0

Introduction: What we talk about when we talk about mind

Are we justified in supposing that there is some one type, sort, or kind of thing[…] that is capable of playing approximately the same unifying role in cognitive science that DNA plays in modern biology? If so, then the name it would be plausible to assign to this type is 'mind'.

— David Martel Johnson (1997:4)

It is a fact readily acknowledged yet widely ignored among those with specialist knowledge that, despite half a century of singularly concentrated effort, there is no agreement about what mind or cognition is. This is not merely to say that there is no scientific consensus about what the descriptive details mean, but also that there is no agreement about the more basic issue of identification. Cognition, or mentality, may be the target of the many disciplines that comprise the cognitive sciences, but so far these sciences have proved curiously recalcitrant about providing even a rough characterization of their special subject such that we can say, with a reasonable degree of confidence, that this organism is cognitive and that one is not, in the way that biologists can say that this organism reproduces sexually and that one does not. To ordinary folk, who don't know how science works and don't appreciate how hard this particular nut is to crack, this is rather shocking to hear and hard to believe. Nevertheless, it is the case.

This is also something of a scandal. Easily hundreds of millions, and very probably billions of dollars have been spent on research in the cognitive sciences since the commencement of the 'cognitive revolution' in the mid-20th century. The investment has yielded a phenomenal return in terms of detailed knowledge about the brain and about the behaviour of humans and other animals, although much (quite understandably) is still unknown. The aim of all this effort, needless to say, has not been to arrive at a general theory of cognition that, implicitly or explicitly, would include a characterization. Quite the opposite, in fact. Apparently, most cognitive scientists believe that we can get along quite well without a general theory, or even that a general theory may not be possible, which is why the issue has languished at or near the bottom of the agenda. The upshot is that new empirical findings about the brain and behaviour are being churned out at a truly staggering rate and in fine detail, but there is no framework for making sense of them.

We speak confidently of the mind as the brain—often hyphenated nowadays as 'mind-brain'—but this tells us virtually nothing about what cognition is and what it does. The brain controls almost everything in the body. In the end the putatively 'physical' parts of the brain and the putatively 'mental' parts may not really be all that separable. Well-reasoned ideas abound about cognition: as computation, representation, information processing, or the engine of complex adaptive behaviour. But each of these concepts is equally afflicted with a high degree of ambiguity and lack of consensus about its meaning. Thus the state of the cognitive sciences is at once methodologically highly advanced and theoretically primitive. This is a problem, or so I claim. I am not alone in believing that the theoretical disorder of the cognitive sciences is a problem (de Waal 2002; Silver 1998; Green 2000; Praetorius 2000; Rose 1999; Ross 1997), but it remains a minority opinion.

The aim of this thesis is to make a start in the direction of a general theory of cognition. Each of the five chapters that follow has a very specific job to do in the service of this goal. The first chapter provides evidence for the strong claims I have made in the preceding paragraphs. It also shows how the lack of a general theory and the concomitant absence of an agreed theoretical vocabulary can retard progress, first by rendering data sets incommensurable so that different findings don't 'hang together' well and, second, by prolonging interminable, seemingly intractable metaphysical debates, a feature typical of what Stephen Toulmin (1972) calls "a would-be discipline" and Thomas Kuhn (1970) refers to as "pre-paradigm" science. Metaphysics, it should be remembered, has more than one meaning. Often metaphysics refers to the branch of philosophy concerned with the abstract nature of reality (ontology-with-a-capital-O) or with *a priori* speculations upon questions that are unanswerable to science. This is not the meaning of metaphysics in play here. Metaphysics also refers to a *disciplinary ontology*, namely, the entities, processes and other phenomena that are the special concern of a particular knowledge-generating enterprise (e.g., physics, biology, psychology) or field of study within such an enterprise (e.g., quantum optics, cellular signalling, comparative psychology of imitative behaviour). It is this sort of metaphysics with which the thesis, and Chapter 1 in particular, is concerned.[1]

Clarity regarding the ontology of a discipline is not a trivial matter, although defining the meaning of theoretical terms is sometimes regarded with suspicion (by contemporary philosophers, at least) as merely a matter of stipulation (Sober 1993). Philosophers of science may differ over many things, but they appear to agree that a consensual theoretical vocabulary is indispensable for disciplinary coherence and meaningful progress (Toulmin 1972; Torretti 1999; Popper 2002).[2] A field need not be especially venerable or well advanced to address the issue of ontological clarity. A textbook on *Ontologies for Bioinformatics* (Baclawski and Niu 2005) was recently published to provide a framework—the need for which is described as 'critical'—for a young, rapidly developing field trying to cope with the exponential growth of data by high-throughput technologies (e.g., microarrays, proteomics, metagenomics). The

[1] The two sorts of metaphysics are very easily conflated. I suspect this is one reason progress on the fundamental issues in the cognitive sciences has been excruciatingly slow. Indirect support for this unargued-for proposition can be found in (Haugeland 1997)

[2] On the other hand, David Hull argues that the uncritical adoption by one scientific community of another's terminology can exaggerate the degree of agreement between them and paper over significant differences (Hull 1988:113-117).

authors' stated aim is for precise, 'computer-readable' formulations of concepts, and the relationships between them, which can be used for querying, viewing and transforming data from different disciplines for various purposes. Because an effective ontological vocabulary cannot be contrived in the absence of a theoretical framework in which it does work, Baclawski and Niu survey the "best-known ontologies" in the two main disciplines with which bioinformatics is concerned, biology and medicine.

