sequence of movements, coupled with sensory feedback, which allows the animal to steer its tool towards its goal) – i.e. the animal remembers how to manipulate the tool to get the goal and the feedback it will get along the way;

Action selection, fine tuning and ability to compare input vs. output:

(i) an *action selection mechanism*, which allows the animal to make a selection from its suite of possible motor response patterns and pick the one that is the most appropriate for the tool it is using and object it is used to get;

(ii) *fine-tuning behaviour:* an ability to stabilise one of its motor patterns within a narrow range of values, to enable the animal to achieve its goal by using the tool;

Self-correction of behaviour and beliefs:

self-correction, that is:

(i) an ability to *rectify any deviations* in motor output from the range which is appropriate for attaining the goal – i.e. the animal can adjust the tool;

(ii) *abandonment* of behaviour that increases, and *continuation* of behaviour that reduces, the animal's "distance" (or deviation) from its current goal – i.e. the animal stops using the tool if it is a hindrance and keeps using it if it helps the animal to obtain what it wants; and

(iii) an ability to *form new associations* and alter its internal representations (i.e. *update its minimal map)* in line with variations in surrounding circumstances that are relevant to the animal's attainment of its goal – i.e. the animal can learn to make new moves with its tool;

Primitive concept:

a *primitive instrumental concept* of a tool, which represents the manner in which it can be manipulated in order to obtain the animal's current goal.

If the above conditions are all met, then we can legitimately speak of the animal as an *intentional agent* which *believes* that it will get what it *wants*, by doing what its internal map tells it to do.

9.5.2 Case study: Tool agency in cephalopods

Taxonomy and comparative anatomy

Cephalopods are a class of the phylum Mollusca (molluscs) and are therefore related to bivalves, scallops, oysters, clams, snails and slugs, tusk shells and chitons. Cephalopods include the pelagic, shelled nautiloids and the coeleoids (cuttlefish, squid and octopods, the group to which octopuses belong).

Among the molluscs, there is an enormous degree of variability in the complexity of the nervous system. Cephalopods are renowned for their large brains, while other molluscs (e.g. bivalves) lack even a head, let alone a proper brain. Most molluscs have a relatively "simple" central nervous system, with five or six pairs of ganglia.

> In the cephalopods alone among the molluscs, evolution has also constructed a brain. It has greatly expanded the forwardmost pairs of ganglia and moved them closer together to create a tightly packed mass of lobes that lies between the eyes and encircles the oesophagus (Hamilton, 1997, p. 32).

The brain-to-body weight ratios of cephalopods exceed those of other invertebrates, as well as most fish and reptiles. Additionally, their brains are anatomically complex. However, mammals and birds far outstrip cephalopods in the complexity of their brains (Anderson and Wood, 2001; Hamilton, 1997; Seth, Baars and Edelman, 2005).

In contrast with molluscs such as clams and oysters, which are passive filter feeders, cephalopods live in a challenging environment, where they have to hunt down mobile prey and avoid predators. They have sophisticated sense organs, a complex rapid movement system, an ability to rapidly change colour and (in the case of cuttlefish and squid) a wide range of social signals (Hanlon and Messenger, 1996; Moynihan, 1985; Wells, 1962).

Most cephalopods have very flexible limbs, with unlimited degrees of freedom. Scientists have recently discovered that octopuses *control* the movement of their limbs by using a decentralised system, where most of the *fine-tuning* occurs in the limb itself:

> ...[A]n octopus moves its arms simply by sending a "move" command from its brain to its arm and telling it how far to move.
>
> The arm does the rest, controlling its own movement as it extends.

> "There appears to be an underlying motor program... which does not require continuous central control," the researchers write (Noble, 2001, online).

The learning abilities and adaptive behaviour of cephalopods compare favourably with those of insects and some vertebrates. The following discussion focuses principally on the well-studied common octopus, *Octopus vulgaris*.

Tool agency in octopuses?

According to Beck's criterion for tool use, octopuses use both rocks and water jets from their siphons as tools to modify their environment. After selecting a place for a home, an octopus usually has to modify it extensively to render it suitable for habitation, as it is usually clogged with sand and the shape may not be appropriate. The octopus gathers up excess sand and small rocks and carries them out to the entrance of its new home. Once there, it lets go of the rocks and blasts them all away with a jet of water from its siphon. After removing sand and small rocks from its burrow, an octopus may be left with a large entrance, so it goes outside, picks up some some small rocks and brings them back to the home, piling them up at the entrance (Mather and Anderson, 1998).

Octopuses also use *water jets* to get rid of nuisances:

> After capturing crab prey, the octopus will usually kill them and hold one or several under the arm web, dissolve the cartilage holding the joints together, digest out the meat, and keep the exoskeleton bits. When it's finished, it will take the remains to the den entrance and jet the lot out into what becomes a midden. If a scavenging Serranid

> fish comes by to eat these remains, the octopus may jet a blast of water to remove the 'pest' from the vicinity (Mather and Anderson, 1998, online).

However, the behaviour described above may turn out to be a *fixed pattern* of behaviour, and even Mather and Anderson (1998) admit that octopuses use tools "in a very simple way". While there is no doubt that octopuses are anatomically capable of fine-tuning their motor movements, what needs to be shown here is that they actually do so while manipulating tools.

Although octopuses do not appear to fine-tune their behaviour when cleaning their dens, they aim their jets carefully at passing scavengers or human observers - rather like a water gun (Mather and Anderson, 2000). This could be an instance of fine-tuning, and hence *bona fide* agency.

Flexible behaviour in octopuses?

Mather and Anderson (2000) describe how octopuses will use a variety of techniques to open a clam shell, switching readily from one to another in the event of failure. Giant Pacific octopuses switch strategies to open different shellfish - smashing thin mussels, prying open clams, and drilling tougher-shelled clams. When clams were wired shut with stainless steel wire, the octopuses couldn't pull them apart, so they switched to drilling and chipping. The authors comment:

> They were intelligently adapting the penetration technique to the clam species presented and the situation in which they were placed…
>
> The above interpretation is reasonable. Unfortunately, the range of behaviours involved

> here is too narrow to decide whether the octopuses were acting intentionally or in a hit-and-miss fashion (Mather and Anderson, 2000, online).

Mather and Anderson (1998) describe a recent experiment where two octopuses were able to fine-tune the force of their water jets to control the movement of some coloured toys floating in their tank. The octopuses did not manipulate the toys in a stereotypical fashion; indeed, their behaviour seemed spontaneous, even playful. This tool-using behaviour by octopuses appears to have been both flexible and fine-tuned, but follow-up studies are warranted.

9.6 Social Agency

9.6.1 Definition of Social Agency

Definition - "Agency in a social context"

An animal can be described as displaying agency in a social context if the following features can be identified:

Innate behaviour:

(i) *innate preferences* or drives;

(ii) *innate motor programs*, which are stored in the brain, and generate the suite of the animal's motor output;

(iii) a tendency on the animal's part to engage in *exploratory behaviour*;

Sensory and discriminatory capacities:

(i) *sensory inputs* that inform the animal whether it has attained its current goal, and if not, whether it is getting closer to achieving it;

(ii) the ability to discriminate between individual members of its own species (conspecifics), as well as between members and non-members of its group;

Memory:

the ability to keep track of the status of individuals within one's group, and remember one's past interactions with them *(book-keeping)*. It is an open question as to whether these memories would be encoded in the animal’s *procedural* memory (e.g. as an aggregative resultant of the animal’s prior interactions with individual X, which is updated regularly but contains no “running history”) or in its *declarative* memory (e.g. a memory of past “good” and “bad” turns done by X).

However, the former hypothesis will be assumed here, as it is more parsimonious*;*

Learning-related requirements: associations and a correlation mechanism:

(i) *associations* between stored memories of the different individuals in the animal's group and the (good or bad) consequences of following their example, as well as direct *associations* between different motor commands and their consequences, which are stored in the animal's memory;

(ii) the ability to learn from observing the behaviour of other individuals *(observational learning)* and to acquire new knowledge that is specific to one's group (traditions);

Representation - minimal map:

an internal representation *(minimal map)* which includes the following features:

> (i) the animal's *current state*, which includes both its current spatial relation to its role model, and its current motor output (represented as its *efference copy*);
>
> (ii) the animal's *current goal* (represented as a stored memory of a sensory stimulus which the animal associates with the attainment of that goal); and
>
> (iii) the animal's *pathway* to its current goal (represented as a stored memory of the individual which can reliably lead the animal to its goal - i.e. the role model);
>
> (iv) the animal’s current *role model:* a useful, reliable individual in its group who is to be followed in the pursuit of important objectives, such as food;

Action selection, fine tuning and ability to compare input vs. output:

(i) an *action selection mechanism*, which allows the animal to make a selection from its suite of possible motor response patterns and pick the one that is the most appropriate to its current social setting;

(ii) *fine-tuning* (controlled, modulated activity): the ability to *model* its behaviour on that of a knowledgeable individual (the *role model)*, and to adjust its social behaviour to take account of differences between the individuals in its group, as well as changes in a given individual's behaviour;

(iii) the ability to store and compare internal representations of its current motor output (i.e. its *efferent copy*, which represents its current "position" on its internal map) and its *afferent* sensory inputs;

Self-correction of behaviour and beliefs:

(i) an ability to *rectify any deviations* in its social behaviour from that which is appropriate for attaining its current goal;

(ii) *abandonment* of social behaviour that proves to be unproductive (e.g. when the animal's expectations of another individual are disappointed), and *continuation* of behaviour that helps the animal obtain its current goal; and

(iii) an ability to *form new associations* and alter its internal representations (i.e. *update its minimal map)* in line with variations in surrounding circumstances that are relevant to the animal's attainment of its goal.

Primitive Concepts:

a *primitive instrumental concept* of a knowledgeable individual, representing the manner in which that individual can be used to obtain the animal’s current goal (e.g. “Always keep the individual’s rear in sight, when following that individual”).

If the above conditions are all met, then we can legitimately speak of the animal as an *intentional agent* which *believes* that it will get what it *wants*, by doing what its internal map tells it to do.

9.6.2 Case study: Social Agency in fish

According to Brown and Laland, "research over the last 50 years has demonstrated that social learning is common amongst fish, birds and mammals, and should now be regarded as a regular feature of vertebrate life... Documented cases of social learning in fish are now commonplace" (2003, p. 281). In particular, "social learning plays a role in fish (i) anti-predator behaviour; (ii) migration and orientation; (iii) foraging; and (iv) mate choice, and ... it is facilitated by eavesdropping" (2003, p. 281).

Brown and Laland (2003) explain the rationale for social learning in fishes:

> In making decisions, such as how to find food and mates or avoid predators, many animals utilise information that is produced by others. Such individuals are referred to as 'eavesdroppers' in the signal-receiver literature (McGregor 1993) and 'observers' in the social-learning literature (Heyes and Galef 1996). Socially transmitted information may simply be a by-product of the demonstrating individual's behaviour or a signal targeted towards a particular audience (2003, p. 280).

Bshary, Wickler and Fricke (2002) argue that most kinds of social behaviour interpreted as cognitive in primates can be found in certain fish as well. For instance, features such as individual recognition, acquisition of new behaviour patterns by observational learning, transmission of group traditions, co-operative hunting, tactical deception (cheating), tit-for-tat punishment strategies, reconciliation, altruism and social prestige, formerly thought to be unique to primates or at least mammals, can all be found in fish societies. Specifically, research cited by Bshary, Wickler and Fricke (2002) shows that each kind of behaviour described in my model of social agency above can be found in fish. My only qualification is that (as the authors point out), the evidence is assembled together from *different species* of fish, mainly because there have been relatively few studies of fish cognition to date. Further research is needed to determine whether any particular species satisfies all of the criteria. Still, the current evidence for a rudimentary social intelligence in fish looks strong.

In my model of social agency in fish, I outlined a proposed set of sufficient criteria for agency in a social context, looked at the rationale for social learning, and noted its widespread occurrence across vertebrate species. Evidence that at least some species of fish satisfy my criteria for social agency is provided below. To date, I have not been able to establish whether any *particular* species meets all of the criteria, although the evidence for cleaner fish, sticklebacks and guppies looks impressive.

Behaviour modelled on that of a knowledgeable individual

Individuals (especially juveniles) learn to model their behaviour on that of experienced adults, mainly by accompanying them and observing how they behave (Bshary, Wickler and Fricke, 2002).

Sensory discrimination between individuals and/or categories of individuals

There is abundant evidence in the literature of individual recognition in fish:

> Individual recognition based primarily on optical cues ... has been demonstrated experimentally in a variety of species... There is even evidence that in damselfish, individuals can recognise one another on purely acoustical cues... In summary, individual recognition can safely be assumed to be widespread across fish families (Bshary, Wickler and Fricke, 2002, pp. 2-3).

In addition to individual recognition, cleaning symbiosis provides an example of a case where the ability to categorise individuals on the basis of their observed characteristics is especially useful:

> In cleaning symbiosis, so-called client fish trade the removal of parasites and dead or infected tissue against an easy meal for so-called cleaner fish... Cleaning symbiosis is particularly promising for comparative studies as cleaner fish are found in many different fish families and can differ markedly in the degree to which they depend on interactions with clients for their diet... Full-time cleaners like the cleaner wrasse (*Labroides dimidiatus*) may have about 2,300 interactions per day with clients belonging to over 100 different species... There is strong evidence that cleaners can categorise their 100-or-so client species into resident species that have access to their local cleaner only, due to their small territory or home range, and other species that have home ranges that cover several cleaning stations. As predicted by biological market theory (Noe et al. 1991), clients with choice options between cleaners almost invariably have priority of access over clients without choice at cleaning stations (Bshary, Wickler and Fricke, 2002, p. 5).

Memory for individuals and their track record

The ability to remember individuals over long periods of time is of fundamental importance for social learning. Bshary, Wickler and Fricke (2002) cite evidence that an anemonefish can recognise an individual that it has not seen for 30 days. (So much for the myth that fish have only a 3-second memory!)

According to Bshary, Wickler and Fricke (2002), some fish can also monitor changes in the status of individuals and track relationships within their groups.

Bshary, Wickler and Fricke (2002) describe experiments showing that some fish species are capable of engaging in book-keeping (remembering their partners' behaviour during past interations) with several partners at once:

> The most famous example of co-operation in fish is probably the inspection of nearby predators by one or several fish that leave the relative safety of their school to do so (Pitcher et al. 1986). During inspection, pairs of sticklebacks, *Gasterosteus aculeatus*, and guppies, *Poecilia reticulata*, among others, approach the predator in alternating moves. A series of experiments led to the conclusion that these fish solve a so-called "prisoner's dilemma" (Luce and Raiffa 1957). In a prisoner's dilemma, two players have the option of either co-operating with or cheating their partner. Cheating the partner yields a higher benefit than co-operation irrespective of what the partner does, but if both partners co-operate then they receive a higher benefit than if both cheat, hence the dilemma. Milinski (1987) and Dugatkin (1988) proposed that fish solve the

> prisoner's dilemma by playing a "tit-for-tat" strategy, which states that a player starts co-operatively and does in all further rounds what the partner did in the previous round (Axelrod and Hamilton 1981). This interpretation is not yet entirely resolved (see review in Dugatkin 1997) but discussions about the interpretation led to a few experiments with very interesting additional results. Milinski et al. (1990a) could show that *individual sticklebacks prefer specific partners to others*, which implies that school members recognise each other. In addition, *partners build up trust in each other during repeated inspections*, that is, they hesitate less in approaching a predator when accompanied by a partner that co-operated in the past (Milinski et al. 1990b). Similar results have been found in guppies (see review in Dugatkin 1997). These data imply that these fish species are capable of *book-keeping (remembering their partners' behaviour during past interactions) with several partners simultaneously* (Bshary, Wickler and Fricke, 2002, p. 3, italics mine).

Observational learning of new practices

Juvenile fish learn what to eat by observing adult conspecifics:

> There is some evidence that young fish learn what to eat by observing adults. Fish definitely learn horizontally from conspecifics what to eat under lab conditions. Templeton (1987, unpublished Ph.D. thesis, cited in Suboski and Templeton, 1989) found that juvenile rock bass ... that saw a trained conspecific eating a novel food item would readily consume that food later when, alone, they were tested for the first time. Without prior observations, these juveniles did not attack the prey... (Bshary, Wickler and Fricke, 2002, p. 4).

Hatchery-reared Atlantic salmon acquire new kinds of feeding behaviour and learn to target new kinds of prey, simply by observing knowledgeable conspecifics (Brown, Markula and Laland, 2003).

Social enhancement of foraging has been reported in species as different as salmon, rock bass, Alaska pollack and brown trout (Brown and Laland, 2003).

Fish can also learn *novel techniques* for obtaining food from observation of knowledgeable conspecifics. Juvenile European sea bass learned to press a lever to get food, simply by watching other fish that had been previously trained to do this (Brown and Laland, 2003).

Learning (group traditions)

Schools of fish have their own "traditions" relating to their choice of sites for resting sites, migration routes and food sources, and this knowledge is transmitted through social learning (Bshary, Wickler and Fricke, 2002). For instance, juvenile French grunts learn the migration route from their resting grounds to feeding sites by following older individuals, and bluehead wrasse have prefrred mating sites that stay the same over many generations (Brown and Laland 2003).

Bshary, Wickler and Fricke describe the mechanism by which traditions are perpetuated in guppies:

> Laland and Williams ... conducted laboratory experiments and showed experimentally that guppies learn the way to hidden food sourced from knowledgeable conspecifics.

> The conspecifics had been trained to use only one of two ways to the food source. Naive fish were added and learned the way to the food source by schooling with the others. Members of the original school could be replaced successively and the school still preferentially took the originally learned way to the food source. The fish thus built up a tradition. Using principally the same experimental set up, Laland and Williams ... went one step further and showed that even maladaptive behaviour can spread through a population due to social learning. In their study, a longer and therefore more costly way to a foraging site was still preferred over a short way 3 days after all original trainers had been removed" (Bshary, Wickler and Fricke, 2002, p. 4).

Innate goals

Brown and Laland (2003) mention four general categories of goals, in relation to which *social learning* is known to take place amongst fish: predator avoidance; migration and orientation; foraging for food; and mate choice. There is an ever-growing body of evidence that juvenile fish engage in extensive social learning of skills relating to all of these goals (Laland, Brown and Krause, 2003).

Fine-tuning (controlled, modulated activity):

Individuals carefully tailor their own social behaviour towards an individual, in accordance with their observations of that individual's past interactions with other individuals.

> Male Siamese fighting fish ... monitor aggressive interactions between neighbouring conspecifics and use the information on relative fighting ability in subsequent aggressive interactions with the males they have observed...(Brown and Laland, 2003,

p. 285).

Individuals also adjust their behaviour towards a specific individual on the basis of their own *previous interactions* with that individual - a practice known as *book-keeping.* Cleaner fish engage in book-keeping: they provide better than average service to dissatisfied clients that "punished" (aggressively chased) them during their last interaction. As Bshary, Wickler and Fricke (2002) point out, punishment can only work if there is *individual recognition.* This means that cleaner fish must be able to keep track of the behaviour of each their clients (up to 100 individuals!), and modulate their behaviour towards each of them.

Cleaner fish also provide tactile stimulation to predatory clients, possibly as a form of pre-conflict management, or towards clients it has cheated in the past (Bshary, Wickler and Fricke, 2002).

Additionally, cleaner fish behave much more attentively (or "altruistically") towards their clients *if they are being watched* by bystanders who have the option of switching to another cleaning station. Research has shown that an observer will copy the behaviour of the previous client, and either invite for inspection if it witnessed a positive interaction, or flee the approaching cleaner if it saw the last client run away as well. The true rationale for cleaner "altruism" is thus a *selfish* one: the opportunity to recruit a new customer and get access to more food (Bshary, Wickler and Fricke, 2002).

Internal representations

The mechanisms by which fish represent their social interactions with other individuals are not known, but fish are certainly able to form internal representations of the status and fighting ability of other individuals in their group, as well as the reliability of former partners (Bshary, Wickler and Fricke, 2002). Presumably, when copying the goal-oriented behaviour of a knowledgeable individual, they must be able to represent the activity of *following the role model's example* as a *means* of attaining its own ends, which are (qualitatively) the same as its own. Alternatively, in simpler cases (e.g finding hidden food by following a knowledgeable individual), the observer may simply represent the model itself as a kind of "moving signpost" pointing to its goal (i.e. the model itself is viewed as a means to the individual's end).

Self-correction

Fish are certainly capable of altering their social behaviour when their expectations of another individual are disappointed. As we saw above, Bshary, Wickler and Fricke (2002) cited evidence that sticklebacks and guppies adopt a *tit-for-tat strategy* towards their partners: a partner that fails to co-operate is punished the next time round. Cleaner fish who cheat their clients by removing extra food (healthy tissue) as well as dead or infected tissue, are "punished" (chased aggressively).

Other evidence for intentional agency in fish

According to Bshary, Wickler and Fricke (2002):

- there is an array of behaviours found in a variety of fish families (categorisation, cheating, punishment, manipulation of individuals and altruism) which are commonly thought of as unique to primates;
- there are instances of interspecific cooperative hunting between giant moray eels and red sea coral groupers;
- co-operative hunting between conspecific predators is widespread in fish, and different

individuals play different roles;

- some fish appear to be able to use *cognitive maps* of their environment. For instance, inter-tidal gobies acquire an effective memory of the topography of its home pool as well as that of surrounding pools, because at low tide, it often has to jump into these pools without being able to see where it is going. Other fish appear to use landmarks for homing;
- some fish use advanced foraging techniques - e.g. removing obstacles to reach hidden prey, and using their spatial intelligence to gain access to prey;
- a few fish are capable of tool-using behaviour in the strict sense of the word (Beck, 1980), where an animal directly handles an object in order to obtain a goal. South American cichlids are a case in point;
- some fish also build complex nests and bowers (Laland, Brown and Krause, 2003).