Figure 0.1 Sciences concerned with cognition*

PSYCHOLOGY	COGNITIVE	COMPUTATIONAL**	ANIMAL	PHYSIOLOGY
1. theoretical	psychology	artificial intelligence	behaviour (individual)	neuroscience
2. experimental	neuroscience	autonomous agents	behaviour (social)	neurobiology
3. developmental	linguistics	modelling	communication	neuroanatomy
4. social	ethology	architectures	behavioural ecology	neurolinguistics
5. clinical	philosophy	signal processing	sensory systems	neuroeconomics
6. psychiatry[α]	anthropology	artificial life	navigation	neuropsychology
7. biological	sociology		culture	neuroendocrinology
8. comparative	archaeology		primatology	
9. evolutionary	ergonomics			
10. ecological	economics[β]			
11. parapsychology				

* It is assumed that the chart is incomplete and considerable overlap exists between categories. For example, 'social cognitive neuroscience' (Gilbert 2002) and 'neuroethoendocrinology' are missing, not least because they are genuinely interdisciplinary and do not fit neatly into any particular category.

** 'Computational' is more or less synonymous with 'cognitive' in the column at left. My intention is to draw a distinction between (cognitive) disciplines and problems or approaches that explicitly fall within computer science. Whether the distinction works is another matter.

[α] Psychiatry, normally distinguished from psychology on the basis of its traditional affiliation with the practice of medicine, is included here for pragmatic purposes.

[β] Perhaps more often referred to as behavioural economics.

Considering the number of fields encompassed by their embrace (see Figure 0.1), a similar effort in the cognitive sciences would be daunting indeed. This is so even were psychology alone considered. The experience of Kleinginna and Kleinginna (1981a; 1981b) in their surveys of how 'emotion' and 'motivation' are used in the psychological literature is instructive. A review through 1980 unearthed no fewer than 102 definitions of motivation, which they classified into nine categories based on the phenomena described or the theoretical issue involved, and 92 definitions of emotion, in eleven categories. Perhaps not surprisingly, an additional category was included in both cases for sceptical arguments regarding the usefulness of the terms. By comparison, the troublesome plurality of the term 'species'—which at a mere 22 definitions has been described as "drowning in a sea of species concepts" (Hull 1999:44)—looks not only tame but almost tractable. Kleinginna and Kleinginna tried to bring some order to this cacophony by proposing their own definitions of emotion and motivation. As far as I can tell, their proposals have had little or no influence. That is because a definition, by itself, is never enough.

Definitions of scientific concepts are verbal, as against formal, techniques for representing the phenomena of a domain of enquiry and thus are inextricably bound to theory (Toulmin 1972). 'Natural selection' means what it does in the context of Darwin's theory of evolution; 'quark' means what it does in the context of the Standard Model of atomic physics. As part of the "empirical content" of a theoretical framework, scientific concepts connect not only to a science's explanatory goals but also to its explanatory procedures and standards of adequacy (Popper 2002:96). The significance of a concept within a historically developing scientific enterprise is thus determinable only within a complex context that includes, according to Toulmin (1972), "subject-matter, formal entailments, [and] explanatory procedures," which are themselves embedded within a larger framework that specifies "on what conditions, in what kinds of case, and with what degree of precision" the concept has meaning at all (p. 185). Chapter 1 shows that it is precisely these fundamental issues that remain currently ambiguous in the sciences concerned with cognition. If anything, cognition is even more vague and polysemic than emotion and motivation, if only because it all too often appears to serve as "little more than a synonym for sophisticated information processing" (Toates 1995:239).

After delineating the shape of the gap, Chapter 1 closes with an outline of the capacities and properties for which a general theory of cognition, ideally, should account. In total, a dozen capacities and properties are identified: affect, anticipation, awareness, discrimination, memory, perception, self-reference, motivation, decision making, error correction, learning, and communication. Collectively these capacities yield two properties—intentionality and normativity—that have been widely discussed by philosophers as suggestive (if not indicative) of mind. Phenomenal consciousness—in Damasio's sense of "the feeling of what happens" or Nagel's what-it-is-like-to-be—is bracketed on the basis of the 'problem of other minds'.[3] As we will see in Chapter 2, liberality or parsimony in attributing phenomenal consciousness to nonhuman animals is arguably a function of usually implicit underlying presuppositions.

If a general theory of cognition is the aim, how to proceed? Chapter 2 identifies two general approaches to the problem of cognition that differ in underlying methodological assumptions. These assumptions relate to the most productive starting point of enquiry and are usually implicit. What I call the *anthropogenic* approach assumes that the best way to approach cognition is to start with the most certain example—human cognition—and abstract the relevant features. Depending upon what is taken to be diagnostic, the resulting theoretical construct may (or may not) apply to other animal species and intelligent machines. The *biogenic* approach, by contrast, assumes that the best way to approach cognition is to recognize that it is, first and foremost, a biological function—it assists an organism to make a living in an ever-changing environment—and should be approached as such. In other words, the biogenic approach assumes that the best way to build a characterization of *what cognition is* and *what cognition does* is to begin with the principles of biology. Psychological questions are posed as biological questions; for example, *what is it that organisms do such that they might* require

[3] The 'problem of other minds' arises from the privacy of human thought and personal experience. We cannot know for certain that another person is conscious except by virtue of her verbal confirmation. Yet, she might be lying. The assumption is, therefore, that similar brain structure—the argument from analogy—guarantees more or less similar experience. But how similar need another brain be? We cannot know in the absence of verbal confirmation, and so on infinitely. To date, the problem has been 'solved' effectively by convention, which many are loath to apply to nonhuman animals.

cognition? The overarching theory of biology is evolutionary theory, so a potentially high degree of continuity is assumed.

For obvious reasons, the anthropogenic approach has predominated historically. The Cartesian approach to mentality is a perfect example of an anthropogenic approach, both in its methodology (introspection) and its presupposition that human cognition is a thing apart from the rest of Nature. Chapter 2 argues that the dominant approach to cognition in the latter half of the 20th century, cognitivism, is an anthropogenic approach. There are many reasons this is so, not least being the well-documented debt of contemporary cognitive science to Descartes (Wheeler 2005; see in particular, Bennett and Hacker 2003). The fact that most cognitive scientists considered the biological basis of cognition quite irrelevant to theorizing about the phenomenon—not just for a little while but for decades—is perhaps the most telling indicator.[4] However, it should be emphasized that the anthropogenic/biogenic distinction relates not to how 'biological' a specific proposal is but to starting assumptions about the problem space. I claim that some of the most difficult problems faced by the computational-representational theory of mind—notably the frame problem and the symbol-grounding problem—are largely the result of this anthropogenic bias. Chapter 2 concludes that it is worth giving a biogenic approach a chance to see if it can do better on the fundamental issues.

Chapter 3 takes a more detailed look at the biogenic approach to cognition, which has its historical roots in the philosophy of Aristotle. Biogenic explanations are currently clustered around three main frameworks for understanding biology: self-organizing complex systems (SOCS), autopoiesis, and biosemiotics. Each framework derives its distinctive character from the aspect of biology it emphasizes. SOCS approaches to cognition emphasize the *physics* of organisms as complex, dynamic, self-organizing and thermodynamically open systems. Autopoietic approaches focus on the *biology* of biology, namely, the distinctive recursive, self-producing organization of vitality.[5] Biosemiotic approaches focus on the *intentionality* of biology, the ubiquity of sign action in living systems, from molecules to whole organisms. From these three frameworks I infer 13 empirical principles, the biogenic 'family traits', which necessarily constrain biogenic theorizing. Because the anthropogenic approach to cognition is not empirically constrained to nearly the same degree, I argue that the biogenic approach is superior for approaching the what-it-is and what-it-does questions relating to natural cognition.

Biogenic approaches should not be conflated with Embodied Cognition (EC), the philosophical movement associated with the belated recognition that knowledge-generating mechanisms (such as we know them) are embodied in an organism (human or otherwise) that is situated in an environment with which it must continually contend and which, in doing so, it also modifies (Varela, Thompson, and Rosch 1991; Clark 1997a). EC holds that cognition cannot be explained adequately if divorced from consideration of the environment with which an organism is coupled. It also recognizes that the mechanisms of cognition have evolved to assist an organism in its interaction with an environment (Dreyfus 1972; Godfrey-Smith 1996; Gibson

---

[4] This neglect of biology often is attributed to the influence upon Cognitive Science of analytic philosophy. (See in particular, (Johnson 1987; Varela, Thompson, and Rosch 1991; Dupuy 2000)) As Mark Johnson observed, "Anglo-American analytic philosophy…mistakenly assumed that only a viewpoint that transcends human embodiment, cultural embeddedness...and location within historically evolving traditions can guarantee the possibility of objectivity" (Johnson 1987:175). It also had to do, I believe, with one of the traditional aims of Cognitive Science, to establish a theoretical framework general enough to cover both organisms and intelligent machines. Searle (1980) advanced his 'biological naturalism' in part as a corrective to both of these tendencies.