Finally, contrary to claims by Varner (1998, p. 32), fish are indeed capable of progressive adjustments in *multiple reversal trials* - as long as olfactory stimuli are used (Mackintosh and Cauty, 1971, cited in Wakelin, 2003). In chapter eight, we examined arguments that creatures which show improvements in serial reversal learning were capable of *meta-learning*, insofar as they had to develop primitive hypotheses about changes in their surroundings. We tentatively concluded (Conclusion 8.3) that these improvements constituted good *prima facie* evidence that an animal is trying to adjust to sudden changes in its environment, by rapidly revising its expectations. If this interpretation is correct, then improvements in serial reversal learning should be described using an *agent-centred* intentional stance.

9.7 A special case of combined agency in the jumping spider Portia

Wilcox (2002) has uncovered evidence of sophisticated mimicry and deception among jumping spiders. I shall quote the abstract in full, as the evidence is so impressive:

> Jumping spiders of the Genus Portia are aggressive *mimics* which prey primarily on other species of spiders, especially web-building spiders. The basic tactics Portia exhibits are invasion of a web and signaling deceitfully on it as if Portia were struggling insect prey, to attract the resident spider closer, wherepon Portia kills and eats the resident. In addition to such basic stalking motions on the web, Portia opportunistically uses background noise on the web to mask its stalking motions on the web; uses *trial-and-error learning* to choose an appropriate signal for a particular prey spider, and to determine a different ploy which may deceive the prey spider better; and makes *detours* which involve evaluation of a situation, *planning ahead* to execute a pathway which may take it initially away from and out of sight of the prey spider, and executing the detour, using a *cognitive map, during a period of at least two hours.* The flexible learning, planning ahead, and persistent maintenance of cognitive map abilities of Portia place it among the foremost of invertebrate groups in the cognitive arena (Wilcox, 2002, online, italics mine).

We need not suppose that the spider is engaging in deception proper, with an intention to induce a false belief in another individual's mind, as we established in chapter one that spiders and most other animals clearly lack the neurological wherewithal for phenomenal consciousness. Even so, the behaviour described here surely qualifies as mentalistic. If we examine the key features of *operant conditioning*, most or all seem to be present:

- a *drive* (to prey on other spiders);
- *motor programs*, an action selection mechanism and fine-tuning behaviour (the spider engages in trial-and-error learning or which allows it to fine-tune its body movements on the web, so that they match those of a struggling insect);
- *internal representations* of its current motor output or efferent copy, and its afferent sensory inputs (not directly verified by Wilcox, but there is good reason to suppose this ability is present, given the fine-tuning behaviour described above);
- *associations* between different motor commands and their consequences, which are stored in the animal's memory and updated when circumstances change (the spider "uses trial-and-error learning to choose an appropriate signal for a particular prey spider, and to determine a different ploy which may deceive the prey spider better");
- a *goal* or end-state (getting close enough to the host spider to eat it);
- a *pathway* for getting there (the detour, encoded in the spider's spatial memory);
- *sensory inputs* (the spider can see and feel its way around the web);
- a *temporal correlation mechanism* (without this, the spider would be unable to identify the right signal to attract the host spider); and
- *self-correction* (the spider adjusts its motor behaviour to suit its host).

The fact that the spider appears to use a *spatial map* while navigating its way around the host spider's web, over a two-hour time period, coupled with the fact that it makes *detours*, suggests that it is making use of *internal representations* to plan its attack. This evidence lends further support to a mentalistic interpretation.

The kind of agency displayed here seems to be a combination of operant agency and navigational agency.

I would suggest that we can also speak of the spider as having *beliefs* - e.g. that its prey lies at the end of the path it is following - and *desires* (to eat other spiders).

Conclusion

One of the aims of this thesis was to describe the structure of and conditions for the simplest kinds of minds that could exist in nature, which I referred to as minimal minds. These conditions have been set forth clearly in the last chapter of this thesis, together with case studies that help give the reader some idea of the distribution of each of the four kinds of minimal minds in the animal kingdom. Each of the four forms of agency described (with the exception of social agency) seems to be found among invertebrates, notably insects and cephalopods. A higher form of agency, involving the capacity to abstract rules that are independent of perceptual stimuli, can be found among honeybees. It is an utter mystery how bees accomplish this feat despite lacking phenomenal consciousness.

I have argued that the key reason why we can reasonably impute mental states to these creatures, and describe them as having minimal minds, is that both their internal representations of the outside world (minimal maps) and their patterns of bodily movement robustly instantiate a key feature that was formerly thought to be the hallmark of mental states: *intrinsic* intentionality. They instantiate this feature insofar as they possess a variety of *normative* features, many of which can be accurately described as *conditions of satisfaction.* The normativity of these minimal maps need not be biological; creatures can pursue non-biological goals as well as biological ones, and their pursuit of those goals is subject to conditions of satisfaction. (There are some philosophers, of course, who reject the alleged distinction between intrinsic and derived forms of intentionality, but these philosophers also tend to hold that there is no clear dividing line between entities that have mental states and those that lack them.) I assert that intrinsic intentionality is quite distinct from the derived variety, but that it can be found in creatures lacking phenomenal consciousness. However, I do not claim to have thereby explained away the mystery of phenomenal consciousness; all I have shown is that consciousness is not what accounts for the unity of mental states.

Looking at the ingredients of intentional agency in chapters two to eight, we found that the common thread they shared was their intentionality – in the sense described by Searle (1983, 1999). Conditions of satisfaction are widespread in the biological world – so much so that I was able to put forward a detailed model of a mentalistic representation in chapter nine, which incorporated these conditions. None of the arguments advanced by Searle (1999) for making consciousness, rather than intentionality, the hallmark of mental states proved to be convincing; one of the major conclusions of this thesis was that non-conscious states could indeed represent states of affairs and have conditions of satisfaction, just like conscious ones.

Conditions of satisfaction can thus serve to define the domain of the mental, after all: they are not an artificial category, as some philosophers had proposed. For the time being, the domain of mental states appears to be in safe hands.

References

NOTE: This bibliography is up-to-date as of January 2007 and includes references to ALL works cited and consulted for this thesis. All links are current, unless explicitly stated otherwise.

Aboitiz F., Morales D. and Montiel J. 2000. Online Paper: "The Evolutionary Origin of the Mammalian Isocortex: Insights from Molecular Developmental Biology." Web address: http://cogprints.org/132/00/BBS12.html

Aboitiz F., Morales D. and Montiel J. 2003. "The Evolutionary Origin of the Mammalian Isocortex: towards an integrated developmental and functional approach." In Behavior and Brain Science, October 2003, vol. 26 no. 5, pp.535-552. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=15179935

Abramson, C. I. 1994. A Primer of Invertebrate Learning. Washington, DC: American Psychological Association.

Abramson, C. I. and Buckbee, D. A. 1995. “Pseudoconditioning in earthworms *(Lumbricus terrestris):* Support for nonassociative explanations of classical conditioning phenomena through an olfactory paradigm.” In Journal of Comparative Psychology, 109, pp. 390-393.

Abramson, C., Stone S., Ortez R., Luccardi A., Vann K., Hanig K. and Rice J. 2000. "The Development of an Ethanol Model Using Social Insects I: Behavior Studies of the Honey Bee *(Apis mellifera).*" In Alcoholism: Clinical and Experimental Research, 24(8), 1153-1166. Web address of abstract:

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?itool=abstractplus&db=pubmed&cmd=Retrieve&dopt=abstractplus&list_uids=10968652

Abramson C.I., Kirkpatrick D. E., Bollinger N., Odde R. and Lambert S. 2001. "Planarians in the Psychology Classroom: Habituation and Instrumental Conditioning Demonstrations." In L. Benjamin, B. Nodine, R. Ernst, C. Blair-Broeker (Eds.). Activities handbook for the teaching of psychology, Volume 4. Washington, DC: American Psychological Association. Web address: http://psychology.okstate.edu/faculty/abramson/plan.html

Abramson C. I., Garrido D. J., Lawson A. L., Browne B. L. and Thomas D. G. 2002. "Bioelectrical Potentials of Philodendron cordatum: a New Method for Investigation of Behavior in Plants." In Psychological Reports, August 2002, 91(1), pp. 173 - 185. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=12353777&dopt=Abstract

Abramson, C. 2003. Personal email, 2 February 2003.

Aegerter, M. 1997. "Fundamental Response." Article in "Universe" magazine, Fall 1997. Universe magazine is published by Washington State University. Web address: http://www.wsu.edu/NIS/Universe/funda.htm

Albrecht-Buehler, G. 2003a (last update). Web page: "Cell Intelligence." Web address: http://www.basic.northwestern.edu/g-buehler/cellint0.htm

Albrecht-Buehler, G. 2003b. Web page: "Group migration." Web address: http://www.basic.northwestern.edu/g-buehler/groups.htm

Albrecht-Buehler, G. 2003. Personal email: "Memory in unicellular organisms?" 30 September 2003.

Allen C. 1999. “Animal Concepts Revisited: The Use of Self-Monitoring as an Empirical Approach.” In Erkenntnis 51: 33-40.

Allen, C. 2004. "Animal Pain." In Nous 38:4 (2004), pp. 617-643. Web address: http://grimpeur.tamu.edu/~colin/Papers/animalpain.pdf

Allen, C. 2004. "Transitive inference in animals: Reasoning or conditioned associations?" In S. Hurley and M. Nudds (eds.) Rational Animals? Oxford University Press, forthcoming. Web address: http://grimpeur.tamu.edu/~colin/Papers/transitive%20inference.pdf

Allen, C. 2005. "Animal Consciousness." In Stanford Encyclopedia of Philosophy (Winter 2005 edition). Web address: http://plato.stanford.edu/archives/win2005/entries/consciousness-animal/

Allen, T. 1986. Online article: "A Summary of the principles of Hierarchy Theory." Web address: http://www.isss.org/hierarchy.htm

Allman J., Hakeem A., Erwin J., Nimchinsky E., Hof P. 2001. "The Anterior Cingulate Cortex: The Evolution of an Interface between Emotion and Cognition." In Annals of the New York Academy of Sciences, 2001, 935: 107-117. Web address of abstract: http://www.annalsnyas.org/cgi/content/abstract/935/1/107

Ancel Meyers, L. and Bull, J. 2002. "Fighting change with change: adaptive variation in an uncertain world." In TRENDS in Ecology & Evolution, Vol. 17, No. 12, December 2002. Web address: http://www.santafe.edu/~ancel/PAPERS/TREE.pdf or http://cluster3.biosci.utexas.edu/research/meyers/LaurenM/MeyersTREE02.pdf

Ancel Meyers, L. 2003. Personal email, 18 May 2003.

Anderson A., Holmes B. and Else L. 1996. "Zombies, dolphins and blindsight" (summary of 2nd Tucson Conference on Consciousness). In New Scientist, 4 May, 1996, pp. 20-27. Web address: http://www.fortunecity.com/emachines/e11/86/zombie.html

Anderson R. and Wood J. 2001. "Enrichment for Giant Pacific Octopuses: Happy as a Clam?" In Journal of Applied Animal Welfare Science, vol. 4(2), pp. 157-168. Web address: http://www.cephbase.dal.ca/refdb/pdf/6405.pdf

Anscombe, G. E. M. 1957. Intentions. Cambridge University Press.

Anscombe, G.E.M.. 1965. “The intentionality of sensation: a grammatical feature.” In R.J. Butler (ed.), Analytical Philosophy: First Series. Oxford: Blackwell. Reprinted in No, A. and Thompson, E. (eds.) 2002. Vision and Mind, Selected Readings in the Philosophy of Perception. Cambridge, Mass.: MIT Press.

Aristotle. 1959. Politics. London: Everyman. Translation by J. Warrington.

Aristotle. 1986. De Anima. London: Penguin Books. Translation by Hugh Lawson-Tancred.

Aristotle. 2001. Nicomachean Ethics. Ed. J. Bywater. Oxford, 1894. Translated by William David Ross, Clarendon Press 1908. Web address: http://www.mikrosapoplous.gr/aristotle/nicom0.htm

ArtificialBrains.Com (now the SETI Institute). 2002. The Honeybee Brain. Web address: http://www.setiai.com/archives/cat_honey_bee_brain.html

Axelrod R., Hamilton W. D. 1981. “On the evolution of cooperation.” In Science 211:1390–1396. Cited in Bshary, Wickler and Fricke, 2002. Web address of abstract: http://www.sciencemag.org/cgi/content/abstract/211/4489/1390

Aydede, M. 2005. "Pain." The Stanford Encyclopedia of Philosophy (Winter 2005 Edition), Edward N. Zalta (ed.). Web address: http://plato.stanford.edu/archives/win2005/entries/pain/

Baars, B. 2001. "There are no known differences in fundamental brain structures between humans and other animals." In Animal Welfare, 2000. Web address: http://cogprints.ecs.soton.ac.uk/archive/00000912/00/UFAW_PAPER_final_version_July_12_2000.txt

Baars, B. 2003a. "Unconscious states cast light on consciousness: Evidence from persistent vegetative states (PVS)." In Science and Consciousness Review, 2003, January, No. 3. Web address: http://www.sci-con.org/2003/01/unconscious-states-cast-light-on-consciousness/

Baars, B. 2003b. "The Global Brainweb - an update on global workspace theory." In Science and Consciousness Review, 2003, October, No. 2. Web address: http://cogweb.ucla.edu/CogSci/Baars-update_03.html

Backer, A. 1999. Lecture: "Evolution of Sleep." From a course on Sleep and Dreams, Bi/CNS 163, conducted at Caltech from 1998-2000. Web address: http://www.its.caltech.edu/~alex/Evolution%20of%20Sleep%20Lecture.htm

Baker G. and Morris K. 1996. Descartes' Dualism. London: Routledge.

Baker G. and Morris K. 1997. Reply to Nadler's review ("Descartes' Dualism?") In Philosophical Books, 38, 3, July 1997, pp. 157-169. Web address: http://www.newdualism.org/papers/S.Nadler-et-al/Baker-Morris-review.htm

Balkenius, C. 2000. "Attention, Habituation and Conditioning: A Computational Model." In Cognitive Science Quarterly, 2000. Vol. 1 (2), pp. 171-214. Web address of draft: http://www.lucs.lu.se/People/Christian.Balkenius/PDF/Balkenius.2000.CSQ.pdf

Ball, J. 1998. "The Question of Non-Human Intelligence." Web address: http://serendip.brynmawr.edu/bb/neuro/neuro98/202s98-paper2/Ball2.html

Balda R., Pepperberg, I. and Kamil A. (eds.) 1998. Animal Cognition in Nature: The Convergence of Psychology and Biology in Laboratory and Field. Academic Press: New York.

Barkay T. and Wagner-Dobler I. 2005. “Microbial Transformations of Mercury: Potentials, Challenges, and Achievements in Controlling Mercury Toxicity in the Environment.” In Advances in Applied Microbiology, Volume 57, pp. 1-52. Web address: http://aesop.rutgers.edu/~barkay/barkay_wagner-dobler.pdf

Bateson, P. 1991. “Assessment of pain in animals.” In Animal Behaviour, 42(5), pp. 827-839.

Bavidge, M. & Ground, I. 1994. Can We Understand Animal Minds? London: Bristol Classical Press.

BBC News report, 30 April 2003. "Fish do feel pain, scientists say." Web address: http://news.bbc.co.uk/2/hi/science/nature/2983045.stm

BBC News report, 31 August 2003. "Scientists highlight fish 'intelligence'." Web address: http://news.bbc.co.uk/go/pr/fr/-/1/hi/england/west_yorkshire/3189941.stm

BBC News report, 26 January 2004, 15:27 GMT. "Parrot's oratory stuns scientists." Web address: http://news.bbc.co.uk/2/hi/science/nature/3430481.stm

BBC News report, 6 October 2005, 00:49 GMT. “New breed of ‘fish-bot’ unveiled.” Web address: http://news.bbc.co.uk/2/hi/science/nature/4313266.stm

Beck, B. B. 1980. Animal Tool Behavior: The Use and Manufacture of Tools. Garland Press: New York, NY.

Bedau, M. "The Nature of Life." 1996. In Margaret Boden, ed., The Philosophy of Artificial Life (Oxford University Press), pp. 332-357. Web address of article: http://www.reed.edu/~mab/papers/life.OXFORD.html

Bedau, M. 1998. "Four Puzzles about Life." In Artificial Life vol. 4 (1998), pp. 125-140. Web address of article: http://www.reed.edu/~mab/papers/4.puzzles.htm

Behavioural Ecology Research Group. 2002. Web page: General Cognition. Web address: http://users.ox.ac.uk/~kgroup/tools/tool_selection.shtml

Beisecker, D. 1999. "The Importance of Being Erroneous: Prospects for Animal Intentionality." In Philosophical Topics, vol. 27:1, Spring 1999, pp. 281-308. Web address: http://www.scsv.nevada.edu/~beiseckd/Erroneous.html

Beisecker, D. 2000. The Meaning of Life and More: A Constructive Account of Non-Biological Intentionality. Talk presented at the June 2000 meeting of the Society for Philosophy and Biology, New York, N.Y. Web address: http://www.scsv.nevada.edu/~beiseckd/LifeandMore.html

Bekoff M., Allen C., and Burghardt G. (eds). 2002. The Cognitive Animal. Cambridge, Massachusetts: MIT. Paperback edition.

Ben-Shahar Y., Thompson C., Hartz S., Smith B. and Robinson G. 2000. "Differences in performance on a reversal learning test and division of labor in honey bee colonies." In Animal Cognition Vol. 3: 119-125. Web addresss of abstract: http://www.springerlink.com/content/nwt9n452yc576jlu/
Web address of full article: http://www.life.uiuc.edu/robinson/research/AnCog3.pdf

Benjamini, L. 1983. "Studies in the learning abilities of brown-necked ravens and herring gulls." I. Oddity learning. In Behaviour, 84, 173-194. Cited in Young and Wasserman, 2001 (see below).

Bennett, A. 1996. “Do Animals Have Cognitive Maps?” In The Journal of Experimental Biology 199:11, pp. 219-224. Web address of abstract: http://jeb.biologists.org/cgi/content/abstract/199/1/219
Web address of full article: http://jeb.biologists.org/cgi/reprint/199/1/219.pdf

Bermudez, J. 2000. Consciousness, Higher-order thought and Stimulus Reinforcement: Comments on Rolls' The Brain and Emotion. In Behavioral and Brain Sciences 23, 2000, pp. 194-195. Web address: http://www.artsci.wustl.edu/~pnp/Papers/bermudez2000b.pdf

Berridge, K. 2003a. "Irrational Pursuit: Hyper-Incentives from a Visceral Brain." In Brocas I. and Carrillo J. (eds.) The Psychology of Economic Decisions. Volume 1: Rationality and Well-being, 2003, pp. 17-40. Oxford University Press. Web address:

http://www.lsa.umich.edu/psych/research&labs/berridge/publications/Berridge%202003%20Irrational%20pursuit.pdf

Berridge, K. 2003b. "Pleasure, Unconscious Affect and Irrational Desire." In Manstead A., Frijda N. & Fisher A. (eds.), Feelings and Emotions: The Amsterdam Symposium, 2004, pp. 43-62. Cambridge University Press. Web address: http://www.lsa.umich.edu/psych/research&labs/berridge/publications/Berridge%20chapter%20Feelings%20&%20Emotions%20book%202004.pdf

Berridge K., Winkielman, P. 2003. "What is an unconscious emotion? (The case for unconscious 'liking'.)" In Cognition and Emotion, 2003, 17(2), pp. 181-211. Web address: http://psy.ucsd.edu/~pwinkiel/Berridge%20&%20Winkielman-unconscious%20emotion-CE-2003.pdf

Berridge, K. 2004. “Motivation concepts in behavioral neuroscience.” In Physiology and Behavior, 81, pp. 179-209. Web address: http://www.lsa.umich.edu/psych/research&labs/berridge/publications/Berridge%20Motivation%20concepts%20Physio%20&%20Beh%202004.pdf

Beugnon, G. 1999. "How ants find their way home." Press release from CNRS-Universite Paul Sabatier, 25 June 1999. Web address: http://www.cnrs.fr/cw/en/pres/compress/ants250699.html

BioMedia Associates. 2003. Web pages: "The Domains of Life" (Web address: http://ebiomedia.com/prod/DOL.html) and "Five Kingdoms and Domains of Life Compared" (Web address: http://ebiomedia.com/downloads/K_D_Table.pdf).

Birch, C. 2001. "Why I Became a Panexperientialist". Paper given at the Australian Association for Process Thought - 2001 Conference, 29 November - 1 December 2001, at Latrobe University, Melbourne. Web address: http://www.nerdshit.com/archive/2004/12/23/why_i_became_a_/.

Bitterman, M. 1965. "The evolution of intelligence." In Scientific American, vol. 212, pp. 92-100.

Bittner, R. 2001. Doing Things for Reasons. Oxford University Press.

Blair, D. F. 1995. "How Bacteria Sense and Swim." In Annual Review of Microbiology, volume 49, pp. 489-522. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?holding=npg&cmd=Retrieve&db=PubMed&list_uids=8561469&dopt=Abstract

Block, N. 1995. "On a Confusion about a Function of Consciousness". In Behavioral and Brain Sciences 18 (2): 227-287. Re-printed in The Nature of Consciousness, 1997, edited by Block. N., Flanagan O. and Guzeldere G. MIT Press. Web address of unedited pre-print (not a quotable final draft): http://www.bbsonline.org/documents/a/00/00/04/31/bbs00000431-00/bbs.block.html

Block, N. 1998. "How Not To Find The Neural Correlate of Consciousness". In a volume of Royal Institute of Philosophy lectures, ed. Anthony O'Hear, pp. 1-11. Web address: http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/NeuralCorrelate.html

Block, N. 2001. "Paradox and Cross Purposes in Recent Work on Consciousness." In Cognition, Volume 79, Issues 1-2, pp. 197-219 (April 2001). Web address of updated and revised version: http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/Elsevier.html

Block, N. 2003. Article: "Consciousness." In Encyclopedia of Cognitive Science, edited by Lynn Nadel. New York, NY, Nature Publishing Group, 2003. Web address: http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/ecs.pdf

Block, N. 2005. “Two Neural Correlates of Consciousness.” In Trends in Cognitive Sciences, Vol. 9, No. 2, pp. 46-52. Web address: http://www.psy.ku.dk/cvc/Meetings/Blo2005.pdf

Borthwick, C., 1996. "The permanent vegetative state; ethical crux, medical fiction?" In Issues in Law and Medicine, vol. 12, no. 2, pp. 167-188. Web address: http://home.vicnet.net.au/%7Eborth/PVSILM.HTM

Brainard M. and Knudsen E. 1998. Journal of Neuroscience, 1998, May 15; 18(10): 3929-42. “Sensitive periods for visual calibration of the auditory space map of the barn owl optic tectum.” Web address: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=9570820&dopt=Abstract

Brannon, E. and Terrace, H. 2000. “Representation of the Numerosities 1-9 by Rhesus Macaques.” In Journal of Experimental Psychology: Animal Behavior Processes. Volume 26, no. 1. January 2000. News report: Azar, B. 2000. "Monkeying around with numbers." In Monitor on

Psychology, Volume 31, no. 1. January 2000. (Published by the APA.) Web page: http://www.apa.org/monitor/jan00/sc4.html

Brentano, F. 1874/1911/1973. Psychology from an Empirical Standpoint. London: Routledge and Kegan Paul.

Brembs, B. 1996. Diplomaarbeit. "Operant and Classical Conditioning of Drosophila at the Flight Simulator." Faculty of Biology, Julius Maximilians University of Wuerzburg, Germany.Web address: http://www.brembs.net/papers/diploma.pdf

Brembs, B. 2000. Doctoral dissertation: An Analysis of Associative Learning in Drosophila at the Flight Simulator. Julius Maximilians University of Wuerzburg, Germany. Web address: http://www.brembs.net/papers/dissertation.pdf

Brembs, B. and Heisenberg, M. 2001. "Conditioning with Compound Stimuli in Drosophila melanogaster in the Flight Simulator." In The Journal of Experimental Biology, volume 204, pp. 2849-2859 (2001). Web address: http://jeb.biologists.org/cgi/content/full/204/16/2849

Brembs, B. Personal email communication, 22 December 2002.

Brembs, B. Personal email communication, 11 August 2003.

Brembs, B. 2003. "Thrust Can Operate Angular Motion." Web page attached to the Department of Neurophysiology and Genetic Behavior, University of Wuerzburg, Germany. Web address: http://www.biozentrum.uni-wuerzburg.de/genetics/behavior/learning/thrust.html

Bridges, M. 2000. Biology 2430 Lecture Notes. Chapter 14: Regulation of Gene Expression. Web address: http://www.langara.bc.ca/biology/mario/Biol2430notes/biol2430chap14.html

Brittan, G. G. (Jr.) 1999. “The Secrets of Antelope.” In Erkenntnis 51: 59-77.

Brooks, R. A. 1986. "A robust layered control system for a mobile robot." In IEEE Journal on Robotics and Automation, Vol. 2, No. 1, March 1986, pp. 14–23; also MIT AI Memo 864, September 1985. Web address of abstract: http://portal.acm.org/citation.cfm?id=888753&dl=acm&coll=&CFID=15151515&CFTOKEN=6184618
Web address of full article: http://people.csail.mit.edu/brooks/papers/AIM-864.pdf

Brosnan S. and de Waal F. 2003. "Monkeys reject unequal pay." In Nature, vol. 425, 18 September 2003, pp. 297-299. Web address: http://www.emory.edu/LIVING_LINKS/capuchins/BrosnanNature/Manuscripts/BrosnanUnequalPay.pdf

Brown C. and Laland K. N. 2003. "Social learning in fishes: a review." In Fish and Fisheries, 2003, vol. 4, pp. 280-288. Web address: http://www.geocities.com/culumbrown/BrownLalandFAF2003

Brown C., Markula A. and Laland, K. 2003. "Social learning of prey location in hatchery-reared Atlantic salmon." In Journal of Fish Biology (2003), 63, 738-745. Web address: http://www.geocities.com/culumbrown/BrownMarkulaLaland2003

Brown C. 2004. "Not just a pretty face." In New Scientist, 12 June 2004, pp. 42-43.

Bryson J. and Hauser M. 2002. “What Monkeys See and Don’t Do: Agent Models of Safe Learning in Primates.” In Barley, M. and Guesgen, H. W., (Eds.), AAAI Spring Symposium on Safe Learning Agents." Web address: http://www.cs.bath.ac.uk/~jjb/ftp/bryson-safe02.pdf

Bshary R., Wickler W. and Fricke H. 2002. "Fish cognition: a primate's eye view." In Animal Cognition, vol. 5, March 2002, pp. 1-13. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=11957395&dopt=Abstract

Web address of full article:

http://www2.unine.ch/webdav/site/ethol/shared/documents/bswfr02.pdf

Budiansky, S. 1998. If a Lion Could Talk. New York: The Free Press.

Buller, D. J. (ed.) 1999a. Function, Selection, and Design. Albany: State University of New York Press. Cited in Cameron, 2000.

Buller, D. J. 1999b. Introduction: Natural Teleology. In Function, Selection, and Design, edited by D. J. Buller. Albany: State University of New York Press. Cited in Cameron, 2000.

Bullock, T. H. 2003. "Have brain dynamics evolved? - Should we look for unique dynamics in the sapient species?" Current version of paper: 17 February 2003. Based on a paper for the Brain Dynamics Workshop, May 10-12, 2002, Rancho Santa Fe, CA. "From Microscopic to Macroscopic Brain Dynamics" organized by T. Sejnowski. Web address: http://cogprints.ecs.soton.ac.uk/archive/00002777/01/Evol_brain_dynamics_2-14.htm

Bunsey, M. and Eichenbaum H. 1996. "Conservation of hippocampal memory function in rats and humans." In Nature, January 18 1996, vol. 379, pp. 255-257. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?holding=npg&cmd=Retrieve&db=PubMed&list_uids=8538790&dopt=Abstract

Butler A., Manger P., Lindahl B. and Arhem P. 2005. "Evolution of the neural basis of consciousness: a bird-mammal comparison." In Bioessays (September 2005), 27(9): 923-36. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=16108067&dopt=Abstract

Butler A. and Cotterill R. 2006. "Mammalian and Avian Neuroanatomy and the Question of Consciousness in Birds." In The Biological Bulletin 211: 106-127. (October 2006). Web address of abstract: http://www.biolbull.org/cgi/content/abstract/211/2/106, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=17062871.

Byrne, A. 2006. “Intentionality.” In *Philosophy of Science: An Encyclopedia*, ed. J. Pfeifer and S. Sarkar (Routledge, forthcoming).

Cabanac, M. 1996. "On the Origin of Consciousness, a Postulate and its Corollary." In Neuroscience and Biobehavioral Reviews, Vol. 20, Issue 1, Spring 1996, pp. 33-40. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=8622827&dopt=Abstract

Cabanac, M. 1999. "Emotion and Phylogeny." In Japanese Journal of Physiology, 1999 Feb., 49(1):1-10. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=10219103&dopt=Abstract

Cabanac, M. 2002. "What is emotion?" In Behavioral Processes, Vol. 60, No. 2, November-December 2002, pp. 69-83. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=12426062&dopt=Abstract

Cabanac, M. 2003. "Pleasure of Taste and Phylogeny." Presentation given at Columbia University Seminar on Appetitive Behavior (#529) on Thursday, May 8, 2003. Web address: http://www.nyorc.org/may_2003.html

Cameron, R. 2000. Unpublished Ph.D. dissertation: "Teleology in Aristotle and Contemporary Philosophy of Biology: An Account of the Nature of Life." University of Colorado, Boulder. Formerly available online at http://www.colorado.edu/philosophy/fac/cameron/diss/diss.pdf and now available on the Internet Archive at

http://web.archive.org/web/20010101-20050101re_/http://www.colorado.edu/philosophy/fac/cameron/diss/diss.pdf

Cameron, R. 2004. "How to be a Realist about *sui generis* Teleology Yet Feel at Home in the 21st century." In The Monist 87(1), January 2004, pp. 72-95.

Campellone, J. 2002. Article: "*Petit mal* seizure". In Medline Plus Medical Encyclopedia, a service of the U.S. National Library of Medicine and the National Institutes of Health. Web address: http://www.nlm.nih.gov/medlineplus/ency/article/000696.htm

Carruthers, P. 1992. The Animals Issue. Cambridge University Press.

Carruthers, P. 1998. "Animal Subjectivity." In Psyche, 4(3), April 1998. Web address: http://psyche.cs.monash.edu.au/v4/psyche-4-03-carruthers.html

Carruthers, P. 1999. "Sympathy and Subjectivity." In Australasian Journal of Philosophy, vol. 77 (1999), pp. 465-482. Web address: http://cogprints.org/1204/00/Sympathy.htm

Carruthers, P. 2000a. "Replies to Critics: Explaining Subjectivity." In Psyche, 6(3), February 2000. Web address: http://psyche.cs.monash.edu.au/v6/psyche-6-03-carruthers.html

Carruthers, P. 2000b. Phenomenal Consciousness. Cambridge University Press.

Carruthers, P. 2000c. The evolution of consciousness. In Carruthers P. and Chamberlain A. (eds.), Evolution and the Human Mind. Cambridge: Cambridge University Press, pp. 254-275.

Carruthers, P. 2001. Article: "Consciousness, Higher-Order Theories of." In Stanford Encyclopedia of Philosophy. Web address of current article: http://plato.stanford.edu/entries/consciousness-higher/ . Web address of original article: http://plato.stanford.edu/archives/sum2001/entries/consciousness-higher/

Carruthers, P. 2004a. "On being simple minded". In American Philosophical Quarterly, 2004, 41(33):205-220. Also in Carruthers, P. 2005. Consciousness: Essays from a Higher-Order Perspective. OUP. (Chapter 12, pp. 215-248). Web address: http://www.philosophy.umd.edu/Faculty/pcarruthers/Simple-minds.pdf
Note: the online PDF version is cited in this thesis.

Carruthers, P. 2004b. "Why the question of animal consciousness might not matter very much." In Philosophical Psychology 17 (2004). Web address: http://www.philosophy.umd.edu/Faculty/pcarruthers/animal-consciousness-might-not-matter.pdf

Cartmill, M. 2000. "Do horses gallop in their sleep?" In Key Reporter 66.1 (autumn 2000), pp. 6-9. Web address: http://www.comet.columbia.edu/~shane/words/animalconsciousness.htm

Catherall, D. 2003. "How Fear Differs From Anxiety." In Traumatology, Vol. 9, No. 2, June 2003. Web address: http://www.fsu.edu/~trauma/V9/V9i2_Fear-Anxiety.PDF

Chalmers, D. 1996. The Conscious Mind: In Search of a Fundamental Theory. New York: Oxford University Press.

Chalmers, D. 1998. "What is a Neural Correlate of Consciousness?" Published in Neural Correlates of Consciousness: Empirical and Conceptual Questions (T. Metzinger, ed), published with MIT Press in 2000. First presented at Association for the Scientific Study of Consciousness conference on "Neural Correlates of Consciousness" in Bremen, June 1998. Web address: http://consc.net/papers/ncc2.html

Chandra S., Hosler J. and Smith B. 2000. "Heritable variation for latent inhibition and its correlation to reversal learning in the honeybee, Apis mellifera". In The Journal of Comparative Psychology, 114:86-97.
Web address:
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?holding=npg&cmd=Retrieve&db=PubMed&list_uids=10739314&dopt=Abstract

Chappell J. and Kacelnik A. 2004. "Selection of Tool Diameter by New Caledonian Crows *Corvus moneduloides*." In Animal Cognition 7:121-127. Web address: http://users.ox.ac.uk/~kgroup/diameter_select.pdf

Cheney D. and Seyfarth R. 1990. How Monkeys See the World: Inside the Mind of Another Species. Chicago: University of Chicago Press.

Chisholm, R. M. 1957. Perceiving. Ithaca: Cornell.

Chittka, L. 2004. “Dances as Windows into Human Perception.” PLoS Biol. 2004, 2(7)e216. (Volume 2, issue 7, pp. 0898-0900.) Web address of abstract: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=449901

Web address of full article:

http://www.biology.qmul.ac.uk/research/staff/chittka/2004/chittka_plos04.pdf

Choi, C. 2001, 2002. Article: "Faces only wasps could love." Science Now, 12 July 2002. Copyright American Association for the Advancement of Science, 2001. Web address: http://bric.postech.ac.kr/science/97now/02_7now/020712c.html

Chudler, E. 2003. "Brain Facts and Figures." Web address: http://faculty.washington.edu/chudler/facts.html#brain

Chui, G. 1999. "Tree of life proposal divides scientists". In San Jose Mercury, Monday, September 20, 1999. Web address: http://ucjeps.berkeley.edu/DeepGreen/SJMercury.html

Church, J. 2000. "Understanding Pain and its Relevance to Animals". Livestock industry research paper prepared for Alberta Farm Animal Care. Web address: http://www.afac.ab.ca/Reports/UnderstandingPain.pdf

Churchland, P.M. 1981. “Eliminative Materialism and the Propositional Attitudes.” In Journal of Philosophy 78: 67-90.

Churchland, P.S. 1986. Neurophilosophy: Toward a Unified Science of the Mind/Brain. Cambridge, MA: MIT Press.

Clark, S. 1988. "Is Humanity a Natural Kind?" In Ingold, T. (ed), What is an Animal? London: Unwin and Hyman.

Clayton N., Bussey T., Emery N., Dickinson A. 2003. "Prometheus to Proust: the case for behavioural criteria for 'mental time travel'." In Trends in Cognitive Science, October 2003, 7(10):436-437. Web address: http://www.zoo.cam.ac.uk/zoostaff/madingley/library/member_papers/nemery/from_prometheus_to_proust.pdf

Coghlan, A. 1998. "Sensitive Flower." In New Scientist, 2153 (26 September 1998), pp. 24-28.

Collett M., Collett T., Bisch S. and Wehner R. 1998. "Local and global vectors in desert ant navigation." In Nature, vol. 394, 16 July 1998, pp. 269-272. Web address: http://www.cogs.susx.ac.uk/users/paulgr/Nat%20Collett%20Collett.pdf

Collett, M. and Collett, T. 2000. "How do insects use path integration for their navigation?" In Biology and Cybernetics, September 2000, vol. 83(3): 245-59. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?holding=npg&cmd=Retrieve&db=PubMed&list_uids=11007299&dopt=Abstract

Collett, T. and Collett, M. 2002. "Memory use in insect visual navigation." Nature Reviews: Neuroscience, 3: 542-552. Web address: http://www.cogs.susx.ac.uk/users/paulgr/NatNeuroSci%20Collett%20Collett.pdf

Collins, A. and Waggoner, B. 1994, 2001. Web page "Metazoa: Systematics" on the University of California Museum of Paleontology Web site. Web address: http://www.ucmp.berkeley.edu/phyla/metazoasy.html

Cook, R. 2002. "Same-Different Concept Formation in Pigeons." In Bekoff M., Allen C. and Burghardt G. (eds). The Cognitive Animal. Cambridge, Massachusetts: MIT. Paperback edition, pages 229-238. Web address: http://grimpeur.tamu.edu/~colin/TCA/Ch/Cook/cook-final.pdf

Corry, R. 2003. "Insect Navigation - Evidence for symbol systems approach (Gallistel)." Lecture Notes for HPSC2610 (Computers, Brains and Minds). Lecture 15. **Former** Web address: http://www.phil.mq.edu.au/staff/rcorry/hpsc2610/lecture15.pdf. Not available on the Internet archive.

Cosmides L. and Tooby J. 1997. Online article: "Evolutionary Psychology: A Primer." Web address: http://www.psych.ucsb.edu/research/cep/primer.html

Cotterill, R. 1997. "Navigation, Consciousness and the Body/Mind 'Problem'." In Psyke and Logos, 1997, vol. 18, pp. 337-341. Web address of abstract: http://dcwww.fys.dtu.dk/~rodney/publications/1997cabstract.html
Web address of full article: http://dcwww.fys.dtu.dk/~rodney/publications/1997c.pdf

Cotterill, R. 2001. "Co-operation of the basal ganglia, cerebellum, sensory cerebrum and hippocampus: possible implications for cognition, consciousness, intelligence and creativity." In Progress in Neurobiology, 64, (2001), 1-33. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=11250060&dopt=Abstract Web address of full article: http://dcwww.fys.dtu.dk/~rodney/publications/pn_cotterill2001.pdf

Cotterill, R. 2002. "Mind - A Moving Story." In Science and Consciousness Review 2002, August No. 2.

Crawford R. 2000. "A Reference Source for Analgesia & Analgesics in Animals." Web address: http://www.nal.usda.gov/awic/pubs/awic200002.htm#fish

Crick F., Koch, C. 1998. "Consciousness and Neuroscience." In Cerebral Cortex, 8:97-107, 1998. Web address: http://www.klab.caltech.edu/~koch/crick-koch-cc-97.html

Crick F., Koch C. 2003. "A Framework for Consciousness." In Nature Neuroscience, Vol. 6, No. 2, February 2003, pp. 119-126. Web address: http://www.klab.caltech.edu/~koch/crick-koch-03.pdf

Crist, E. 2002. "The inner life of earthworms: Darwin's argument and its implications." In Bekoff M., Allen C. and Burghardt G. (eds). The Cognitive Animal. Cambridge, Massachusetts: MIT. Paperback edition. Web address: http://grimpeur.tamu.edu/~colin/TCA/Ch/Crist/crist-final.pdf

Cruchet R. 1905. Tics et sommeil. Presse Med. 1905; 13:33-36.

Davidson, D. 1975. "Thought and talk." In S. Guttenplan (ed.), Mind and Language. Oxford University Press. Reprinted in Davidson D. 1984. Inquiries into Truth and Interpretation. Clarendon Press, Oxford, pp. 155-170.

Davidson, D. 1985. “Rational Animals.” In E. LePore and B. McLaughlin (eds.) Actions and Events. Blackwell, Oxford, pp. 473-480.

Davidson, D. 1999. “The Emergence of Thought.” In Erkenntnis 51: 7-17.

Davidson, K. 2001. "Bees brighter than we knew, study finds - They pass cognitive tests usually given apes, people." In San Francisco Chronicle, Thursday April 19, 2001. Web address: http://www.cyberbee.net/news/beebright.html

Davidson, M. and the Florida State University Web site. Molecular Expressions (Trade Mark) - Cell Biology and Microscopy. Structure and Function of Cells and Viruses. Web page: Virus Structure. Last modified: January 9, 2005. Copyright 1995 - 2005. Web address: http://micro.magnet.fsu.edu/cells/virus.html

Davies, P. 1999. The Fifth Miracle: The Search for the Origin and Meaning of Life. New York: Simon and Schuster.

Davison, S. 1999. "Why Machines Deserve Respect, Too." Paper given at the Notre Dame Center for Ethics and Culture 2003 Formation and Renewal Conference. Web address: http://ethicscenter.nd.edu/archives/documents/ScottDavison.doc

Dawkins, M. 1994. Through Our Eyes Only? The Search for Animal Consciousness. W. H. Freeman / Spektrum, Oxford.

Dawkins, R. 1976. The Selfish Gene. Oxford: Oxford University Press.

Dawkins, R. 1986. The Blind Watchmaker. New York: W. W. Norton and Company.

De Bono M., Maricq A.V. 2005. “Neuronal substrates of complex behaviors in C. elegans.” Annual Review of Neuroscience 28:451-501. Web address: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Abstract&list_uids=16022603

DeGrazia, D. 2002. Online Article: "Philosophical and Ethical Issues Concerning Pain". Copyright: the John Hopkins Center for Alternatives to Animal Testing. Web address: http://altweb.jhsph.edu/meetings/pain/degrazia.htm

De Sousa, R. 2003. Article: "Emotion". In The Stanford Encyclopedia of Philosophy. Web address: http://plato.stanford.edu/entries/emotion/ . Web address of original article: http://plato.stanford.edu/archives/spr2003/entries/emotion/

Dennett, D. 1995a. “Do Animals Have Beliefs?” In Roitblat, H., ed., Comparative Approaches to Cognitive Sciences, MIT Press, 1995. Web address: http://ase.tufts.edu/cogstud/papers/doanimal.htm

Dennett, D. 1995b. "Animal consciousness: what matters and why." In Social Research, Fall 1995, vol. 62 no. 3 p. 691. Web address: http://instruct.westvalley.edu/lafave/dennett_anim_csness.html

Dennett, D. 1997. Kinds of Minds. London: Phoenix. Paperback edition.