[5] Artificial Life is not counted separately because research conducted under this rubric usually falls within SOCS, autopoiesis or biosemiotics.

1979; Dreyfus, Dreyfus, and Athanasiou 1986; Beer 1990). By acknowledging that the only cognitive mechanisms about which we are certain are those that have evolved in biological systems constantly engaged in body-world interactions, EC exhorts us to take prevailing biological knowledge and evolutionary theory seriously. But biology can be taken seriously in all sorts of ways. One need not *begin* with the principles of biology to ensure at a minimum that one's theory of mind does not *contravene* those principles.

Whereas a biogenic approach to cognition is intrinsically embodied, an embodied approach need not be biogenic. Assuming that beliefs and desires are the *sine qua non* of cognition and then building a plausible biological case for how they function in the economy of an animal, and how they might have evolved, may be a thoroughly embodied approach, but it is not biogenic. Beliefs and desires derive their privileged status in cognitive science from folk psychology, and 'the folk' are necessarily human. It is quite another thing to derive a picture of what cognition is and what it does from the principles of biology and then see how beliefs and desires (such as we experience them) arise from that matrix.

Simply saying, "Let's look over there," is not very useful without an example of what we might find if we actually *did* look over there. Chapter 4 provides an example of a biogenic framework for investigating cognition, the *agentive framework*. The purpose of the exercise is not to produce the theory we've all been waiting for (although if it helps anyone in thinking about their own work, the effort will have been worthwhile), but, rather, to walk through the logic of a possible hypothesis, to show how biological principles can be enlisted to yield a characterization of cognition that not only captures the necessary features of the phenomenon but also connects to the theoretical structure of the larger knowledge enterprises in which natural cognition is ultimately embedded, namely, biology and psychology. Theories in psychology often do not connect with the larger knowledge enterprise (Mesoudi, Whiten, and Laland in press), sometimes defiantly in the name of disciplinary autonomy (Fodor 1968) but more often because "psychologists often ignore work outside their own sub-specialities, and almost always ignore work outside their own discipline" (Gilbert 2002:3). Although the search for a definitive 'mark of the mental' is regarded as misguided by some, at least within psychology (Kimble 1994), the agentive framework (as its name implies) suggests that *agency*, defined in a particular way, is a less problematic candidate than what has been offered to date. From a series of biological premises the agentive framework yields a concept of cognition that accommodates the capacities outlined in Chapter 1, applies to a number of different data sets in multiple disciplines and makes a number of empirical predictions. It is, therefore, testable and subject to falsification, features considered necessary for a theoretical construct to pull its weight in a scientific ontology (Popper 2002; Hull 1999).

The agentive framework has uncomfortable consequences, however. The main consequence is that all organisms, not just some animals, are cognitive. This includes, counter-intuitively, bacteria. Cognition, on this story, turns out to be a necessary *systemic* function of how biological creatures make a living, in the same way that respiration, ingestion, digestion and elimination are necessary systemic functions. To do what living creatures do, I argue, they *need* cognition; it is not an optional extra found in the lucky few. To demonstrate that the claim has substance and is not vacuous, Chapter 5 examines in detail several types of behaviour exhibited by organisms at the putative taproot of the phylogenetic tree: bacteria. Drawing on the latest research in microbiology, a discipline undergoing revolutionary change due to advances in imaging technology, I show how three classic arguments against bacterial cognition—that it is not flexible, complex or distal enough—are no longer sustainable.

In brief, the bacterial sensorium turns out to be far more diverse than traditionally believed, with some species[6] having different signal transduction systems for many dozens of environmental states of affairs as well as internal physical states (Ulrich, Koonin, and Zhulin 2005). Bacterial chemotaxis thus is not pinball-like but, rather, results from the integration of multiple cues, both internal and external to the organism (Koshland 1980). Bacteria are not normally clonal loners but live mostly social lives, and for the same reasons that other social species do: there is safety and greater access to resources in consortium (Shapiro 1997). Bacteria have complex systems of chemical communication, both species-specific and species-general, that facilitate a variety of social behaviours (Miller and Bassler 2001), including life in multicultural colonies that have been compared (non-facetiously) to cities, due to divisions of labour and real estate (Watnick and Kolter 2000). Bacteria are so social, in fact, that they are now being considered as possible models of social evolution (Queller 2004). The picture that is emerging is of a class of organisms that may well be more sensitive and *adaptable* in the face of environmental change—not merely well-adapted to an existing niche—than are so-called 'higher' vertebrates (Shapiro and Dworkin 1997), including humans. In short, there is now compelling evidence that most—very possibly all—of the cognitive capacities outlined in Chapter 1 can be found in bacteria, albeit in a far simpler and more circumscribed form than that found in humans.

The good news, in my view, is that a liberal account such as the agentive framework is precisely what Darwinian theory leads us to expect, that mental capacities evolved in the same way that physical capacities have. Fins and limbs may have made certain sophisticated types of motility possible, but motility as a biological function was already a going concern before fins and limbs arrived on the scene. According to the agentive framework, human cognitive capacities likewise can be seen as highly sophisticated elaborations on a basic *Bauplan* of capacities that were already promoting the survival, flourishing and reproduction of long-distant evolutionary ancestors before the advent of the brain or even a nervous system.

The broadening of perspective has substantial advantages. Two will suffice for now. First, it aligns the cognitive sciences with the natural sciences rather better than they are at present, and provides a vast new source of data to tap for the study of cognition, most importantly the literature on (social and non-social) cellular signalling, the ground-floor of biological information processing. Second, it means that we can expect to understand many principles of our own cognitive functioning by studying the cognitive capacities of vastly simpler biological systems. While primate studies are clearly necessary for potential insights into peculiarly human cognitive styles, simpler creatures can be used to understand the logic of the function of cognition more generally. Needless to say, this is exactly how most scientific understanding of physiology, genetics and molecular biology has been achieved. The bad news, of course, is that thinking of cells as cognitive provides another counter-intuitive challenge to our ordinary ways of conceiving the world, which appears to be an unfortunate side effect of much scientific endeavour—or of any deep, systematic scrutiny, for that matter.

Background assumptions

Before we proceed, let me briefly situate the project with respect to a few technical issues. First, I view the thesis as an exercise in natural philosophy, by which I mean the rather old-

[6] The concept of species in bacteriology is even more fraught than it is in biology generally. I will strive to use the terms 'variants' or 'isolates' as substitutes for bacterial 'species' but sometimes—as here—these terms don't fit well and create unnecessary ambiguity.

fashioned sort of "philosophy about nature" that originated with Aristotle and was practised by Descartes, Leibniz, Newton, Spinoza, Locke, Kant and, more recently, by James, Pierce, Dewey, and Popper. (Luckily, genius is not required to do natural philosophy, only to be remembered for having done it.) Gary Hatfield (2002) notes that the canonical natural sciences (e.g., astronomy, physics, chemistry, optics, zoology, biology, psychology) gestated in the "old science" of natural philosophy, which provided the conceptual background against which a discipline could develop into a "new science", namely, an autonomous, empirical knowledge enterprise using up-to-date experimental methods and mathematical formalisms (p.212). Natural philosophy is philosophical "in that there [is] discussion of the basic classification of natural things into kinds, characterization of the properties of those kinds, and an explanatory framework involving those kinds and their relations" (ibid:210). In other words, the natural philosophers saw themselves as part of a broad knowledge enterprise aimed at understanding the world and how it works, and they drew upon the latest scientific evidence available to them, in addition to the evidence provided by their own rumination. What they were doing was partly philosophy and partly science, but, given a contemporary Anglophone reading of the special role of philosophy,[7] they might well have characterized their work as more the latter than the former. On the whole their accounts, however speculative, were positive rather than critical. The thesis is also empirical because, to the extent possible, the arguments rest on evidence derived from scientific observation and experiment, rather than on traditional conceptual analysis.

Second, this account is principally positive rather than critical; it aims, perhaps immodestly, to contribute to science as much as to philosophy. Although the various sciences have parted ways from their philosophical parent over the past few hundred years (the last, arguably, being psychology), there is still work that ordinary philosophers can do in the service of science. I agree with Kim Sterelny (2003) that one of the things philosophers do well, which can profit science, is to synthesize data from a wide variety of sources. I have tried to do that here. The discussions I have had with scientists during the course of my research suggest this is a job that needs doing in many areas. Whether or not I have done the job well, in this particular domain, is for others to judge.

Third, there will be much talk of 'function' in the chapters that follow. What we know incontrovertibly about natural cognition is that it does something to support the continued existence, wellbeing and reproduction of an organism. In the case of other biological functions, like respiration and digestion, *what it is* has everything to do with *what it does*. The same, presumably, is true of cognition. This in no way implies that cognition is capable of a univocal, T-shirt-worthy analysis, any more than is any other complex, systemic biological function. Equally, it is not to foreclose the possibility. The concept 'function' does not itself have a singular meaning in biology, to say nothing of the profusion of meanings in philosophy (Wright 1973; Griffiths 1998; Neander 1998; Nagel 1977; Bigelow and Pargetter 1987; Enç and Adams 1998; Cummins 1975; Godfrey-Smith 1998; Millikan 1989a; Sarkar 2005; Hardcastle 1999a). Here we will stick to the biological meanings.