Denton D., McKinley M. and Weisinger R. 1996. "Hypothalamic integration of body fluid regulation." In Proceedings of the National Academy of Sciences, Vol. 93, Issue 14, pp. 7397-7404, July 9, 1996. Web address of abstract: http://www.pnas.org/cgi/content/abstract/93/14/7397?ijkey=58d694a4b4a371d5848df5770b4d85b46b8506ee&keytype2=tf_ipsecsha

Denton D., Shade R., Zamarippa F., Egan G., Blair-West J., McKinley M., Lancaster J., and Fox P. 1999. "Neuroimaging of genesis and satiation of thirst and an interoceptor-driven theory of origins of primary consciousness." In Proceedings of the National Academy of Sciences, Vol. 96, Issue 9, pp. 5304-5309, April 27, 1999. Web address of abstract: http://www.pnas.org/cgi/content/full/96/9/5304

Descartes, R. 1968. Discourse on Method and the Meditations. London: Penguin.

Dickinson, A. 1980. Contemporary Animal Learning Theory. Cambridge, England: Cambridge University Press.

Di Primio F., Muller B. S. and Lengeler J. W. "Minimal Cognition in Unicellular Organisms." 2000. In: J.-A. Meyer, A. Berthoz, D. Floreano, H. L. Roitblat and S. W. Wilson (eds). SAB (Simulation of Adaptive Behavior) 2000 Proceedings Supplement, International Society for Adaptive Behavior, Honolulu, 2000, pp. 3-12. Web address of abstract: http://www.ais.fraunhofer.de/~diprimio/publications/diprimio_MinCog-abstract.html . Web address of full article: http://www.ais.fraunhofer.de/~diprimio/publications/diprimio_MinCog.pdf

Digital Life Lab at CalTech, Web address: http://dllab.caltech.edu/avida/ cited in http://www.alife.org and posted on Tuesday, October 1, 2002.

Discover magazine. 1998. "Platypus dreams - REM sleep discovered in ancient platypus - Breakthroughs - Brief article." March 1998 issue. Web address: http://articles.findarticles.com/p/articles/mi_m1511/is_n3_v19/ai_20324733

Dolan, R. 2000. "Emotional processing in the human brain revealed through functional neuroimaging." In The New Cognitive Neurosciences (Gazzaniga M., ed.), pp. 1115-1132. Cambridge, Massachusetts: MIT Press. Cited in Rose, 2002a.

Doolittle, W. F. 2000. "Phylogenetic Classification and the Universal Tree". In Science 284:2124-9. Web address: http://cas.bellarmine.edu/tietjen/Ecology/phylogenetic_classification_and_.htm

Doolittle, W. F. 2002. Web page: Life on Earth, accessed from The Tree of Life home page. Web address: http://doolittle.biochem.dal.ca/index.html.

Dreckmann, F. 1999. “Explaining Behaviour.” In Erkenntnis 51: 93-111.

Dretske, F. 1980. “The intentionality of cognitive states.” In Rosenthal, D. (ed.)(1990) The Nature of Mind. Oxford: Oxford University Press.

Dretske, F. 1981. Knowledge and the Flow of Information. Cambridge, Mass.: MIT Press.

Dretske, F. 1988. Explaining behavior. Cambridge, Massachusetts: MIT.

Dretske, F. 1993. “Misrepresentation.” In Readings in Philosophy and Cognitive Science, ed. A. Goldman. (Cambridge: MIT Press, 1993, pp. 297-313.

Dretske, F. 1995. "What good is consciousness?" Talk given to the Center for the Study of Language and Information, Stanford University, 2 November 1995. Web address: http://www.ecs.soton.ac.uk/~harnad/Papers/Py104/dretske.good.html

Dretske, F. 1999. "Machines, Plants and Animals: The Origins of Agency." In Erkenntnis, 51:19-31.

Dretske, F. 2001. "First Person Warrant." In Psyche 7(11), April 2001. Web address: http://psyche.cs.monash.edu.au/v7/psyche-7-11-dretske.html

Duane M., Fortenberry G. and Horowitz A. "Inquiry in Science Using an Animal Behavior Model." 1998 Woodrow Wilson Leadership Program for Teachers (WWLPT) Biology Institute: Life in Motion. Web address: http://www.woodrow.org/teachers/bi/1998/planaria/ (contains references to scientific articles) and http://www.woodrow.org/teachers/bi/1998/planaria/backgroundpage.htm (describes experiments on planaria).

Dubnau, J. Personal email communication, 14 August 2003.

Dugatkin L. A. 1988. “Do guppies play Tit for Tat during predator inspection visits?” In Behavioral Ecology and Sociobiology. 23:395–399. Cited in Bshary, Wickler and Fricke, 2002. Web address of abstract: http://www.springerlink.com/content/h6131654251r7138/

Dugatkin L. A. 1997. Cooperation among animals. Oxford University Press: New York. Cited in Bshary, Wickler and Fricke, 2002.

Dusek J., Eichenbaum H. 1997. "The hippocampus and memory for orderly stimulus relations." In Proc Natl Academy of Sciences, USA., 24 June 1997; vol.94 no.13, pp. 7109-7114. Web address: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9192700

Eaton M. 2003. "Common animal models for spasticity and pain." In Journal of Rehabilitation Research and Development, Vol. 40, No. 4, July/August 2003, Supplement 1, pp. 41-54. Web address: http://www.vard.org/jour/03/40/4Sup1/eaton.html

Ebert, T. 1983. "Aristotle on what is done in perceiving." In Zeitschrift fur Philosophische Forschung, 37, 181-198.

Edelman D., Baars B. and Seth A. 2005. "Identifying Hallmarks of Consciousness in Non-Mammalian Species." In Consciousness and Cognition 2005 Mar;14(1):169-87. Web address: http://www.nsi.edu/users/seth/Papers/HallmarksWeb.pdf

Edelman, D. 2004. Personal email, 19 July 2004.

Edelman G. and Tononi G. 2000. A Universe of Consciousness. New York: Basic Books.

Edelman, G. 2003. "Naturalizing consciousness: A theoretical framework." In Proceedings of the National Academy of Sciences U.S.A., 100(9): 5520-5524, 29 April 2003.
Web address: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=154377

Egan G., Silk T., Zamarripa F., Williams J., Federico P., Cunnington R., Carabott L., Blair-West J., Shade R., McKinley M., Farrell M., Lancaster J., Jackson G., Fox P. and Denton D. 2003. Article: "Neural correlates of the emergence of consciousness of thirst." In Proceedings of the National Academy of Sciences of the USA, December 9, 2003, vol. 100, no. 25, pp. 15241-15246. Web address: http://www.pnas.org/cgi/reprint/100/25/15241.pdf

Emerson, E. 1998. Article: "Researcher teases out secrets from surprisingly 'intelligent' viruses." In HSC Weekly, the weekly journal of the Health Sciences Campus at University of Southern California. October 30, 1998. Volume 4, Number 30. Web address: http://www.usc.edu/hsc/info/pr/1vol4/430/viruses.html

Emery N. and Clayton N. 2004. Article: "Comparing the Complex Cognition of Birds and Primates." In Comparative Vertebrate Cognition, ed. by Lesley J. Rogers and Gisela Kaplan. Kluwer Academic/Plenum Publishers, 2004. Web address: http://www.zoo.cam.ac.uk/zoostaff/madingley/library/member_papers/nemery/comparing_birds_and_primates_chapter.pdf

Encyclopedia Britannica. 1989. Article: "Animal Learning." Macropedia, Volume 1, p. 872. University of Chicago.

Epstein, A. 1982. “Instinct and Motivation as explanations for complex behaviour.” In The Physiological Mechanisms for Motivation, ed. D. W. Pfaff, Springer-Verlag, New York, pp. 25-58.

Esch H. E., Zhang S., Srinivasan, M. V. and Tautz J. 2001. "Honeybee dances communicate distances measured by optic flow." In Nature, vol. 411, pp. 581-583. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=11385571&dopt=Abstract

Evans, D. 2001. Emotion: The Science of Sentiment. Oxford University Press.

Evans, C. 2002. "Cracking the Code: Communication and Cognition in Birds." In Bekoff M., Allen C. and Burghardt G. (eds). The Cognitive Animal. Cambridge, Massachusetts: MIT. Paperback edition, pages 315-322. Web address: http://grimpeur.tamu.edu/~colin/TCA/Ch/Evans/evans-final.pdf

Faraco, J. 2000. Posting: "Re: Are there animals who don't sleep or that sleep very little?". Posting on Zoology, ID 964164818.Zo, August 1, 2000. Web address: http://www.madsci.org/posts/archives/aug2000/965504574.Zo.r.html

Federmeier K. 1997. “Perceiving a New Category: The Neurobiological Basis of Perceptual Categorization.” Cogsci.UCSD Technical Report 97.05. Department of Cognitive Science, University of California, San Diego. Web address: http://www-cogsci.ucsd.edu/research/files/technical/97_05.pdf

Feldman, R. 2003. Essentials of Understanding Psychology, 5th edition, University of Massachusetts, Amherst. See glossary for chapter 6, definition of "procedural memory". Web address of glossary: http://highered.mcgraw-hill.com/sites/0072494263/student_view0/chapter6/glossary.html

Ferguson H., Cobey S and Smith B. 2001. "Sensitivity to a change in reward is heritable in the honeybee, Apis mellifera". In Animal Behavior 61:527-534. Web address: http://www.culturaapicola.com.ar/apuntes/genetica/34_sencibilidad_cambios_hereditarios.pdf

Finlay, B.F., Darlington, R.B. & Nicastro, N. 2001. Developmental structure in brain evolution. In Behavioral and Brain Sciences, 24(2): 263-308. Web address: http://www.bbsonline.org/Preprints/OldArchive/bbs.finlay.html

Finsen, S. 1990. "On Moderation." In Bekoff, M. and Jamieson, D. (eds), Interpretation and Explanation in the Study of Behavior. Boulder, Colorado: Westview.

Fiorito G. and Scotto P. "Observational Learning In Octopus Vulgaris." In Science, vol. 256, issue 5056, April 24 1992, pp. 545-547. Web address of abstract: http://www.sciencemag.org/cgi/content/abstract/256/5056/545

Web address of full article: http://www.sciencemag.org/cgi/reprint/256/5056/545.pdf?ijkey=e87a7731d8d4e3e388d9e9b958bd696cde341a9c

Fisheries Society of the British Isles. 2002. "Fish Welfare." Briefing Paper 2, Fisheries Society of the British Isles, Granta Information Systems, Granta Information Systems, 82A High Street, Sawston, Cambridge, CB2 4H, England. Web address: http://www.nal.usda.gov/awic/pubs/Fishwelfare/FSBI.pdf

Forster, M. and Saidel, E. 1994. “Connectionism and the Fate of Folk Psychology.” In Philosophical Psychology 7: 437-452.

Forterre P. and Philippe H. 1999. "Where Is the Root of the Universal Tree of Life?" In BioEssays 21.10, pp. 871-879. Web address: http://imbs.massey.ac.nz/bio_evol/Topic4/Forterre&Philippe99.pdf

Fowler, T. 1995. Online article: “More than Meets the Eye: Aristotle on Sense Perception.” Web address: http://www.perseus.tufts.edu/GreekScience/Students/Tom/Anima.html

Fox, D. 2004. "Do fruit flies dream of electric bananas?" In New Scientist, vol. 181, issue 2434, 14 February 2004, page 32. Web address: http://grimpeur.tamu.edu/pipermail/animals/2004-February/000068.html

Franklin S., Baars B., Ramamurthy U. and Ventura M. 2005. “The Role of Consciousness in Memory.” In the online journal Brains, Minds and Media (http://www.brains-minds-media.org/). July 4, 2005 (bmm150). Web address: http://www.brains-minds-media.org/archive/150/bmm150.pdf

Freeman, W. 1991. “The Physiology of Perception.” In Scientific American, Vol. 264 (2), pp. 78-85. Web address: http://sulcus.berkeley.edu/FLM/MS/Physio.Percept.html

Frey R. G. 1980. Interests and Rights: The Case Against Animals. Oxford: Clarendon Press.

Friedlander B. 2002. "CU behaviorist: Facial markings help paper wasps identify each other." In Cornell Chronicle, 7 November 2002. Web address: http://www.news.cornell.edu/Chronicle/02/11.7.02/wasp_faces.html

Frye M. and Gray J. 2005. “Mechanosensory Integration for Flight Control in Insects.” In Methods in Insect Sensory Neuroscience (Thomas Christensen, ed.), CRC Press, pp. 108-129. Web address: http://www.physci.ucla.edu/research/frye/papers/CRC__chapter.pdf

Furlow, B. 1999. "Flower Power!" In New Scientist magazine (9 January 1999).

Galizia C. G, Menzel R. 2000. "Odor perception in honeybees: coding information in glomerular patterns." In Current Opinion in Neurobiology, August 2000, 10(4):504-510. Web address of abstract:
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=10981621&dopt=Abstract

Gallistel, C. R. 1998. "Insect Navigation: Brains as Symbol-Processors." In Scarborough D. and Sternberg S. (eds.) An Invitation to Cognitive Science, Vol. 4: Methods, Models, and Conceptual Issues. MIT Press. Web address: http://www.cogsci.msu.edu/insectnav.pdf

Gallistel, C. R. and Gibbon, J. 2001. "Computational versus Associative Models of Simple Conditioning." In Current Directions in Psychological Science, vol. 10, pp. 146-150. Web address: http://ruccs.rutgers.edu/faculty/GnG/Comp__vs_Assoc__Models.pdf

Gallistel, C. R. 2003. "Conditioning from an Information Processing Perspective." In Behavioural Processes, vol. 61(3), pp. 1-13. Web address:

http://ruccs.rutgers.edu/faculty/GnG/Cond%20from%20Info%20Proc%20Persp.pdf

Gallup, G. 1998. "Animal Self-Awareness: A Debate - Can Animals Empathize? - Yes." In Scientific American, Winter 1998 special issue, pp. 66-71. Web address: http://www.zoosemiotics.helsinki.fi/Can%20Animals%20EmpathizeYES.doc

Gallup G. Jr., Anderson J. and Shillito D. 2002. "The Mirror Test." In Bekoff M., Allen C. and Burghardt G. (eds). The Cognitive Animal. Cambridge, Massachusetts: MIT. Paperback edition, pages 325-334.

Geer D. 2002. "Organic Computing: Life that Computes." In TechWorthy.com (2002) magazine, Bedford Communications, 2002. Back issue available from Bedford Communications, Inc., 1410 Broadway, 21st floor, New York, NY 10018. Also in the March 2002 edition of Laptop magazine, published by Bedford Communications. Available on the Internet archive at http://web.archive.org/web/20010101-20050101re_/http://www.techworthy.com/magarchives/scitech/96646.html

Gentle, M. 2002. "Comparative vertebrate nociception and pain." In: Proceedings of the Mayday Conference: A Cross-Species Approach to Pain and Analgesia - 2002, Ludders J.W., Paul-Murphy J., Robertson S., Gaynor J., Hellyer P.W., Wong P.L. and Barakatt C. (eds.). International Veterinary Information Service, Ithaca NY (www.ivis.org). Web address: http://www.ivis.org/proceedings/mayday/Gentle/ivis.pdf

German Research Centre for Biotechnology (GBF). 2002. "Microbial Hg resistance." Web page contact: Dr. I. Wagner-Dobler. Web address: http://mik.gbf.de/mercury_remediation1/introduction2.html

Geyer M. and Markou A. 2000. "Animal Models of Psychiatric Disorders." Online article. American College of Neuropsychopharmacology (ACNP) - Psychopharmacology: the Fourth Generation of Progress. Web address: http://www.acnp.org/g4/GN401000076/Default.htm

Giurfa M., Eichmann B. and Menzel R. 1996. Symmetry perception in an insect. In Nature, volume 382, pp. 458-461. Web address of abstract: http://www.nature.com/nature/journal/v382/n6590/abs/382458a0.html

Giurfa, M. and Capaldi E. A. 1999. "Vectors, routes and maps: new discoveries about navigation in insects." In Trends in Neuroscience (TINS), volume 22, no. 6, 1999. Web address: http://www.fbmc.fcen.uba.ar/fca/Navegacion/Giurfa&Capaldi1999.pdf

Giurfa, M., Zhang, S., Jenett, A., Menzel, R., & Srinivasam, M.V. 2001. "The concepts of 'sameness' and 'difference' in an insect." In Nature, 19 April 2001, volume 410, pp. 930-933. Web address of abstract: http://www.nature.com/nature/journal/v410/n6831/abs/410930a0_fs.html . See also http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=11309617&dopt=Abstract

Giurfa, M. 2003. "The amazing mini-brain: lessons from a honey bee." In Bee World, volume 84(1), pp. 5-18. Web address of abstract: http://cat.inist.fr/?aModele=afficheN&cpsidt=15025844

Giurfa, M. 2005. “Left with the association, naked as it were? Ideas from honeybee learning.” Paper given at Workshop on Causality, organized by the Institute for Cognitive Science in Lyon and the University of Geneva. Date of publication: 6 June 2005. Web address: http://www.interdisciplines.org/causality/papers/8

Giurfa M. and Menzel R. 2003. "Human spatial representation derived from a honeybee compass". In Trends in Cognitive Sciences, vol. 7, no. 2, February 2003. No online abstract available.

Godfrey-Smith, P. 2001. Environmental Complexity and the Evolution of Cognition. In Sternberg R. and Kaufman J. (eds.) The Evolution of Intelligence. Lawrence Erlbaum Associates, 2001. Web address of article: http://www.its.caltech.edu/~theory/Godfrey-Smith.pdf

Gould, J. 1986. "The locale map of bees: do insects have cognitive maps?" In Science, 232:861-863. Web address of abstract: http://adsabs.harvard.edu/abs/1986Sci...232..861G

Gould, J. 1988. "A mirror-image ambiguity in honey bee visual memory." In Animal Behaviour, 36:487-492. No abstract located online.

Gould, J. 1990. “Honey bee cognition.” In Cognition, November 1990, 37(1-2):83-103. Web address of abstract:
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=2269009&dopt=Abstract

Gould J. L. and Gould C. G. 1994. The Animal Mind. New York: Scientific Library.

Gould, J. 2002. "Can Honey Bees Create Cognitive Maps?" In Bekoff M., Allen C. and Burghardt G. (eds). The Cognitive Animal. Cambridge, Massachusetts: MIT. Paperback edition, pages 41-45. Web address: http://grimpeur.tamu.edu/~colin/TCA/Ch/Gould/gould-final.pdf

Grandin, T. 1997. "Thinking the Way Animals Do." In Western Horseman, Nov. 1997, pp. 140-145. Web address: http://www.grandin.com/references/thinking.animals.html

Grandin, T. 1998a. "Consciousness in Animals and People with Autism." Online article, October 1998. Web address: http://www.grandin.com/references/animal.consciousness.html

Grandin T. & Deesing M. 1998b. Chapter: "Genetics and the Behaviour of Domestic Animals." In Behavioral Genetics and Animal Science, Academic Press, 1998. Web address: http://www.grandin.com/references/genetics.html

Grandin, T. & Deesing M. 2003. "Distress in Animals: Is it Fear, Pain or Physical Stress?" Paper given at American Board of Veterinary Practitioners - Symposium 2002, May 17, 2002, Manhattan Beach, California. Special Session: Pain, Stress, Distress and Fear. Updated in September 2003. Web address: http://www.grandin.com/welfare/fear.pain.stress.html

Grau, J. 2002. "Learning and Memory Without a Brain." In Bekoff M., Allen C. and Burghardt G. (eds). The Cognitive Animal. Cambridge, Massachusetts: MIT. Paperback edition, pages 77-88. Web address: http://grimpeur.tamu.edu/~colin/TCA/Ch/Grau/grau-final.pdf

Gray, J. A. 1995. “The contents of consciousness: A neuropsychological conjecture.” In Brain and Behavioral Sciences, 18: 659-722. Web address of abstract:
http://www.bbsonline.org/documents/a/00/00/05/52/index.html
Web address of article (minus illustrations):
http://www.bbsonline.org/documents/a/00/00/05/52/bbs00000552-00/bbs.gray.html

Gray, L. 2003. "A Mathematician Looks at Wolfram's New Kind of Science". In Notices of the American Mathematical Society, February 2003, pp. 200-211. Web address:
http://www.ams.org/notices/200302/fea-gray.pdf

Greenwood, J. 1991. The Future of Folk Psychology. Cambridge: Cambridge University Press.

Griffin, D. 1976. The Question of Animal Awareness: Evolutionary Continuity of Mental Experience. New York: The Rockefeller University Press.

Griffin, D. 1994. Animal Minds. University of Chicago Press. Paperback edition.

Griswood, L. 2001. "Lamb welfare in relation to castration and docking." In University of Sydney, Faculty of Veterinary Science, Veterinary Education and Information Network, Links - Essays. Web address: http://vein.library.usyd.edu.au/links/Essays/2001/griswood.html

Gruen, L. 2003. Article: "The Moral Status of Animals". In The Stanford Encyclopedia of Philosophy. Web address*:* http://plato.stanford.edu/entries/moral-animal/ (no earlier copies).

Guerrieri F., Lachnit H., Gerber B., and Giurfa M. 2005. “Olfactory blocking and odorant similarity in the honeybee”. In Learning and Memory, 2005 March; 12(2): 86–95. Web address of abstract: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1074325

Guzeldere G., Flanagan O. and Hardcastle V. 2000. "The nature and function of consciousness: Lessons from blindsight." In The New Cognitive Neurosciences (Gazzaniga M., ed.), pp. 1277-1284. Cambridge, Massachusetts: MIT Press. Cited in Rose, 2002a.

Haldane E. S. and Ross G. R. T. (translators). 1970. The Philosophical Works of Descartes, 2 vols. Cambridge: Cambridge University Press.

Haldane, J. B. S. 1947. What is Life? New York: Boni and Gaer. Web address of critical excerpt: http://www.marxists.org/archive/haldane/works/1940s/life.htm

Hall, S. 1999. "Fear Itself." Article in New York Times magazine, February 28, 1999. Web address: http://query.nytimes.com/gst/fullpage.html?sec=health&res=9A00E2DD133DF93BA15751C0A96F958260

Hamilton, G. "What is this octopus thinking?" 1997. In New Scientist, 7 June 1997, pp. 30-35. Web address: http://members.fortunecity.com/templarseries/cephpod.html

Hanlon, R.T. and Messenger, J.B. 1996. Cephalopod Behaviour. Cambridge: Cambridge University Press.

Hardie, J. 1997. "To take a decision, first you have to be angry." In New Scientist, Emotions supplement, 3 May 1997.
Web address: http://www.fortunecity.com/emachines/e11/86/hardie.html

Harrison A. and Schunn C. 2003. "ACT-R/S: Look Ma, No 'cognitive-map'!" Paper given at International Conference on Cognitive Modeling, 2003. Web address: http://simon.lrdc.pitt.edu/~harrison/pubs/iccm2003.pdf

Hardcastle V. 1997. Online paper: "Attention versus Consciousness: A Distinction with a Difference." Web address: http://www.phil.vt.edu/Valerie/papers/attencons.html

Hart, S. 1996. The Language of Animals. St Leonards, Sydney: Allen and Unwin.

Hartwell, S. 2001, 2003. "Do Cats Have Emotions?" Web address: http://messybeast.com/emoticat.html

Hauser M., Chomsky N. and Fitch W. T. 2002. "The Faculty of Language: What Is it, Who Has It, and How Did It Evolve?" In Science, vol. 298, 22 November 2002. Web address: http://www.ai.mit.edu/projects/dm/hauser-etal.pdf.

(For a response to this paper, see Jackendorff R. and Pinker, S. 2005. “The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky).” In Cognition, Volume 97, Issue 2, September 2005, Pages 211-225. Web address: http://pinker.wjh.harvard.edu/articles/papers/pinker_jackendoff.pdf. For a counter-response, see Fitch, W. T., Hauser M., Chomsky, N. 2005. “The Evolution of the Language Faculty: Clarifications and Implications.” In Cognition, Volume 97, Issue 2, September 2005, pp. 179-210. Web address: http://www.st-andrews.ac.uk/%7Ewtsf/FitchHauserChomsky.doc.)

Hauser M., Fitch W. T. 2003. “What are the Uniquely Human Components of the Language Faculty?” In M.H. Christiansen and S. Kirby (eds.), Language Evolution: The States of the Art, pp. 158-181. Oxford University Press. Web address of article: http://www.wjh.harvard.edu/~mnkylab/publications/languagespeech/langfaculty.pdf

Hawes, A. 1995. "What the Buzz is all about". In ZooGoer 24(6), September/October 1995. Published by Smithsonian National Zoological Park. Web address: http://natzoo.si.edu/Publications/ZooGoer/1995/6/buzzabout.cfm

Heisenberg M., Wolf R. and Brembs B. 2001. Flexibility in a Single Behavioral Variable of Drosophila. In Learning and Memory, 8:1-10. Web address: http://www.brembs.net/papers/dfsreview01.pdf

Hennessey T. M., Rucker W. B. and McDiarmid C. G. 1979. "Classical conditioning in paramecia." In Animal Learning and Behavior, 7, 417-423.

Hennig, B. 2004. Draft: "Cartesian Conscientia." [Final version to appear in: British Journal for the History of Philosophy, 2007.] Web address of abstract: http://www.borishennig.de/texte/descartes/diss (Until early 2006, the site carried a 50-page article in English, which summarised the conclusions of Hennig's German dissertation, "Was bedeutet 'conscientia' bei Descartes?" (Institut fur Philosophie der Universitat Leipzig). Page references in chapter 1 are citations from this English article, which is available on the Internet Wayback Machine at http://web.archive.org/web/20060209190515/http://www.borishennig.de/texte/descartes/diss/cartes_04b.pdf.)

Hennig, B. 2005. "Corrigibility as the Mark of the Mental" (12-page essay). Web address: http://www.borishennig.de/texte/2005/corrigibility.draft.pdf

Hennig, B. 2006. Ph.D. Dissertation: "Conscientia bei Descartes." Published by Verlag Karl Alber, 2006.

Herman L. 2002. "Exploring the Cognitive World of the Bottlenosed Dolphin." In Bekoff M., Allen C. and Burghardt G. (eds). The Cognitive Animal. Cambridge, Massachusetts: MIT. Paperback edition, pages, 275-284. Web address: http://grimpeur.tamu.edu/~colin/TCA/Ch/Herman/herman-final.pdf

Hermann B., Seidenberg M., Sears L., Hansen R., Bayless K., Rutecki P. and Dow C. 2004. “Cerebellar atrophy in temporal lobe epilepsy affects procedural memory.” In Neurology 2004, 63: 2129-2131. Web address of abstract: http://www.neurology.org/cgi/content/abstract/63/11/2129

Heyes C.M. and Galef B.G. 1996. Social Learning in Animals: The Roots of Culture. Academic Press: London. Cited in Brown and Laland, 2003.

Hinkle D. J. and Wood D. C. 1994. "Is Tube-Escape Learning by Protozoa Associative Learning?" in Behavioral Neuroscience, volume 108, no. 1 (1994), pp. 94 - 99.

Hoage, RJ, and Goldman L., eds. 1986. Animal Intelligence: Insights into the Animal Mind. National Zoological Park Symposia for the Public Series. Washington, DC: Smithsonian Institution Press.

Hodos W. and Campbell C. 1969. "Scala naturae: Why is there no theory in comparative psychology?" In Psychological Review 76: 337-350.

Holldobler, B. and Wilson, E. O. 1994. Journey to the Ants. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

Honore P., Rogers S., Schwei M., Salak-Johnson J., Luger N., Sabino M., Clohisy D. and Mantyh P. 1999. "Murine Models of Inflammatory, Neuropathic and Cancer Pain: Each

Generates a Unique Set of Neurochemical Changes in the Spinal Cord and Sensory Neurons." In Neuroscience, Vol. 98 No. 3, pp. 585-598, 2000. Web address: http://neurosystems.umn.edu/murine%20models.pdf

Horowitz A. 2002. Doctoral dissertation. "The behaviors of theories of mind, and A case study of dogs at play." Department of Cognitive Science, University of Southern California, San Diego. Web address: http://crl.ucsd.edu/%7Eahorowit/horowitz.pdf

Horgan, T. and Woodward, J. 1985. “Folk Psychology is Here to Stay.” In Philosophical Review 94: 197-226.

Huber, L. 2001. "Visual Categorization in Pigeons." In Cook, R. G. (ed.) 2001. Avian Visual Cognition. Tufts University. Web address: http://www.pigeon.psy.tufts.edu/avc/huber/default.htm

Humphrey, N. 1993. A History of the Mind. London: Vintage.

Humphreys M., Guerney K. and Prescott T. 2005. “Is There an Integrative Center in the Vertebrate Brain-Stem? A Robotic Evaluation of a Model of the Reticular Formation Viewed as an Action Selection Device.” In Adaptive Behavior, Vol. 13, No. 2, pp. 97-113. Web address of abstract: http://adb.sagepub.com/cgi/content/abstract/13/2/97

Illingworth, J. 1999. University of Leeds course notes: Biochemistry 3390. Web address: http://www.bmb.leeds.ac.uk/illingworth/motors/flagella.htm

Indiana University. 2003. A Moment of Science. Web address: http://amos.indiana.edu/library/scripts/waspid.html

Isaacs A., Daintith J. and Martin E. (eds.). 1999. 4th edition. A Dictionary of Science. Oxford University Press.

Jackson, P. 2002. Lecture notes for Psychology 333 - Principles of Learning. Powerpoint presentation: Rescorla-Wagner Theory (1972) of Classical Conditioning. Web address: http://pjackson.asp.radford.edu/rescorlawagner.ppt

Jacob, P. 2003. Article: “Intentionality.” In The Stanford Encyclopedia of Philosophy (Fall 2003 Edition), Edward N. Zalta (ed.). Web address of original article: http://plato.stanford.edu/archives/fall2003/entries/intentionality/

JAMA. 1990. "Persistent Vegetative State and the Decision to Withdraw or Withhold Life Support." In Journal of the American Association, January 19, 1990; 263(3): 426-430. Web address: http://www.ama-assn.org/ama/upload/mm/369/ceja_report_019.pdf

Jiva T. and Masoody M. 2003. "Case Report: A Case of Parasomnia." In Sleep Review, May/June 2003. Web address: http://www.sleepreviewmag.com/issues/articles/2003-05_11.asp

Jordan, A. 2002. "Pain without 'Pain'." Web address: http://serendip.brynmawr.edu/bb/neuro/neuro00/web2/Jordan.html

Jouvet, M. 1969. "Coma and other disorders of consciousness." pp. 62-69. In: Handbook of Clinical Neurology, Vol. 3 (Vinken, P.J. and G. W. Bruyn, eds.). New York: Elsevier Science Publishers. (Cited by Rose, 2002a.) Web address of article: http://sommeil.univ-lyon1.fr/articles/jouvet/hcn_69/contents.html

Jouvet M. and Jouvet D. 1963. "A Study of the Neurophysiological Mechanisms of Dreaming." In Electroenceph. Clin. Neurophysiol., 1963, Suppl. 24. Web address: http://sommeil.univ-lyon1.fr/articles/jouvet/ecn_63/lesions.html

Kacelnik A. 2004. "Meanings of Rationality." Chapter 2 in Rational Animals? (Nudd M. and Hurley S. eds.). Oxford University Press. Web address of preprint: http://www.cogsci.msu.edu/DSS/2004-2005/Kacelnik/Kacelnik%20Rationality.pdf

Kahn, C. S. 1979. "Sensation and Consciousness in Aristotle's Psychology." In Articles on Aristotle. Eds. Barnes J., Schofield, M. and Sorabji, R. London: Duckworth.

Kaiser, G. 2001. On-line Biology 230 (Microbiology) Lecture Guide - Unit 1 (Introduction, the Prokaryotic Cell and the Eukaryotic Cell). Web address: http://student.ccbcmd.edu/courses/bio141/lecguide/unit1/index.html (to access the material on differences between eukaryotes and prokaryotes, click on part B in the Introduction: http://student.ccbcmd.edu/courses/bio141/lecguide/unit1/proeu/proeu.html).

Kant, I. (ed. Paton, H. J.). 1983. The Moral Law. London: Hutchison & Co.

Kaplan, M. 2000. On-line article: "Pain and Analgesia in Herps." Web address: http://www.anapsid.org/herppain.html

Kaplan, M. 2000. On-line article: "Signs of Pain and Discomfort in Herps." Web address: http://www.anapsid.org/herppain3.html

Kauffman S. 1993. "The Origins of Order: Self-Organization and Selection in Evolution." New York: Oxford University Press.

Kauffman S. 2000. Investigations. New York: Oxford University Press.

Kavanau, J. L. 1997. "Origin and Evolution of Sleep: Roles of Vision and Endothermy." In Brain Research Bulletin, Vol. 42, No. 4, pp. 245-264. Web address: http://www.obee.ucla.edu/Faculty/Kavanau/originofsleep.html. Address of article in PDF format: http://www.obee.ucla.edu/Faculty/Kavanau/OriginEvolutionSleep.pdf.

Kazlev, M. A. 2002. Web page: "The Classification of Living Organisms." Web address: http://www.palaeos.com/Kingdoms/kingdoms.htm

Kennedy, J. S. 1992. The New Anthropomorphism. Cambridge University Press.

Kenny, A. J. P. 1975. Will, Freedom and Power. Oxford: Basil Blackwell.

Kentridge, R. 1995. Comparative Psychology lecture notes (lecture 5). Web address: http://www.dur.ac.uk/robert.kentridge/comp5.html

Kihlstrom J., Couture L., Schacter D. and Cork R. 1999. "Anesthesia, effects on cognitive functions." In Encyclopedia of Neuroscience (Adelman G. and Smith B., eds.). Amsterdam: Elsevier. Cited in Rose, 2002a.

Kihlstrom, J. 2000. “Consciousness and memory during surgical anesthesia”. An expansion of Kihlstrom et al., 1999. Web address: http://www.institute-shot.com/anesthesia_and_surgery.htm

Kilian A. E. and Muller B. S. 2002. “Life-like Learning in Technical Artefacts: Biochemical vs. Neuronal Mechanisms.” In: Conference Proceedings of the 9th International Conference on Neural Information Processing (ICONIP'02) / CD-ROM / Wang, Lipo [Hrsg.], (2002), pp. 296 – 300. Web address: http://www.ais.fraunhofer.de/%7Ebsm/CR1714.pdf

Kinser, P. 2000. "Brains on the Web - Comparative Neuroanatomy and Intelligence." Web address: http://serendip.brynmawr.edu/bb/kinser/Home1.html

Klingenberg, R. 1999. "Perception of Pain in Reptiles." In The Vivarium 10 (4): 45-49. Web address: http://www.anapsid.org/herppain2.html

Klinowska, M. 1994. "Brains, Behaviour and Intelligence in Cetaceans (Whales, Dolphins and Porpoises)." In 11 Essays on Whales and Man, second edition, 1994. Web address: http://www.highnorth.no/Library/Myths/br-be-an.htm

Knight, W. 2003. Article: "'Digital organisms' illuminate evolution". In New Scientist, 7 May 2003. Web address: http://www.newscientist.com/news/news.jsp?id=ns99993706

Koch C. and Crick F. 2000. "Some thoughts on consciousness and neuroscience." In The New Cognitive Neurosciences (Gazzaniga M., ed.), pp. 1285-1294. Cambridge, Massachusetts: MIT Press. Cited in Rose, 2002a.

Koch C., Crick F. 2001. "The neural basis of consciousness." In: The International Encyclopedia of the Social and Behavioral Sciences, Smelser, N. and Baltes, P., eds., Vol. 4, pp. 2600-2604. Elsevier, Oxford, United Kingdom. Web address: http://www.klab.caltech.edu/~koch/Elsevier-NCC.html

Koch, C. 2003. "The Quest for Consciousness - A Neurobiological Approach". 2003. Englewood (Colorado): Roberts and Company Publishers. Web address: http://www.questforconsciousness.com/

Kolb B. and Whishaw I. 1995. Fundamentals of Human Neuropsychology. New York: W. H. Freeman. Cited in Rose, 2002a.

Komischke B., Giurfa M., Lachnit H., & Malun D. 2002. Successive olfactory reversal learning in honeybees. Learning & Memory, 9, 122-129. Web address: http://staff-www.uni-marburg.de/~ag-lachn/Abstracts.htm#Komischke%20et%20al.%20 (abstract), http://staff-www.uni-marburg.de/~lachnit/KoGiLaMa02.pdf (full article).

Komischke B., Sandoz J.-C., Lachnit H. and Giurfa M. 2003. “Non-elemental processing in olfactory discrimination tasks needs bilateral input in honeybees.” In Behavioural Brain Research, 2003 Oct 17; 145(1-2):135-43. Web address: http://www.staff.uni-marburg.de/~lachnit/Abstracts[d].html#KoSaLaGi03 or http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=14529812&dopt=Abstract (abstract only), http://www.staff.uni-marburg.de/~lachnit/KoSaLaGi03.pdf (full article).

Koob, G. 2000. "Animal Models of Drug Addiction." Online article. American College of Neuropsychopharmacology (ACNP) - Psychopharmacology: the Fourth Generation of Progress. Web address: http://www.acnp.org/g4/GN401000072/Default.htm

Koopowitz, H. and Keenan, L. 1982. “The primitive brains of platyhelminthes.” In Trends In Neurosciences 5(3): 77-79. Cited in Prescott (2007), below.

Koshland, Daniel E., Jr. 2002. "The Seven Pillars of Life." In Science (March 22, 2002), pp. 2215-2216. Web address: http://www.sciencemag.org/cgi/reprint/295/5563/2215.pdf

Kuba M., Meisel D., Byrne R., Griebel U., & Mather J. 2003. "Looking at Play in *Octopus vulgaris*". In Proceedings of the International Symposium Coleoid Cephalopods Through Time : 17-19 September 2002. International Symposium Coleoid Cephalopods Through Time (2002 : Berlin, Germany). Edited by Kerstin Warnke, Helmut Keupp & Sigurd von Boletzky. Berlin : Fachrichtung Palaontologie, Institut fur Geologische Wissenschaften, Freie Universitat Berlin,

2003. (Berliner palaobiologische Abhandlungen; Bd. 3). Web address: http://www.ulrikegriebel.com/images/KubaBerlinPlay-Paper.pdf

Laird, J. 1994. Web site: "Brooks' Subsumption Architecture." Web address: http://ai.eecs.umich.edu/cogarch3/Brooks/Brooks.html For a short extract on the background to Brooks' subsumption architecture, see: http://ai.eecs.umich.edu/cogarch3/Brooks/Brooks_Background.html

Lakoff G. and Johnson M. 1999. Philosophy in the Flesh. New York: Basic Books.

Laland K. N., Brown C. and Krause J. 2003. "Learning in Fishes: From three-second memory to culture." In Fish and Fisheries, 2003, vol. 4, pp. 199-202. Web address: http://www.geocities.com/culumbrown/LalandBrownKrauseFAF2003

Laland, K. N., Odling-Smee, J. and Feldman, M. W. 1999. “Evolutionary consequences of niche construction and their implications for ecology.” In Proceedings of the National Academy of Sciences, U.S.A. August 1999. Volume 96, pp. 10242 - 10747. Web address: http://www.pnas.org/cgi/reprint/96/18/10242.pdf

Laland K. N., Williams K. 1997. “Shoaling generates social learning of foraging information in guppies.” In Animal Behavior 53: 1161-1169. Cited in Bshary, Wickler and Fricke, 2002. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=9236013&dopt=Abstract

Langley, D. 2002. On-line article: "The Octopus." Web address: http://helium.vancouver.wsu.edu/~langleyd/octopus/octopushomepage.htm

Laureys S., Goldman S., Phillips C., Van Bogaert P., Aerts J., Luxen A., Franck G., Maquet P. 1999. "Impaired effective cortical connectivity in vegetative state: preliminary investigation using PET." In Neuroimage 1999 April, vol. 9(4): 377-382. Cited in Rose, 2002a. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=10191166&dopt=Abstract

Laureys S., Faymonville M., Luxen A., Lamy M., Franck G., Maquet P. 2000a. "Restoration of thalamocortical activity after recovery from persistent vegetative state." In The Lancet, 355:1790-1791. Cited in Rose, 2002a. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=10832834&query_hl=3&itool=pubmed_docsum

Laureys S., Faymonville M., Janssens N., Del Fiore G., Degueldre C., Aerts J., Luxen A., Moonen G., Lamy M., Maquet P. 2000b. "Functional neuroanatomy of vegetative state: a lesional study." In Soc. Neurosci. Abstr., 26:1236. (Abstract from a Society for Neuroscience Annual Meeting Presentation.) Cited in Rose, 2002a.

Laureys S., Faymonville M., Degueldre C., Del Fiore G., Damas P., Lambermont B., Janssens N., Aerts J., Franck G., Luxen A., Moonen G., Lamy M., Maquet P. 2000c. "Auditory processing in

the vegetative state." In Brain, 123: 1589-1601. Cited in Rose, 2002a. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=10908189&query_hl=3&itool=pubmed_docsum
Web address of full article: http://brain.oxfordjournals.org/cgi/reprint/123/8/1589

Laureys S., Owen A. M. and Schiff N. D. 2004. "Brain function in coma, vegetative state, and related disorders." In *The Lancet,* Neurology, Vol. 3, pp. 537-546, September 2004. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15324722
A reprint of the full article is available from the author on request.

Laureys S. 2005. "The neural correlate of (un)awareness: lessons from the vegative state." In *Trends in Cognitive Science,* Vol. 9, No. 12, pp. 556-559, December 2005. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16271507
A reprint of the full article is available from the author on request.

Lawson-Tancred H. 1986. Translation of Aristotle's De Anima. London: Penguin Books.

Leahy, M. P. T. 1994. Against Liberation. Revised paperback edition. London: Routledge.

LeDoux, J. 1999. The Emotional Brain: The Mysterious Underpinnings of Emotional Life. Paperback edition. London: Phoenix.

LeDoux, J. 2003. "The Emotional Brain, Fear and the Amygdala." In Cellular and Molecular Neurobiology, Vol. 23, Nos. 4/5, October 2003. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=14514027&dopt=Abstract

Legrand, D. 2001. "Action and theories of movement, definition of intentional action." Paper given at C.N.I.A. (Cognitique - Naturalizing Intention in Action) video-conference (14 May 2001) between the University of Provence (Aix-en-Provence, France), the University of Jussieu (Paris, France) and the University of Parma (Parma, Italy). Web address: http://www.up.univ-mrs.fr/wcnia/preparationb.htm

Leibniz, G.W. 1714. Monadology. Cited in The Stanford Encyclopedia of Philosophy, article "Leibniz's Philosophy of Mind". Last modified: July 15, 2002. Web address of current article: http://plato.stanford.edu/entries/leibniz-mind/ Web address of original article: http://plato.stanford.edu/archives/fall2002/entries/leibniz-mind/

Lenski R., Ofria C., Pennock R. and Adami C. 2003. Article: "The Evolutionary Origin of Complex Features". In Nature, vol. 423, 8 May 2003, pp. 139-144. Web address: http://myxo.css.msu.edu/papers/nature2003/Nature03_Complex.pdf

Leontev A. 1978. Activity, Consciousness, and Personality. Web address: http://lchc.ucsd.edu/MCA/Paper/leontev/

Lewis, D. 1972. "Psychophysical and Theoretical Identifications". Australasian Journal of Philosophy 50: 249-58.