---

[7] It is sometimes observed that in the 20th century a consensus formed within the Anglo-analytic tradition around the unique role of philosophy as a knowledge enterprise, which is to uncover hidden assumptions and reveal the meaning of concepts (Ragland and Heidt 2001). While the need to comprehend underlying assumptions and concepts is important to both philosophy and science, on this view the excavation of meaning is philosophy's special function. The natural philosophers, I think it is safe to say, did not view their primary labour as the excavation of meaning but, rather, a positive attempt to advance knowledge about the world that did not previously exist—at least not in the form in which they offered it. Clarifying meaning was a necessary adjunct to the positive program, but not an end in itself.

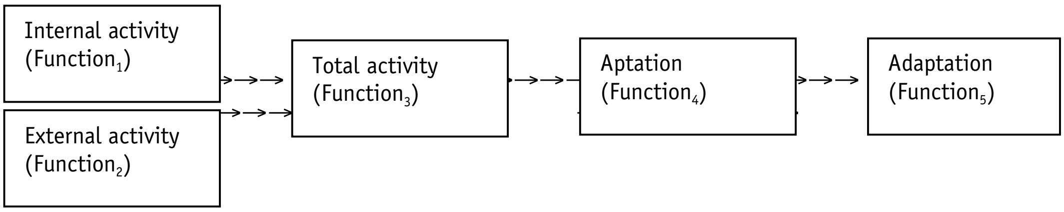
Mahner and Bunge (2001) identify five meanings of function in biology, each of which is appropriate to a particular explanatory context. None is mutually exclusive of the others. Indeed, they "are logically related, namely, by the relation of implication" (p.78), one of the advantages of the classification. The first three meanings in the Mahner-Bunge taxonomy refer simply to 'biotic' or 'biologically relevant' activity, which is distinguished from, say, the quantum mechanical activity of sub-atomic entities or processes (p. 78). *Function*$_1$ refers to biotic activity *internal* to a subsystem of an organism, for example, the biochemical processes that bring about the rhythmic contraction of the heart. *Function*$_2$ refers to biotic activity *external* to a subsystem. For example, the rhythmic contraction of the heart muscle pumps blood. *Function*$_3$ refers to *internal* cum *external activity*, or internal and external activities whose roles are interdependent. The role of the rhythmically contracting heart as a blood pump, which facilitates the circulation of oxygen and other nutrients, is an example of function$_3$.

Functions$_{1\text{-}3}$ are sometimes referred to as 'effects' because, as Mahner and Bunge put it, they "entail nothing as to the value or usefulness of the corresponding activities" (p.78). *Function*$_{4\text{-}5}$ refer to the value of the subsystem to the organism as a whole, which can be considered without reference to evolution (*function*$_4$) or explicitly in terms of natural selection (*function*$_5$). *Function*$_4$ relates to the current value (positive, negative, neutral) of a subsystem to the organism. An activity or structure is a *function*$_4$ if it contributes to an organism's wellbeing, persistence or reproduction (e.g., a heart). Following Gould and Vrba (1982), Mahner and Bunge refer to a fitness-enhancing *function*$_4$ as an *aptation*.[8] An activity or structure that reduces an organism's wellbeing or impairs its ability to persist and reproduce (e.g., a tumour) is a malaptation (malfunction$_4$, dysfunction); while one that has no effect (e.g., the appendix) is a nullaptation.

Aptation does not necessarily entail anything about the existence of the trait, however. The support of spectacles may be among the various *functions*$_4$ of the human nose, but that is not why the nose exists. Aptations that are "retained or improved on by [natural] selection" are *adaptations*, which comprise the final category, *function*$_5$. While all adaptations are aptations, the reverse is not true (Gould and Vrba 1982; Gould and Lewontin 1979). Malaptations cannot be adaptations because, "[b]y definition, there can be no features that are favoured by selection for being disvaluable" (Mahner and Bunge 2001:78). However, a nullaptation ('spandrel' in Gould and Lewontin's terminology) may be retained in subsequent generations and may prove adaptive in another setting. The taxonomy is set out in Figure 0.2 below.

[8] According to Gould and Vrba, a function is an *exaptation* if its current (positive) contributory role is different from the one it had under different selective pressures at an earlier stage of evolution. The phenotype is coopted for a new use as a result of changes in environmental conditions and selective pressure.

Figure 0.2 Mahner-Bunge taxonomy of biological 'function' concepts



The figure shows the relations among the five concepts of function in biology. The figure can be read from left to right (e.g., *Funtion*$_4$ "is needed to define" *Funtion*$_5$), or else from right to left (e.g, *Funtion*$_5$ "implies" *Funtion*$_4$). (Reproduced with permission.)