Li F. F., VanRullen R., Koch C. and Perona P. 2002. “Rapid natural scene categorization in the near absence of attention.” In Proc. Natl. Acad. Sci. 99 (14): 9596-9601. Web address: http://www.pnas.org/cgi/reprint/99/14/9596

Libet B. 1999. "Consciousness: neural basis of conscious experience." In Encyclopedia of Neuroscience (Adelman G. and Smith B., eds.). Amsterdam: Elsevier. Cited in Rose, 2002a.

Libet, B. 2003. “Can Conscious Experience Affect Brain Activity?” In Journal of Consciousness Studies, 10, No. 12, pp. 24-28. Web address of full article: http://www.imprint.co.uk/pdf/Libet.pdf

Liotti M., Brannan S., Egan G., Shade R., Madden L., Abplanalp, B., Robillard R., Lancaster J., Zamarripa F., Fox P., and Denton D. 2001. "Brain responses associated with consciousness of breathlessness (air hunger)". In PNAS, vol 98, no. 4, pp. 2035-2040. Web address: http://www.pnas.org/cgi/content/full/98/4/2035

Liotti M. and Panksepp J. 2003. "Imaging Human Emotions and Affective Feelings". In Panksepp (ed.), Textbook of Biological Psychiatry, chapter 2. Wiley, New York.

Lipp, O. 1998. Lecture notes: Psychology 269: Human Associative Learning. Web address: http://www2.psy.uq.edu.au/~landcp/PY269/index.html

For the quote on the Rescorla-Wagner model cited in the thesis, see http://www2.psy.uq.edu.au/~landcp/PY269/r-wmodel/r-wmodel.html

Lissek S. and Gunturkun O. 2003. "Dissociation of Extinction and Behavioral Disinhibition: The Role of NMDA Receptors in the Pigeon Associative Forebrain during Extinction." In Journal of Neuroscience, September 3 2003, vol. 22, pp. 8119-8224. Web address: http://www.bio.psy.ruhr-uni-bochum.de/papers/Lissek%20J%20Neurosci%202003.pdf

Luce R.D., Raiffa H. 1957. Games and decisions. Wiley: New York. Cited in Bshary, Wickler and Fricke, 2002.

Lurz, R. 2000. "A Defense of First-Order Representationalist Theories of Mental-State Consciousness." In Psyche, 6(1), February 2000. Web address: http://psyche.cs.monash.edu.au/v6/psyche-6-01-lurz.html

Lurz, R. 2003. "Neither HOT nor COLD: An Alternative Account of Consciousness." In Psyche, 9(1), January 2003. Web address: http://psyche.cs.monash.edu.au/v9/psyche-9-01-lurz.html

Machery, E. 2005. “Concepts Are Not a Natural Kind.” In Philosophy of Science, 72 (July 2005), pp. 444-467. Web address: http://www.pitt.edu/~machery/papers/concepts%20are%20not%20a%20natural%20kind_machery.pdf

Machin, K. 1999. "Amphibian Pain and Analgesia." In Journal of Zoo and Wildlife Medicine, 30 (1): 2-10. Web address of abstract: http://www.biopsychiatry.com/painamph.htm

MacKenzie, D. 2000. "Protein at a Price." In New Scientist magazine, 18 March 2000, pp. 32-36.

Mackintosh, N. J. 1974. The Psychology of Animal Learning. New York: Academic Press.

Macphail, E. 1982. Brain and Intelligence in Vertebrates. Clarendon Press, Oxford.

Macphail, E. 1998. The Evolution of Consciousness. Oxford University Press.

Maher, B. 2002. "Uprooting the Tree of Life." In The Scientist, Volume 16, Issue 18, Sep. 16, 2002. Web address: http://www.the-scientist.com/2002/09/16/26/1

Mallot, H. A. 2000. Chapter 11: Visual Navigation. In Computational Vision: Information Processing in Perception and Visual Behavior, MIT Press, Cambridge, Massachusetts.

Margulis L. and Schwartz K. 1988. Five Kingdoms: An Illustrated Guide to the Phyla of Life on Earth. San Francisco: W. H. Freeman & Co.

Martin C. C. and Gordon R. G. 2001. "The Evolution of Perception." In Cybernetics and Systems, volume 32, pp. 393 - 409. Web address: http://www.biologiateorica.it/papers/05papers.htm

Martin, L. 1998. Online article: "Can sleepwalking be a murder defense?" Web address: http://www.lakesidepress.com/pulmonary/Sleep/sleep-murder.htm

Martin, V. 2000. "Box Jellyfish May Be Able To See." McGraw-Hill Higher Education: Regional Perspectives in Marine Biology, Case Study #9. Web address: http://www.mhhe.com/biosci/pae/marinebiology/casestudies/case_09.mhtml

Masson, J. and McCarthy, S. 1996. When Elephants Weep: The Emotional Lives of Animals. London: Vintage.

Mataric, M. J. 1990. A Distributed Model for Mobile Robot Environment-Learning and Navigation, MIT AI Lab, AITR-1228. (Cited by Prescott, 2007, below.)

Mather, J. 2003. "Concept-formation in octopuses?" Personal email, 8 September 2003.

Mather, J. and Anderson, R. 1998. "What behavior can we expect of octopuses?" In The Cephalopod Page (on-line), edited by Dr. James Wood. Web address: http://www.thecephalopodpage.org/behavior.php

Mather, J. and Anderson, R. 2000. "Octopuses are smart suckers!" In The Cephalopod Page (on-line), edited by Dr. James Wood. June 11, 2000. Web address: http://is.dal.ca/%7Eceph/TCP/smarts.html

McClintock, J. 2003. "The Secret Life of Ants." In Discover, Vol. 24, No. 11, November 2003. Web address: http://www.discover.com/issues/nov-03/features/the-secret-life-of-ants/?page=1

McCourt R. M., Chapman R. L., Buchheim M. A. and Mishler B. D. 1995. "Green Plants." Web address: http://tolweb.org/tree?group=Green_plants&contgroup=Eukaryotes

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

McGovern Institute for Brain Research at MIT. 2003. "C. elegans neural development and behavior." Web address: http://web.mit.edu/mcgovern/html/Areas_of_Research/c_elegans.shtml

McGraw-Hill Dictionary of Scientific and Technical Terms. 2003. Entry: “procedural memory.” McGraw-Hill Companies, Inc. Sixth edition. Web address of definition: http://www.answers.com/topic/procedural-memory

McGraw-Hill Encyclopedia of Science and Technology. 2005. Entry: “animal.” The McGraw-Hill Companies, Inc. Web address of definition: http://www.answers.com/topic/stuffed-animal (scroll down to view the McGraw-Hill definition).

McGregor, P. K. 1993. “Signalling in Territorial Systems – a context for individual identification, ranging and eavesdropping.” In Philosophical Transactions of the Royal Society of London Series B – Biological Sciences, 340: 237-244. Cited in Brown and Laland, 2003.

McKay C. 2004. "What Is Life? and How Do We Search for It in Other Worlds?" In PLOS Biology, Volume 2, Issue 9, e302, September 2004.

Medin D., Lynch E., Solomon K. 2000. “Are There Kinds of Concepts?” In Annual Review of Psychology, 51:121-147. Web address: http://www.psych.northwestern.edu/psych/people/faculty/medin/PDF's/medin,%20lynch,%20solomon%20-Are%20there%20kinds%20of%20concepts.pdf

Medina, L. 2002. "Proposal for New Nomenclature." Talk given at Avian Brain Nomenclature Forum, 18-20 July 2002, Durham (North Carolina, USA). Web address: http://avianbrain.org/nomen/proposals/Medina-july_2.html

Menzel R., Brandt R., Gumbert A., Komischke B. and Kunze J. 2000. "Two spatial memories for honeybee navigation." In Proceedings of the Royal Society of London, Series B, 22 May 2000, volume 267, pp. 961-968. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=10874744&dopt=Abstract
Web address of full paper: http://www.neurobiologie.fu-berlin.de/menzel/Pub_AGmenzel/MenzelBrandtGumbertKomischkeKunze2000ProcRoySoc.pdf

Menzel R. and Giurfa M. 2001. "Cognitive architecture of a mini-brain: the honeybee." In TRENDS in Cognitive Sciences, Vol. 5, No. 2, February 2001.

Menzel, R. 2003. Personal email: Re: Blocking in invertebrates. 21 July 2003.

Merfeld, D. 2001. "Must all action halt during sensorimotor mismatch?" In Behavioral and Brain Sciences, Vol. 24 (1), 2001, pp. 189-190. Web address: http://jvl.dyndns.org/jvpl/dmm1.pdf

Merikle P. and Daneman M. 2000. "Conscious vs. unconscious perception." In The New Cognitive Neurosciences (Gazzaniga M., ed.), pp. 1295-1304. Cambridge, Massachusetts: MIT Press. Cited in Rose, 2002a.

Merriam-Webster Online Dictionary. 2006. Web address: http://www.m-w.com/ (dictionary entry consulted: “learn”).

Microsoft Help and Support Glossary. 2006. Entry: “Memory.” Web address: http://support.microsoft.com/default.aspx?scid=%2Fsupport%2Fglossary%2Fm.asp

Midgley, M. 1984. Animals and Why They Matter. University of Georgia.

Midgley, M. 1993. "'The Four-Leggeds, the Two-Leggeds, and the Wingeds": An Overview of Society and Animals.' In PSYETA, Vol. 1, No. 1. Web address: http://www.psyeta.org/sa/sa1.1/shapiro.html

Milinski M. 1987. “Tit for tat in sticklebacks and the evolution of cooperation.” In Nature 325: 433–435. Cited in Bshary, Wickler and Fricke, 2002. Web address of abstract: http://www.nature.com/nature/journal/v325/n6103/abs/325433a0.html

Milinski M., Külling D., Kettler R. 1990a. “Tit for tat: sticklebacks *Gasterosteus aculeatus* ‘trusting’ a cooperative partner.” In Behavioral Ecology 1(1):7–11. Cited in Bshary, Wickler and Fricke, 2002. Web address of abstract: http://beheco.oxfordjournals.org/cgi/content/abstract/1/1/7

Milinski M., Pfluger D., Külling D., Kettler R. 1990b. “Do sticklebacks cooperate repeatedly in reciprocal pairs?” In Behavioral Ecology and Sociobiology, July 1990, 27(1):17–21. Cited in Bshary, Wickler and Fricke, 2002. Web address of abstract: http://www.springerlink.com/content/pp87722552668180/

Milius, S. 2001. "It's a snake! No, a fish. An octopus?" In Science News, vol. 160, no. 9, September 1, 2001, p. 132. Web address: http://www.sciencenews.org/20010901/fob2.asp

Miller, G. 1999. "Does the savagery of predation in nature show that God either isn't, or at least isn't good-hearted?" Web address: http://www.christian-thinktank.com/pred2.html (The article contains an interesting, well-researched discussion of animal sentience.)

Millikan, R. 1984. Language, Thought, and Other Biological Categories. Cambridge, Massachusetts: MIT Press.

Millikan, R. 1993. White Queen Psychology and Other Essays for Alice. Cambridge, Massachusetts: MIT Press.

Millikan, R. 2000. On Clear and Confused Ideas. Cambridge: Cambridge University Press.

Milner D. and Goodale, M. 1995. The Visual Brain in Action. Oxford University Press.

Miyake A. and Shah P. 1999. Models of Working Memory. Cambridge: Cambridge University Press. Cited in Franklin *et al.,* 2005.

Moller, R. 2001. "Do Insects Use Templates or Parameters for Landmark Navigation?" In Journal of Theoretical Biology, vol. 210, no. 1, pp. 33-45. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=11343429&dopt=Abstract

Web address of full article:

http://www.ti.uni-bielefeld.de/downloads/publications/MoellerTempPara.pdf

Monnin T. and Peeters C. 1999. "Dominance hierarchy and reproductive conflicts among subordinates in a monogynous queenless ant." In Behavioral Ecology, vol. 10, no. 3, pp. 323-332. Web address of abstract: http://beheco.oupjournals.org/cgi/content/abstract/10/3/323

Moren J. & Balkenius C. 2000. "A Computational Model of Emotional Learning in the Amygdala." From animals to animats 6: Proceedings of the 6th International conference on the simulation of adaptive behavior. Cambridge, MA. MIT Press, 2000. Web address: http://www.lucs.lu.se/People/Christian.Balkenius/PDF/Moren.Balkenius.2000.Amygdala.pdf

Morris, P. J. 1993. "The developmental role of the extracellular matrix suggests a monophyletic origin of the kingdom Animalia." In Evolution 47(1): 152-165. Cited by the University of California

Museum of Paleontology web site, web page "Metazoa: More on the Morphology". Web address: http://www.ucmp.berkeley.edu/phyla/metazoamm.html

Morrison G., Kumar N., Wen J., Runciman S., Neghandi H. and van der Kooy D. 1995. "lrn-1 and lrn-2 Block Associative Learning in C. elegans." Early 1995 International Worm Meeting abstract 60. Web address: http://elegans.swmed.edu/wli/[wm95p60]/

Morrison G. E., Wen J. Y. M., Runciman S., van der Kooy D. 1999. "Olfactory associative learning in Caenorhabditis elegans is impaired in lrn-1 and lrn-2 mutants." In Behavioral Neuroscience 113: 358-367. Web address: http://elegans.swmed.edu/wli/[cgc3545]/

Mouritsen H. and Frost B. 2002. "Virtual migration in tethered flying monarch butterflies reveals their orientation mechanisms." In Proceedings of the National Academy of Sciences, vol. 99, no. 15, pp. 10162-10166. Web address: http://www.pnas.org/cgi/reprint/152137299v1.pdf

Moynihan, M. 1985. Communication and Noncommunication by Cephalopods. Bloomington: Indiana University Press.

Myers P. Z. 2005. Online blog: “Bird Brains.” 9 September, 2005. Web address: http://pharyngula.org/index/weblog/comments/bird_brains/

Nadler, S. 1997. “Descartes’ Dualism?” Review of a book by Gordon Baker and Katherine Morris. In Philosophical Books, 48, 3, (July 1997), pp. 157-169. Web address: http://www.newdualism.org/papers/S.Nadler-et-al/Baker-Morris-review.htm

Nagel, E. 1979. "Teleology Revisited." In Teleology Revisited and Other Essays in the Philosophy and History of Science, edited by E. Nagel. New York: Columbia University Press. pp. 275-316.

National Health and Medical Research Council (Australian Government). 2003. Information paper: "Post-Coma Unresponsiveness (Vegetative State): A Clinical Framework for Diagnosis." Endorsed 18 December 2003. Web address: http://www7.health.gov.au/nhmrc/publications/synopses/hpr23syn.htm

Neander K. 1991a. "Functions as Selected Effects: The Conceptual Analyst's Defence." In Philosophy of Science 58:168-184. Reprinted in Allen, Colin, Marc Bekoff, and George Lauder, eds. 1998. Nature’s Purposes: Analyses of Function and Design in Biology: The MIT Press.

Neander K. 1991b. "The Teleological Notion of 'Function'." In Australasian Journal of Philosophy 69:454-468. Reprinted in Buller, D. J., ed. 1999. Function, Selection, and Design. Albany: State University of New York Press. pp. 123-41.

Neander K, 1995. “Misrepresenting and Malfunctioning.” In Philosophical Studies 79, pp. 109-141.

New J.G. 2002. “Multimodal integration in the feeding behavior of predatory teleost fishes.” In Brain Behavior and Evolution 2002; 59(4): 177-189. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=12138338&dopt=Abstract

New J.G. and Kang P.Y. 2000. “Multimodal integration in the strike-feeding behavior of predatory fishes.” In Philosophical Transactions of the Royal Society of London B: Biological Sciences 2000, Sep. 29; 355(1401): 1321-1324. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=11079423
Web address of full article: http://www.journals.royalsoc.ac.uk/media/3d325d0erq3jpv4dyjvw/contributions/h/b/t/f/hbtfnunag8b2u413.pdf

Nieuwenhuys R., ten Donkelaar H. and Nicholson C. 1998. The Central Nervous System of Vertebrates. Berlin: Springer.

Nilsson D., Gislen L., Coates M., Skogh C. and Garm A. 2005. “Advanced optics in a jellyfish eye.” In Nature, 435(7039), 12 May 2005, pp. 201-205. Web address of abstract: http://www.nature.com/nature/journal/v435/n7039/abs/nature03484.html
Web address of article: http://www.thep.lu.se/pub/Preprints/05/lu_tp_05_05.pdf

Nissani M. 2004. "Theory of Mind and Insight in Chimpanzees, Elephants, and Other Animals?" In Rogers L. and Kaplan G. (eds.), Comparative Vertebrate Cognition: Are Primates Superior to

Non-Primates? (a featured volume in Developments in Primatology: Progress and Prospects), Chapter 7 (pp. 227-261). New York: Kluwer Academic/Plenum Publishers, January 2004. Web address: http://www.is.wayne.edu/mnissani/ElephantCorner/Conscrev.htm

Noble, I. "Octopus arms do the thinking." BBC news report, Thursday, 6 September, 2001. Web address: http://news.bbc.co.uk/2/hi/science/nature/1529397.stm

Noe R., Schaik C. P. van, and Hooff J. A. R. A. M. van. 1991. “The market effect: an explanation for pay-off asymmetries among collaborating animals.” In Ethology 87:97–118. Cited in Bshary, Wickler and Fricke, 2002.

Norman M., Finn J., and Tregenza, T. 2001. "Dynamic mimicry in an Indo-Malayan octopus." In Proceedings of the Royal Society of London B vol. 268 (September 7, 2001), pp. 1755-1758. Web address: http://www.biology.leeds.ac.uk/staff/tbt/Papers/MDNetal_PB01.pdf

Nuffield Council on Bioethics. 2001. Web address: http://www.nuffieldbioethics.org/animalresearch/pp_0000001071.asp

Nuttley, W. 2003a. Personal email. "Re: Blocking in invertebrates." 18 July 2003.

Nuttley, W. 2003b. Personal email. "Re: Operant conditioning in C. elegans." 27 August 2003.

O’Callaghan, J. 2000. “From Augustine’s Mind to Aquinas’ Soul.” Talk given at the University of Notre Dame Thomistic Institute, July 14-21, 2000. Web address: http://www2.nd.edu/Departments//Maritain/ti00/ocallagh.htm

Okano H., Hirano T. and Balaban E. 2000. “Learning and Memory.” In Proceedings of the National Academy of Sciences of the United States of America (PNAS), vol. 97, no. 23, pp. 12403-12404. Web address of abstract: http://www.pnas.org/cgi/content/abstract/97/23/12403 Web address of article: http://www.pnas.org/cgi/reprint/97/23/12403

Ong R., Morris J., O'Dwyer J., Barnett J., Hemsworth P. and Clarke I. 1997. "Behavioural and EEG changes in sheep in response to painful electrical stimuli." In Australian Veterinary Journal, vol. 75, no. 3, pp. 189-193. Web address: http://www.ava.com.au/avj/mar97/189.htm

Panksepp, J. 1998. Affective Neuroscience: The Foundations of Human and Animal Emotions. New York: Oxford University Press.

Panksepp, J. 2001. "On the subcortical sources of human emotions and the primacy of emotional-affective (action-perception) processes in human consciousness." In Evolution and Cognition, Volume 7, No. 2, 2001, pp. 134-140. Web address: https://notes.utk.edu/Bio/greenberg.nsf/f5b2cbf2a827c0198525624b00057d30/10ea06f6b78647d985256cdb00531f47/$FILE/Affect%20article%20by%20JP%20-%20introduced.pdf

Panksepp, J., Bernatzky, G. 2002. "Emotional sounds and the brain: the neuro-affective foundations of musical appreciation." In Behavioral Processes, Volume 60, No. 2, November-December 2002, pp. 133-155. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=12426066&dopt=Abstract

Panksepp, J. 2002. "Foreword: The MacLean Legacy and Some Modern Trends in Emotion Research." In The evolutionary neuroethology of Paul MacLean: convergences and frontiers. Eds. Gerald A. Cory, Jr. and Russell Gardner, Jr. Westport, Connecticut: Praeger.

Panksepp J., Moskal J., Panksepp J. and Kroes R. 2002. "Comparative Approaches in Evolutionary Psychology: Molecular Neuroscience Meets the Mind." In Neuroendocrinology Letters Special Issue, Suppl. 4, Vol. 23, December 2002. Web address: http://www.nel.edu/pdf_w/23s4/NEL231002R11_Panksepp_.pdf

Panksepp, J. 2003a. "Can Anthropomorphic Analyses of Separation Cries in Other Animals Inform us about the Emotional Naure of Social Loss in humans?" Comment on Blumberg and Sokoloff (2001). In Psychological Review, 2003, Vol. 110, No. 2, pp. 376-388. Web address: http://www.psych.ufl.edu/~steh/PSB6088/Panksepp.pdf

Panksepp J., Burgdorf J. 2003b. "Toward an Open-Minded Study of the Neuroevolutionary Substrates of Affect: Rejoinder to Blumberg to Blumberg and Sokoloff's (2003) Reply." In Psychological Review, Vol. 110, No. 2, pp. 385-396. Web address: http://www.psych.ufl.edu/~steh/PSB6088/Panksepprep.pdf

Panksepp J., Burgdorf J. 2003c. "'Laughing' rats and the evolutionary antecedents of human joy?" In Physiology and Behavior, 79 (2003), pp. 533-547. Web address: http://caspar.bgsu.edu/~courses/Reading/Papers/2003PanBur.pdf

Panksepp, J. 2003d. "Feeling the Pain of Social Loss." In Science, Volume 302, 10 October 2003, pp. 237-239. Web address: http://www.scn.ucla.edu/pdf/PankseppPerspective.pdf

Panksepp, J. 2003e. Review article: "Damasio's Error?" In Consciousness & Emotion 4:1 (2003), pp. 111-134. New York: Oxford University Press.

Panksepp, J. 2003f. "At the interface of the affective, behavioral, and cognitive neurosciences: Decoding the emotional feelings of the brain." In Brain and Cognition, 52(2003):4-14. Elsevier Science (USA). Web address: http://www.tturc.uci.edu/researchint/panksepp.pdf

Panksepp, J. 2003g. "The Peri-Conscious Substrates of Affective Consciousness." In Psyche, vol. 9 no. 15, December 2003. Web address: http://psyche.cs.monash.edu.au/symposia/mangan/panksepp.html

Panksepp Jules B., Huber R. 2004. "Ethological analyses of crayfish behaviour: a new invertebrate system for measuring the rewarding properties of psychostimulants." In Behavioural Brain Research, 153 (2004): 171-180. Web address: http://ntp.neuroscience.wisc.edu/students/student-art/panksepp153.pdf

Paradis S., Cabanac M. 2004. "Flavor aversion learning induced by lithium chloride in Reptiles but not in Amphibians." In Behavioral Processes, 2004, July 30; 67(1): 11-18. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=15182921&dopt=Abstract

Parsons L., Denton D., Egan G., McKinley M., Shade R., Lancaster J. and Fox P. 2000. "Neuroimaging evidence implicating cerebellum in sensory/cognitive processes associated with thirst." In Proceedings of the National Academy of Sciences USA, 97(5):2332-2336, February 29, 2000. Web address: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=15801

Parsons L., Egan G., Liotti M., Brannan S., Denton D., Shade R., Robillard R., Madden L., Abplanalp B. and Fox P. 2001. "Neuroimaging evidence implicating cerebellum in the experience of hypercapnia and hunger for air." In Proceedings of the National Academy of Sciences USA, 2001, February 13; 98(4):2041-2046. Web address: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=29378

Penfield, W. 1975. The Mystery of the Mind: A Critical Study of Consciousness and the Human Brain. Princeton University Press. (Cited by Block, 1995.)

Pepperberg, I. M. 1987. "Acquisition of the same/different concept by an African Grey parrot *(Psittacus erithacus).* Learning with respect to categories of color, shape, and material." In Animal Learning and Behavior, 15, 423-432. Cited in Young and Wasserman, 2001 (see below).

Pepperberg, I. 1991. "Referential Communication with an African Grey Parrot." In Harvard Graduate Society Newsletter, Spring 1991, pp. 1-4.

Pepperberg, I. 2002. "Cognitive and Communicative Abilities of Grey Parrots." In Bekoff M., Allen C. and Burghardt G. (eds). The Cognitive Animal. Cambridge, Massachusetts: MIT. Paperback edition, pages 247-254. Also in Current Directions in Psychological Science 11 (3), 83–87. Web address of abstract: http://www.blackwell-synergy.com/doi/abs/10.1111/1467-8721.00174 Web address of full article: http://philebus.tamu.edu/~colin/Phil320/tamu/parrot.pdf

Perner, J. 2003. "Dual Control and the Causal Theory of Action: The Case of Non-intentional Action." In N. Eilan & J. Roessler (eds.), Agency and Self-Awareness - Issues in Philosophy and Psychology, pp. 218-243. Oxford: Oxford University Press.

Pest Control Times, 16 November 2004. Article: "Decoy Robots To Rid Us Of Cockroaches." Web address: http://www.pctonline.com/news/news.asp?ID=3071

Petkewich, R. 2002. "Mercury pollution may contribute to antibiotic resistance." In Environmental Science and Technology Online News, 24 June 2002. Web address: http://pubs.acs.org/subscribe/journals/esthag-w/2002/jun/science/rp_mercury.html

Pilcher, H. "Old dog learns new tricks." In Nature Science Update, 11 June 2004. Web address: http://www.nature.com/news/2004/040607/pf/040607-8_pf.html

Pitcher T.J., Green D.A., Magurran A.E. 1986. “Dicing with death: predator inspection behavior in minnow *Phoxinus phoxinus* shoals." In Journal of Fish Biology, 28 (4): 439–448. Cited in Bshary, Wickler and Fricke, 2002. Web address of abstract: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1095-8649.1986.tb05181.x

Pitt, D. 2002. Article "Mental Representation". In The Stanford Encyclopedia of Philosophy (Winter 2002 Edition), ed. Edward N. Zalta. Web address: http://plato.stanford.edu/archives/win2002/entries/mental-representation/

Plotnik J., de Waal F. and Reiss D. 2006. "Self-Recognition in an Asian Elephant." In PNAS (Proceedings of the National Academy of Sciences of the U.S.A.), 7 November 2006, vol. 103, no. 45, pp. 17053-17057. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&list_uids=17075063 Web address of full article: http://www.emory.edu/LIVING_LINKS/pdf_attachments/PlotniketalPNAS.pdf

Povinelli, D. 1998. "Animal Self-Awareness: A Debate - Can Animals Empathize? - Maybe not." In Scientific American, Winter 1998 special issue, pp. 67-75. Web address: http://geowords.com/lostlinks/b36/7.htm

Prescott T., Redgrave P. and Gurney K. 1999. "Layered control architectures in robots and vertebrates." In Adaptive Behavior, vol. 7, pp. 99-127. Web address: http://www.dcs.shef.ac.uk/~amanda/AB99.pdf

Prescott, T. 2007. “Forced moves or good tricks in design space? Landmarks in the evolution of neural mechanisms for action selection.” To appear in Adaptive Behavior: Special Issue on Mechanisms of Action Selection, 15(1). Web address: http://www.abrg.group.shef.ac.uk/publications/prescott_AB07.pdf

Preuss, P. 2000. "Circuits of a Cell." In Berkeley Lab Research Review, Winter 2000 edition. Web address: http://www.lbl.gov/Science-Articles/Research-Review/Magazine/2000/Winter/features/06circuits.html

Preuss T. 2000. "What's human about the human brain?" In The New Cognitive Neurosciences (Gazzaniga M., ed.), 2nd edition, pp. 1219-1234. Cambridge, Massachusetts: MIT Press. Cited in Rose, 2002a. Web address of fullarticle: http://research.yerkes.emory.edu/Preuss/2000PreussWhat%27sHuman.pdf

Procyshyn, A. 2001. Article: "Lucid Representation". In PHICS (PHIlosophy and Cognitive Science) Conference proceedings, pp. 132-145. (The PHICS Conference was held at Carleton University, Ottawa, Canada, from September 28 to 30, 2001.) Web address: http://www.carleton.ca/iis/PHICS/Proceedings/procyshyn.pdf

Proust J. 1999. “Mind, Space and Objectivity in Non-Human Animals.” In Erkenntnis 51: 41-58.

Ramsey, F. 1990. Article: "General Propositions and Causality". In Philosophical Papers, ed. D. H. Mellor. Cambridge: Cambridge University Press.

Ramsey, W., Stich, S. and Garon, J. 1990. “Connectionism, Eliminativism and the Future of Folk Psychology.” In Philosophical Perspectives 4: 499-533.

Ramsey, W. 1991. “Where Does the Self-Refutation Objection Take Us?” In Inquiry 33: 453-65.

Ramsey, W. 2003. "Eliminative Materialism." In The Stanford Encyclopedia of Philosophy (Fall 2003 Edition), Edward N. Zalta (ed.). Web address: http://plato.stanford.edu/archives/fall2003/entries/materialism-eliminative/

Randerson, J. 2001. "Hatchery fish get survival training." In New Scientist, 4 October 2001. Web address: http://www.newscientist.com/article.ns?id=dn1374

Randerson, J. 2002. "Gene study gives language lesson." In New Scientist, 14 August 2002. Web address: http://www.newscientist.com/news/news.jsp?id=ns99992678

Rankin, C. 2001. Wormlab Web page. Web address: http://www.psych.ubc.ca/%7Ecrankin/

Rankin, C. 2003. Personal email, 31 May 2003.

Raup, D. M. and Seilacher, A. 1969. Fossil Foraging Behaviour: Computer Simulation. Science 166: 994-995. (Cited in Prescott, 2007, above.)

Ravenscroft, I. 2003. "Simulation, Collapse and Humean Motivation". Mind and Language 18: 162-74.

Ravenscroft, I. 2004. "Folk Psychology as a Theory." In The Stanford Encyclopedia of Philosophy (Spring 2004 Edition), Edward N. Zalta (ed.). Web address: http://plato.stanford.edu/archives/spr2004/entries/folkpsych-theory/

Regan, T. 1988. The Case for Animal Rights. London: Routledge.

Reiss, D. and Marino, L. 2001. "Mirror Self-Recognition in the Bottlenose Dolphin: A Case of Cognitive Convergence." In PNAS (Proceedings of the National Academy of Sciences of the United States of America), 8 May 2001, Volume 98, no. 10, pp. 5937 - 5942. Web address: http://www.pnas.org/cgi/doi/10.1073/pnas.101086398

Reppert, V. 1992. “Eliminative Materialism, Cognitive Suicide, and Begging the Question.” In Metaphilosophy 23: 378-92.

Rifkin, S. 1995. "The Evolution of Primate Intelligence." In The Harvard BRAIN (Harvard's Undergraduate Neuroscience Magazine), Volume II, Issue I, Spring 1995. Web address: http://hcs.harvard.edu/~husn/BRAIN/vol2/Primate.html

Ristau C, ed. 1991. Cognitive Ethology: The Minds of Other Animals. Hillsdale NJ: Lawrence Erlbaum Associates.

Roediger H., Marsh E. & Lee S. 2002. "Varieties of memory." In D.L. Medin & H. Pashler (Eds.), Stevens' Handbook of Experimental Psychology. Third Edition, Volume 2: Memory and Cognitive

Processes, pp. 1-41. New York: John Wiley & Sons. Web address: http://media.wiley.com/product_data/excerpt/0X/04713803/047138030X.pdf

Rogers L. and Kaplan G. (eds.) 2003. Comparative Vertebrate Cognition: Are Primates Superior to Non-Primates? Kluwer Primatology Series: Developments in Primatology: Progress and Prospect. New York: Kluwer Academic/Plenum Publishers.

Rollin, B. 1989. The Unheeded Cry: Animal Consciousness, Animal Pain, and Science. Oxford: Oxford University Press.

Rorden, C. 2005. University of Nottingham, Course C81HBB Neuropsychology, lecture 2. Web address: http://www.psychology.nottingham.ac.uk/staff/cr1/nsy1b.pdf

Rorty, R. 1979. Philosophy and the Mirror of Nature. Oxford, Blackwell.

Rose, J. D. 2002a. "The Neurobehavioural Nature of Fishes and the Question of Awareness and Pain." In Reviews in Fisheries Science, vol. 10, issue 1, pp. 1-38. Web address: http://www.animal-health-online.de/drms/rosefish.pdf or http://www.nal.usda.gov/awic/pubs/Fishwelfare/Rose.pdf

Rose, J. D. 2002b. "Do Fish Feel Pain?" Online article. Web address: http://www.g-feuerstein.com/Presse/fishpain.htm

Rose, J. D. 2003a. "A critique of the paper: 'Do fish have nociceptors: Evidence for the evolution of a vertebrate sensory system' published in Proceedings of the Royal Society, 2003 by Sneddon, Braithwaite and Gentle." Reprinted online by United States Department of Agriculture, in its list of publications on fish welfare. Web address: http://www.nal.usda.gov/awic/pubs/Fishwelfare/RoseC.pdf

Rose, J. D. 2003b. "A rebuttal to the paper: 'Do fish have nociceptors: Evidence for the evolution of a vertebrate sensory system' published in Proceedings of the Royal Society, 2003 by Sneddon, Braithwaite and Gentle." Published online in In Depth, second quarter, 2003. In Depth is the official newsletter of the Amateur Fishermen’s Association of the Northern Territory. Web address: http://www.afant.com.au/newsletter/pdf/Indepth%202nd%20quarter%202003.pdf.

Rose, J. K. and Rankin, C. 2001. "Analyses of Habituation in Caenorhabditis elegans." In Learning and Memory, Vol. 8, No. 2, pp. 63-69, March/April 2001. Web address: http://www.learnmem.org/cgi/content/full/8/2/63

Rose, S. 1993. Article: "No way to treat the mind." In "New Scientist", April 17, 1993. Web address: http://nootropics.com/sceptic

Rosenthal, D. 1986. "Two concepts of consciousness." In Philosophical Studies 49: 329-359.

Rosenthal, D. 1993. “State Consciousness and Transitive Consciousness.” In Consciousness and Cognition, 2, 4 (December 1993), pp. 355-363. Web address: http://web.gc.cuny.edu/cogsci/sctc.htm

Rosenthal D. 2002a. "How many kinds of consciousness?" In Consciousness and Cognition 11(2002), pp. 653-665. Web address: http://web.gc.cuny.edu/cogsci/how-many.pdf

Rosenthal D. 2002b. "Explaining consciousness." In Philosophy of Mind: Classical and Contemporary Readings, ed. David J. Chalmers, New York: Oxford University Press, 2002, pp. 406-421. Web address: http://web.gc.cuny.edu/cogsci/explg.pdf

Ross, G. M. 1975-1999. Hobbes' Objections to Descartes' Meditations, with Descartes' Replies. Translation copyright George MacDonald Ross. Web address: http://www.philosophy.leeds.ac.uk/GMR/hmp/texts/modern/hobbes/objections/objindex.html

Roth, G. 2003. "The quest to find consciousness." In Scientific American Mind, 2003, 14 (1), pp. 32-39. Web address: http://www.hku.hk/physio/Session28Jan04b.pdf

Russell, B. 1905/1956. “On denoting.” Originally printed in Mind, New Series, Vol. 14, No. 56. (Oct., 1905), pp. 479-493. Reprinted in Marsh, R.C. (ed.) (1956) Bertrand Russell, Logic and Knowledge, Essays 1901-1950, New York: Capricorn Books. Web address: http://cscs.umich.edu/~crshalizi/Russell/denoting/

Ryder, D. and Martin C. B. 1998. Draft paper: "The autonomic nervous system and Dretske on phenomenal consciousness." Web address: http://homepage.mac.com/ancientportraits/drsite/ANS-Ryder_Martin.pdf

Ryder, D. 2003. "Re: The autonomic nervous system and Dretske on phenomenal consciousness." Personal email, 4 September 2003.

Saeki, Yamamoto and Iino. 2001. "Plasticity of chemotaxis revealed by paired presentation of a chemoattractant and starvation in the nematode Caenorhabditis elegans." In Journal of Experimental Biology, Vol. 204, Issue 10, pp. 1757-1764. Web address: http://jeb.biologists.org/cgi/content/abstract/204/10/1757

Sample, I. 2000. "Primitive Intelligence." In New Scientist (27 September 2000). Web address: http://www.newscientist.com/news/news.jsp?id=ns999925
(Source of original report by Sample:
Nakagaki T., Yamada H. and Toth A. 2000. "Intelligence: Maze-solving by an amoeboid organism." In Nature magazine, vol. 407, p. 470. Web address: http://www.nature.com/cgi-taf/Dynapage.taf?file=/nature/journal/v407/n6803/abs/407470a0_fs.html See also Nakagaki T. and Kobayashi R. 2004. Online Presentation: "How does an amoeba tackle some geometrical puzzles?" Web address: http://sbs-xnet.sbs.ox.ac.uk/complexity/complexity_PDFs/talk_Nakagaki.pdf)

Sanabria F., Baker F. and Rachlin H. 2003. "Learning by pigeons playing against tit-for-tat in an operant prisoner's dilemma." In Learning and Behavior, 31(4):318-331, November 2003. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=14733481

Schatz B., Chameron S., Beugnon G. and Collett T. 1999. "The use of path integration to guide route learning in ants." In Nature vol. 399, pp. 769-772, 24 June1999. Web address of abstract: http://www.nature.com/cgi-taf/DynaPage.taf?file=/nature/journal/v399/n6738/abs/399769a0_r.html

Schmajuk, N. 2007. “Computational Models of Classical Conditioning.” Article in Scholarpedia, the free peer-reviewed encyclopedia written by scholars from all around the world. As at 30 January, 2007. Web address: http://www.scholarpedia.org/article/Computational_Models_of_Classical_Conditioning

Schroedinger, E. 1944. What is Life? The Physical Aspect of the Living Cell Cambridge: Cambridge University Press. Web address of chapter 6 (Order, Disorder and Entropy): http://dieoff.org/page150.htm

Schusterman R., Kastak C. R. and Kastak D. 2002. "The Cognitive Sea Lion: Meaning and Memory in the Laboratory and in Nature." In Bekoff M., Allen C. and Burghardt G. (eds). The Cognitive Animal. Cambridge, Massachusetts: MIT. Paperback edition, pages 217-228. Web address: http://grimpeur.tamu.edu/~colin/TCA/Ch/Schusterman/schusterman-final.pdf

Schwei M., Honore P., Rogers S., Salak-Johnson J., Finke M., Ramnaraine M., Clohisy D. and Mantyh P. 1999. "Neurochemical and Cellular Reorganization of the Spinal Cord in a Murine Model of Bone Cancer Pain." In The Journal of Neuroscience, December 15 1999, 19(24):10886-10897. Web address: http://neurosystems.umn.edu/bonecancer.pdf

Schwitzgebel, E. 2006. Article: "Belief". In The Stanford Encyclopedia of Philosophy (Fall 2006 edition), Edward N. Zalta (ed.). Web address: http://plato.stanford.edu/archives/fall2006/entries/belief/

Scigliano, E. 2003. "Through the Eye of an Octopus." In Discover magazine, vol. 24, no. 10, October 2003. Web address: http://www.discover.com/issues/oct-03/features/feateye/

Searle, J. 1983. Intentionality. Cambridge: Cambridge University Press.

Searle, J. 1992. The Rediscovery of the Mind. MIT Press. (Cited by Block, 1995.)

Searle, J. 1999. Mind, Language and Society. Paperback edition. Basic Books: New York.

Seeley, T. 1995. The Wisdom of the Hive: the social physiology of honey bee colonies. Harvard University Press.

Sellars, W. 1956. "Empiricism and the Philosophy of Mind". In H. Feigl and M. Scriven (eds), Minnesota Studies in the Philosophy of Science (vol. 1). Minneapolis: University of Minesota Press. Web address: http://www.ditext.com/sellars/epm.html

Seth A., Baars B. and Edelman D. 2005. "Criteria for consciousness in humans and other mammals." In Consciousness and Cognition, March 2005; 14(1):119-39. Web address: http://www.nsi.edu/users/seth/Papers/ConCog2004aWeb.pdf

Shaw P., Cirelli C., Greenspan R. and Tononi G. 2000. "Correlates of sleep and waking in *Drosophila melanogaster*." In Science, Vol. 287, 10 March 2000, pp. 1834-1837. Web address: http://ntp.neuroscience.wisc.edu/faculty/fac-art/Tononi287.pdf

Shettleworth S. 2001. Review: "Animal cognition and animal behaviour." In Animal Behaviour, 2001, vol. 61, pp. 277-286. Web address: http://www.cogsci.msu.edu/DSS/2003-2004/shettleworth/Shettleworth2001.pdf

Shettleworth S. 2002. Article: "Spatial Behavior, Food Storing, and the Modular Mind." In Bekoff M., Allen C. and Burghardt G. (eds.), The Cognitive Animal, pp. 127-135. MIT Press, Cambridge, Massachusetts. Web address: http://grimpeur.tamu.edu/~colin/TheCognitiveAnimal/Ch/Shettleworth/shettleworth-final.pdf

Shewmon D., Holmes G. and Byrne P. 1999. "Consciousness in congenitally decorticate children: developmental vegetative state as self-fulfilling prophecy?" In Developmental Medicine & Child Neurology (1999), vol. 41, pp. 364-374. Web address: http://hydranencephaly.com/researchresults.htm

Shields, C. 2003. Article: "Aristotle's Psychology". In The Stanford Encyclopedia of Philosophy (Summer 2003 Edition), Edward N. Zalta (ed.). Web address of original article: http://plato.stanford.edu/archives/sum2003/entries/aristotle-psychology/

Shields, C. 2003. Supplementary Article: "Controversies Surrounding Aristotle's Theory of Perception." In The Stanford Encyclopedia of Philosophy (Summer 2003 Edition), Edward N. Zalta (ed.). Web address: http://plato.stanford.edu/entries/aristotle-psychology/suppl3.html/

Silby, B. 1998. E-paper. "On A Distinction Between Access and Phenomenal Consciousness." Web address: http://www.def-logic.com/articles/silby011.html

Siewert, C. 2003. Article: “Consciousness and Intentionality.” In The Stanford Encyclopedia of Philosophy (Summer 2003 Edition), Edward N. Zalta (ed.). Web address of original article: http://plato.stanford.edu/archives/sum2003/entries/consciousness-intentionality/

Singer, P. 1977. Animal Liberation. New York: Avon Books.