While Mahner and Bunge's is not among the most widely cited accounts of function in the philosophy of science literature, their taxonomy is, in my view, the clearest account and, as far as I can tell, the most consistent with biological practice. The taxonomy also provides a useful framework for regimenting the multitude of presentations in the philosophical literature. For example, Valerie Gray Hardcastle's 'pragmatic approach' to understanding functions focuses on biotic activity relative to an explanatory goal—"Functions are simply what *T* is doing in *O*, relative to a domain of inquiry" (Hardcastle 1999a:41)—and thus encompasses *functions*$_{1\text{-}3}$. Value to the organism is not an issue *per se* on Hardcastle's account. Wright's seminal presentation (Wright 1973), which inspired the 'teleological' approach to function popularised by Millikan (1989a), describes functions as biotic processes or structures that have been selected in the course of evolution for their contribution to the system of which they are a part. The value dimension coupled with natural selection places Wright functions squarely in the purview of adaptation and *function*$_5$. Cummins' classic alternative to Wright, causal role functions (Cummins 1975), construes function in terms of the contribution a process or structure makes to a systemic competence but without reference to selection history (etiology), making it roughly equivalent to *function*$_{3\text{-}4}$.

Sahotra Sarkar (2005), on the other hand, distinguishes between two types of function associated with biology: "broad-sense functions" and "narrow-sense functions" (pp. 17-19). *Broad-sense functions* answer *how*-questions and relate to what trait *T* does and how it does it. The critical explanatory principle for broad-sense functions is organism persistence. How do the mechanisms that constitute functional trait *T* contribute to the persistence of organism *O*? How do the networks of structural elements and chemical processes (*T*) operate to enable *O* to carry out the activities that enable its survival, wellbeing and reproduction? As Sarkar defines it, "an effect of some structure, *A*, is a function if it contributes to the persistence of some system, *B*, of which *A* is a part" (Sarkar 2005:18). Broad-sense functions are thus similar to *function*$_4$, which implies *functions*$_{1\text{-}3}$. Sarkar claims (correctly, in my view) that when biologists talk about functions, it is to broad-sense function that they advert. *Narrow-sense functions*, on the other hand, are those effects of a biological feature (e.g., a liver, a red blood cell) that contribute to an organism's fitness—namely, its ability to survive and reproduce—and have been maintained over generations because of this contribution. Narrow-sense functional explanations relate primarily to *why* the trait is maintained, and natural selection is the critical explanatory principle.[9] Narrow-sense functions are thus similar to *function*$_5$. Sarkar claims that fitness-enhancing effects are generally what philosophers of biology mean when they talk about functions.

Finally, Enç and Adams (1992) distinguish between the dispositional and etiological aspects of 'function'. The *dispositional*, or forward-looking, aspect of function indicates that a trait or character attributed with this property has a propensity for producing certain types of activity that has certain types of consequences for the system of which it is a part. The key in a functional account of this kind, which equates with *function*$_4$ (and resonates with Bigelow and Pargetter (1987)), is the specification of the particular activity that displays the propensity. The *etiological*, or backward-looking, aspect of function, according to Enç and Adams, relates to the propensity's causal history—by natural selection in the case of biological functions—and equates to *function*$_5$.

All five types of function in the Mahner-Bunge taxonomy will figure in the following chapters. However, cognition as *function*$_4$—what cognition contributes to the persistence of an organism—is the main quarry. For the purposes of this discussion I will adopt the following definition of *function*$_4$: *Feature (subsystem) T has a function in organism (system) O if its activity contributes to the persistence of O, where T is a structure*[10] *or network(s) of processes*. An example of how this concept of function figures in biological explanation is as follows: "It is well established that the dentate gyrus of the hippocampus is a site of neurogenesis in the adult mammal, but experimental and theoretical explorations of the *functional significance* of this process are just beginning to emerge" (Meltzer, Yabaluri, and Deisseroth 2005:653)(my emphasis). Here functional significance refers to "cognitive utility" (p.655), namely, how the generation of new neurons in the hippocampus contributes to that brain structure's well-established role in memory and mood (*function*$_3$), which are themselves functions that contribute to the persistence of the organism in which they feature (*function*$_4$).

Although the general approach to cognition adopted here is 'functional', it cannot be emphasized too strongly that this does not entail a commitment to a particular brand of functional*ism*. Functionalism, as an interpretive strategy, comes in many flavours—Thomas Polger (2004) claims to have distinguished over 100 variations—and is found in a variety of disciplines within the humanities and social sciences, including philosophy, psychology, anthropology, sociology, hermeneutics, and linguistics.[11] *Philosophical functionalism*, which has been central to much theorizing in Cognitive Science and philosophy of mind over the past half century, is an explicitly metaphysical thesis about the conditions by which a mental state is identified in a cognitive system.[12] The core metaphysical commitment is that a mental state is none other than the role (function) it occupies, which can be specified purely in terms of a set of abstract relations linking stimulus conditions with a behavioural response (Putnam 1975; Wilson 2005; Polger 2004). Cognitive states thus are distinguished by the role they play in

---

[9] In the case of etiological accounts, adaptive utillity is why the trait was selected in the past; in the propensity account, it is the reason the trait will be selected in future.