Sjolander, S. 1993. "Some cognitive breakthroughs in the evolution of cognition and consciousness, and their impact on the biology of language." In Evolution and Cognition 3: 1-10.

Skoyles J. 1999. "Child development and autism: A cerebellar-prefrontal model." Web address: http://cogprints.org/791/00/autism.htm

Slobodchikoff, C. 2002. "Cognition and Communication in Prairie Dogs." In In Bekoff M., Allen C. and Burghardt G. (eds). The Cognitive Animal. Cambridge, Massachusetts: MIT. Paperback edition, pages 257-264. Web address: http://philebus.tamu.edu/~colin/Phil320/tamu/prairiedogs.pdf

Sleepdisorderchannel. 2004. Online article: "Sleepwalking." Web address: http://www.sleepdisorderchannel.net/sleepwalking/index.shtml

Sleep Research Society. 1997. "Basics of Sleep Behavior." (See part B. The Phylogeny of Sleep and part I. Functions of Sleep.)
Web address: http://www.sleephomepages.org/sleepsyllabus/sleephome.html

Slovic P., Finucane M., Peters E. and MacGregor D. 2004. "Risk as Analysis and Risk as Feelings: Some Thoughts about Affect, Reason, Risk, and Rationality." In Risk Analysis, Volume 24, Issue 2, April 2004, Page 311. Web address: http://cancercontrol.cancer.gov/brp/presentations/slovic.pdf

Smith, A. 2002. "Analgesics: just what we wanted." In Signaling Update, 20 December 2002. Web address: http://www.signaling-gateway.org/update/updates/200212/nrd981.html or http://www.nature.com/drugdisc/res_high/articles/nrd981.html

Smith, J. 1991. "A Question of Pain in Invertebrates." In ILAR (Institute for Laboratory Animal Research) Journal (formerly ILAR News), Volume 33 (1 - 2), 1991. Web address: http://dels.nas.edu/ilar_n/ilarjournal/33_1_2/V33_1_2Question.shtml

Sneddon L., Braithwaite V. & Gentle M. (2003) "Do fish have nociceptors? Evidence for the evolution of a vertebrate sensory system." In Proceedings of the Royal Society of London B, (June 7, 2003), vol. 1270 (no. 1520): 1115-1121. Web address of abstract: http://www.journals.royalsoc.ac.uk/(vgvua1v2gjkpra55lsuiffqp)/app/home/contribution.asp?referr er=parent&backto=issue,2,15;journal,96,318;linkingpublicationresults,1:102024,1.Web address of full article: http://www.journals.royalsoc.ac.uk/media/e2egpkurypc9e2jhpq0x/contributions/e/8/v/y/e8vyfa4ndt5c4q0a.pdf.

Sorabji, R. 1993. Animal Minds and Human Morals. London: Gerald Duckworth & Co. Ltd.

Staddon J. E. R. 1983. Adaptive Behaviour and Learning. Cambridge University Press.

Stefano G. B., Salzet B. and Fricchione G. L. 1998. “Enkelytin and opioid peptide association in invertebrates and vertebrates: Immune activation and pain.” In Immunology Today, June 1998, Volume 19, no. 6, pp. 265 - 268. Web address of **full article:** http://www.univ-lille1.fr/lea/Menu_du_Site/Aims/ENKELYTIN%20AND%20OPIOID%20PEPTIDE%20ASSOCIATION%20IN%20INVERTEBRATES%20AND.PDF Web address of **abstract:** http://www.nel.edu/23_2/NEL230202R01_Stefano.htm

Stephan A. 1999. “Are Animals Capable of Concepts?” In Erkenntnis 51: 79-92.

Stephens D. and Anderson D. 2001. "The adaptive value of preference for immediacy: when short-sighted rules have far-sighted consequences." In Behavioral Ecology, 12:330-339, 2001. Web address: http://nash.cbs.umn.edu/lab/pdf/context.pdf

Stevens, C. 1997/98. "A Whole Animal Alternative Model for Pain Research." In Animal Welfare Information Center Newsletter, Winter 1997/98, Volume 8, no. 3 to 4. Web address: http://www.nal.usda.gov/awic/newsletters/v8n3/8n3steve.htm

Stich S. 1979. "Do Animals Have Beliefs?" In Australasian Journal of Philosophy, vol. 57, no. 1 (March 1979): 17-18.

Stich, S. 1983. From Folk Psychology to Cognitive Science. Cambridge, MA: MIT Press.

Stich, S. and Ravenscroft, I. 1994. "What is Folk Psychology?" In Cognition 50: 447-68. Web address: http://ruccs.rutgers.edu/tech_rpt/folkpsych5.pdf

Stich, S. 1996. Deconstructing the Mind. New York: Oxford University Press.

Stoerig, P. and Cowey, A. 1997. "Blind-sight in man and monkey." In Brain (1997), 120(3), pp. 535-559. Web address: http://brain.oxfordjournals.org/cgi/reprint/120/3/535 (**Note:** article takes a while to download.)

Strube, G. 1998. "Modelling Motivation and Action Control in Cognitive Systems." In U. Schmid, J. Krems & F. Wysocki (Eds.), Mind Modelling. 1998. Berlin: Pabst, pp. 89-108. Web address: http://cognition.iig.uni-freiburg.de/team/members/strube/ActionControl98.pdf

Suboski M. D. and Templeton J. J. 1989. “Life skills training for hatchery fish: social learning and survival.” In Fisheries Research 7:343–352. Cited in Bshary, Wickler and Fricke, 2002.

Suddendorf T. and Busby J. 2003. "Like it or not? The mental time travel debate: Reply to Clayton et al." In Trends in Cognitive Science, October 2003, 7(10):437-438. Web address: http://staff.science.uva.nl/~michiell/docs/SuddendorfBusbyTiCSResp.pdf

Super H. *et al.* 2001. “Two distinct modes of sensory processing observed in monkey primary visual cortex (V1).” Nat. Neurosci. 4, 304-310. Web address: http://www.nyu.edu/gsas/dept/philo/courses/consciousness05/LammeSensory.pdf

Tamminga C., Shadmehr R. and Holcomb H. 2000. “Images in Neuroscience. Cognition: procedural memory.” In American Journal of Psychiatry, February 2000, 157(2):162. Web address of article: http://ajp.psychiatryonline.org/cgi/reprint/157/2/162

Tang S. and Guo A. "Choice Behavior of Drosophila Facing Contradictory Visual Cues." 2001. In Science, vol. 294, pp. 1543-1547, 16 November 2001. Web address of abstract: http://www.sciencemag.org/cgi/content/abstract/294/5546/1543. Web address of full article: http://library.ibp.ac.cn/Html/slwj/000172240500054.pdf

Templeton J.J. 1987. Unpublished Ph.D. thesis. Cited in Bshary, Wickler and Fricke, 2002.

Thompson D. 1985. Online paper: “Intentionality and Causality in John Searle.” Web address: http://www.ucs.mun.ca/~davidt/Searle.htm

Thorpe, W. H. 1956 . Learning and instinct in animals. London: Methuen.

Tibbetts, E. 2002. "Visual signals of individual identity in the wasp Polistes fuscatus." In The Proceedings of the Royal Society of London B. Issue 269, number 1499, pages 423-428. Web address of **abstract:**

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=12137570&dopt=Abstract

Tinbergen, N. 1951. The Study of Instinct, Oxford University Press, New York.

Tomioka M., Kunitomo H. and Iino Y. 2002. “The insulin-like signaling pathway participates in learn in salt chemotaxis.” Third Japanese Worm Meeting, 2002. Web address: http://www.wormjp.umin.jp/jp/jwm2002abst/5111.html

Tononi, G. 2004. “Consciousness and the brain: theoretical aspects.” In Encyclopedia of Neuroscience (on-line), 2004. Web address: http://www.jsmf.org/meetings/2003/nov/consciousness_encyclopedia_2003.pdf

Treede R., Kenshalo D., Gracely R. and Jones A. 1999. "The Cortical representation of pain." In Pain, 1999 Feb., 79(2-3):105-111. Cited in Rose, 2002a. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=10068155&dopt=Abstract

Tulving, E. "Episodic and semantic memory." In E. Tulving & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.

Tulving, E. 2002. "Episodic memory and common sense: how far apart?" In Episodic Memory: New Directions in Research. Edited by Baddeley A., Conway M. and Aggleton J. New York: Oxford University Press. pp. 269-288. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=11571040&dopt=Abstract

University of Edinburgh. 2006. “Study shows hummingbirds are smart operators.” In The University of Edinburgh News and Events, 6 March 2006. Web address: http://www.ed.ac.uk/news/060306hummingbird.html

University of Utah. 2002. "How Bacteria Swim to Dinner: Teams of Sensors May Change Faint Food Smells into Marching Orders." In University of Utah News and Public Relations, April 29 2002. Web address: http://www.utah.edu/unews/releases/02/apr/dinner.html

Utah State University. 1995. "Learning About Fooods and Locations." Fact Sheet. Behavioral Principles and Practices, No. 1.1.2. Web address: http://extension.usu.edu/files/factsheets/Learn%20Foods_Location.pdf

Utah State University. 2001. "Training Animals to Avoid Foods." Fact Sheet. Application of Behavioral Principles - Diet Training, No. 2.1.1. Web address: http://extension.usu.edu/files/factsheets/Training_to_avoid.pdf

Van Gulick. 2004. Article: "Consciousness". In The Stanford Encyclopedia of Philosophy (Fall 2004 Edition), Edward N. Zalta (ed.). Web address of original article: http://plato.stanford.edu/archives/fall2004/entries/consciousness/

Van Swinderen B. and Greenspan R. 2003. "Salience modulates 20-30 Hz brain activity in *Drosophila.*" In Nature Neuroscience, Vol. 6, No. 6, pp. 579-586, June 2003. Web address: http://www.ini.unizh.ch/~pfmjv/InsectCognition/nat+neurosci_6_579.fly-salience.pdf

Varner, G. E. 1998. In Nature's Interests? Interests, Animal Rights and Environmental Ethics. New York: Oxford University Press.

Varner, G. E. 2000. "Which animals have desires?" A lecture on Animal Rights v. Animal Welfare. Web address: http://www-phil.tamu.edu/~gary/awvar/lecture/desires.html

Varner, G. E. 2000. Philosophy 314 - Environmental Ethics Course Notes. Web address: http://www-phil.tamu.edu/~gary/ee/index.html

Vince, G. 2004. "Puzzled monkeys reveal key language step." In "New Scientist", 15 January 2004. Web address: http://www.newscientist.com/news/news.jsp?id=ns99994572

Vines, G. 1994. "The emotional chicken." In "New Scientist", 22 January 1994.

Waggoner, B. 1994, 1998. Web page: "The Biosphere - Life on Earth", accessed from the University of California Museum of Paleontology Web site.

Web address: http://www.ucmp.berkeley.edu/alllife/threedomains.html

Waggoner, B. 1996. Web page: "Anthony van Leeuwenhoek", accessed from the University of California Museum of Paleontology Web site. Web address: http://www.ucmp.berkeley.edu/history/leeuwenhoek.html

Wakelin, D. 2003. Psychology lecture notes for course AS110: Learning. Lecture 4: "Macphail vs. Bitterman, A nil-nil draw?" Web address: http://psych.unn.ac.uk/books/as110/2003/l4_2003.html

Walker, S. 1983. Animal Thought. Routledge and Kegan Paul, London. International Library of Psychology Series. Web address of chapter 4: http://www.psyc.bbk.ac.uk/people/academic/walker_s/pubs/an-thought/at-ch4.html

Walton, M. 2002. "Study: Apes lack gene for speech." CNN.com Sci/Tech report, August 15, 2002. Web address: http://edition.cnn.com/2002/TECH/science/08/15/coolsc.speech/index.html

Washington University School of Medicine. 1997. Neuroscience Tutorial (An illustrated guide to the essential basics of clinical neuroscience created in conjunction with the first year course for medical students.) Created by Diana Weedman Molavi, Ph.D., at the Washington University School of Medicine. Home page: http://thalamus.wustl.edu/course/

For information on H.M. and the role of the hippocampus, select the link: Medial Temporal Lobe and Memory. Web address: http://thalamus.wustl.edu/course/limbic.html

Wasserman E. 2002. "General Signs." In Bekoff M., Allen C. and Burghardt G. (eds). The Cognitive Animal. Cambridge, Massachusetts: MIT. Paperback edition, pages 175-182. Web address: http://grimpeur.tamu.edu/~colin/TCA/Ch/Wasserman/wasserman-final.pdf

Web Sciences International and Sleep Research Society (United States). 1997. Web page: "Basics of Sleep Behaviour". See parts A (What is Sleep?) and B (The Phylogeny of Sleep). Web address: http://www.sleephomepages.org/sleepsyllabus/

Webb B. 2004. "Neural mechanisms for prediction: do insects have forward models?" In Trends in Neurosciences 27:278-282. Web address: http://homepages.inf.ed.ac.uk/bwebb/publications/tins_compact.pdf

Weber, B. 2003. Article: "Life". In The Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.) Current Web address: http://plato.stanford.edu/entries/life/. Web address of original article: http://plato.stanford.edu/archives/fall2003/entries/life

Weir A. A. S., Chappell J., & Kacelnik A. 2002. “Shaping of hooks in New Caledonian crows.” Science vol. 297(5583), p. 981 (9 August 2002). Web address of article summary plus online movie: http://www.sciencemag.org/feature/data/crow

Weir, A. A. S. and Kacelnik A. 2006. “A New Caledonian crow *(Corvus moneduloides)* creatively re-designs tools by bending or unbending aluminium strips.” In Animal Cognition 9: 317-334. Web address of draft of full article: http://users.ox.ac.uk/~kgroup/publications/pdf/WeirKacelnik_06_bend_unbend.pdf

Weiss P. and Jeannerod M. 1998. "Getting a Grasp on Coordination." In News in Physiological Sciences, Vol. 13, No. 2, April 1998, pp. 70-75. Web address: http://physiologyonline.physiology.org/cgi/reprint/13/2/70

Wells, M.J. 1962. Brain and Behaviour in Cephalopods. London: Heinemann.

Wepruk, J. 2002. "Pain Management in Farmed Animals." In Livestock Welfare INSIGHTS Issue 3 - September 2002. Web address: http://www.afac.ab.ca/research/species/Articles/painman.htm

White, M. 2000. Lecture: "The Evolution of Sleep." This lecture forms part of Dr White's Psyc 473 course, Sleep and Dreams. Web address: http://psych.fullerton.edu/mwhite/473pdf/473%20Evolution%20of%20Sleep.pdf

Whiten A. and Byrne R. W. 1997. Machiavellian Intelligence II. Cambridge University Press.

Wilcox, R. 2002. "Deceit Amongst The Webbing: Aggressive Mimicry And Cognitive Abilities in Portia Jumping Spiders." Talk given at CISAB (Center for the Integrative Study of Animal Behaviour) at Indiana University, Spring 2002 Graduate Seminar on Learning and Evolution, 1 February 2002. Web address of abstract: http://www.indiana.edu/~animal/speakers/wilcox.html

Wilcox R. S. and Jackson R. 1998. “Cognitive Abilities of Araneophagic Jumping Spiders.” In Balda, Pepperberg and Kamil (eds.), Animal Cognition in Nature: The Convergence of Psychology and Biology in Laboratory and Field. Academic Press: New York, pp. 411-434.

Wilcox R. S. and Jackson R. 2002. “Jumping Spider Tricksters: Deceit, Predation, and Cognition.” In Bekoff M., Allen C., and Burghardt G. (eds). The Cognitive Animal. Cambridge, Massachusetts: MIT Press, pp. 27-33. Web address: http://grimpeur.tamu.edu/~colin/TCA/Ch/Wilcox/wilcox-final.pdf

Williams, R. 2002. "'Bird Brain' Can Be a Compliment." Interview with Dr. Irene Pepperberg on Radio National (The Science Show, 23 February 2002.) Web address: http://www.abc.net.au/rn/science/ss/stories/s487635.htm

Wilson, B., Mackintosh, N. J., & Boakes, R. A. 1985. "Transfer of relational rules in matching and oddity learning by pigeons and corvids." In Quarterly Journal of Experimental Psychology, 37B, 313-332. Cited in Young and Wasserman, 2001 (see below).

Winkielman P. & Berridge K. 2003. "Irrational Wanting and Subrational Liking: How Rudimentary Motivational and Affective Processes Shape Preferences and Choices." In Political Psychology, 24 (4), pp. 657-680. Web address of abstract: http://www.blackwell-synergy.com/links/doi/10.1046/j.1467-9221.2003.00346.x/abs/ Web address of **full article:** http://psy.ucsd.edu/~pwinkiel/winkielman-berridge-political-psychology-2003.pdf

Wittgenstein, L. 1958. Philosophical Investigations. Macmillan: New York.

Wolf, D. M. and Arkin, A. 2003. "Motifs, modules and games in bacteria." In Current Opinion in Microbiology 2003, 6:125-134. Web address: http://people.cs.vt.edu/~murali/teaching/csb/papers/regnets/wolf-arkin-current-opinion-microbio-vol6-motifs-modules-games.pdf

Wolf R. and Heisenberg M. 1991. "Basic organization of operant behavior as revealed in Drosophila flight orientation." In Journal of Comparative Physiology [A], December 1991, vol. 169(6), pp. 699-705. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?holding=npg&cmd=Retrieve&db=PubMed&list_uids=1795235&dopt=Abstract

Wolfram, Stephen. 2002. "A New Kind of Science". Stephen Wolfram, LLC. Web address of online edition: http://www.wolframscience.com/nksonline/toc.html

Wood, D. 1988. "Habituation in Stentor produced by mechanoreceptor channel modification." In Journal of Neuroscience 8: 2254-2258. Web address of abstract: http://www.jneurosci.org/cgi/content/abstract/8/7/2254 Web address of full article: http://www.jneurosci.org/cgi/reprint/8/7/2254

Wood, D. 2003. Home page, University of Pittsburgh. Web address: http://www.pitt.edu/~neurosci/wood.html

Wood, J. 2003. "Observational learning in cephalopods" (a reply to a query by Dr. Richard Katz). On the Cephalopod Page, FAQ, no. 52. Web address: http://www.cephdev.utmb.edu/TCP/faq/TCPfaq2b.cfm?ID=52

Woodward, J. 2003. Article: "Scientific Explanation". In the Stanford Encyclopedia of Philosophy. Current Web address: http://plato.stanford.edu/entries/scientific-explanation/
Original Web address: http://plato.stanford.edu/archives/sum2003/entries/scientific-explanation/

Woodworth, R. S. 1918. Dynamic Psychology. New York: Columbia University Press.

WordNet, Cognitive Science Laboratory, at Princeton University, 2006. Online definition of “sensor.” Web address: http://wordnet.princeton.edu/perl/webwn?s=sensor

Wright, W. 2003. "Attention and Phenomenal Consciousness." Washington University PNP Medical School Seminar Program (December 2003). Web address: http://www.mindstuff.net/AttenPhenomenalConsc.ppt

Wynne C. 2001. “Animal Cognition: The Mental Lives of Animals”. Palgrave: London.

Young G., Ropper A., and Bolton C. 1998. Coma and Impaired Consciousness. New York: McGraw-Hill. Cited in Rose, 2002a.

Young M. E. and Wassermann E. A. 2001. "Stimulus Control in Complex Visual Arrays." In Cook, R. G. (ed.) 2001. Avian Visual Cognition. Tufts University. Web address: http://www.pigeon.psy.tufts.edu/avc/young/default.htm

Young M. E. and Wassermann E. A. 2001. "Stimulus Control in Complex Visual Arrays." Part V: Conclusions. In Cook, R. G. (ed.) 2001. Avian Visual Cognition. Tufts University. Web address: http://www.pigeon.psy.tufts.edu/avc/young/conclusions.html

Zariwala H., Miller A., Faumont S., and Lockery S. 2003. "Step Response Analysis of Thermotaxis in Caenorhabditis elegans." In The Journal of Neuroscience, May 15, 2003. Vol. 23(10), pp. 4369-4377. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=12764126&dopt=Abstract Web address of full article: http://chinook.uoregon.edu/papers/thermoStepjns.pdf

Zhang S. W., Srinivasan M. V. and Collett T. 1995. “Convergent Processing in Honeybee Vision: Multiple Channels for the Recognition of Shape.” In Proceedings of the National Academy of Sciences in the United States of America. March 1995 - Neurobiology. Volume 92, pp. 3029 - 3031. Web address of abstract: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=42352 Web address of full article: http://www.pubmedcentral.gov/picrender.fcgi?artid=42352&blobtype=pdf

Zhang S. W., Bartsch K. and Srinivasan M. V. 1996. “Maze Learning by Honeybees.“ In Neurobiology of Learning and Memory. Volume 66, no. 3, pp. 267 - 282. Web address of abstract: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=8946421&dopt=Abstract Web address of full article: http://www.rsbs.anu.edu.au/Profiles/Shao-Wu_Zhang/documents/1996-Mazelearningbyhoneybees-NLM-Zhang.pdf

Zhang, S. 2002. Subject: The Concepts of 'Sameness' and 'Difference' in an Insect. 30 December 2002. Personal email.

Zhu J. 2003. "Reclaiming volition: An alternative interpretation of Libet's experiment." In Journal of Consciousness Studies. Volume 10, no. 11, pp. 61-77. Web address of abstract: http://www.ingenta.com/isis/searching/Expand/ingenta?pub=infobike://imp/jcs/2003/00000010/00000011/art00004