[10] Examples of structures include limbs, organs, tissues, and types of cells, as well as discernible parts thereof.

[11] Of these, I am most sympathetic to the biologically inspired functionalism of pragmatist psychologists such as William James and the so-called 'Chicago School' headed by John Dewey and James Rowland Angell. These functionalist psychologists saw their approach to psychology as analogous to physiology, in contrast to structuralism (e.g., advocated by Titchener), which they saw as analogous to anatomy.

[12] Polger identifies six types of functionalism employed in philosophy: intentional, semantic, theoretical, explanatory, and methodological, in addition to metaphysical functionalism. However, metaphysical functionalism is the dominant variety and all six types involve the abstract relation of input to output and other mental states. (See Chapter 3 of Polger 2004.)

transforming the inputs of a stimulus, which may be quite complex, into the outputs of a certain (often unspecified) kind of behaviour, either relatively directly or via interaction with other mental states.

Theoretically at least, the set of relations believed to constitute the function (cognition, mentality) can be instantiated in any sort of physical system, natural or artefactual, not just in a brain or nervous system. The function thus is said to be 'multiply realizable', capable of instantiation in different physical media. In short, philosophical functionalism holds that mechanisms (physiological or otherwise) are not what make cognitive states cognitive; what makes cognitive states cognitive is, rather, an (unspecified) abstract set of relations that constitutes the functional role relative to the generation of (unspecified) types of behaviour. Mechanisms, according to philosophical functionalism, are implementation details and need not figure in the elucidation of a function. In biology, by contrast, mechanisms are the main game in just such an enterprise. The status of functionalism within the philosophy of mind, once unassailable, is under growing challenge (Foss 1995; Putnam 1997; Shapiro 2000; Polger 2004; Churchland 2005), not least for its lack of connection with empirical reality. This is as good a reason as any to withhold assent here.

Talk of function invariably leads to talk of teleology, the purposes or goals that an activity or structure fulfils or subserves. Goal-directedness and normativity, both of which have many meanings, occupy a region of philosophy, both of mind and science, that has long stirred controversy and remains highly fraught (see, for example, Davies 2001). I will have more to say about this issue in Chapter 3. For now I will merely advert to three lines of argument that will figure in that discussion. First, even a cursory glance at the technical literature in biology, molecular biology and genetics/genomics demonstrates biologists routinely use teleological language (e.g., purpose, goal, 'for'), which suggests the concept of goal-directedness is explanatorily useful. While I do not mean to suggest that anyone disputes the heuristic utility of such language, it is important to keep this in mind. Second, one of the signatures of biological phenomena—what sets them apart from the phenomena studied by physics and chemistry—is their *qualitative* or relational nature (Mayr 1982); their highly organized, 'oriented', 'coherent', and 'constructive' activity depends on information about other states of affairs (Monod 1972:45). This inescapable, so-far-irreducible aspect of biological processes prompted the introduction of the term *teleonomic*, as an alternative to the philosophically loaded notion of teleology, by C.S. Pittendrigh (1958). Finally, the requirements of self-maintenance and reproduction, combined with a history of evolutionary selection, determine that each type of organism responds to some aspects of its environment but not to others. As a result some environmental features are salient, or have value to the organism while others (indeed, vastly more) do not.

Finally, I should be clear about what I will not do: I will not attempt a philosophical analysis of the natural-kind status of cognition. Two factors, in addition to radical polysemy, suggest this might be an unproductive strategy. First is growing philosophical scepticism about the utility of traditional natural-kind analyses of scientific concepts generally (Dupré 1981; Hull 1965; Beer 1995; Griffiths 1999), but psychological concepts in particular (Charland 2005).[13] Kornblith (1999) summarizes the dilemma thus: "That scientific taxonomies are simply more

[13] Charland (2005) argues that the philosophical debate over the natural kind status of emotion, for example, "appears to have reached an impasse" (p.83). While some may believe this vindicates the notion that emotion is not a natural kind, Charland claims a declaration of victory would be pyrrhic, "because it is a 'victory' that has somehow—incredibly—managed to steer totally clear of the relevant issues in emotion science" (ibid).

messy than was once assumed has thus not only complicated the picture of natural kinds, but also made some doubt the very usefulness of the notion." (p.589) As a result, the concept of 'natural kind' is itself the subject of active, ongoing philosophical revision from its Aristotelian and Lockean roots (Boyd 1991; R. Wilson 1999; Millikan 2000; Griffiths 1997). Second, asking whether mind is a scientific kind "invites a consideration of just about every major problem in the Philosophy of Science and several in Philosophy of Mind" (Clark 1996:17). Paul Griffiths' excellent *What Emotions Really Are* (Griffiths 1997), a detailed exploration of the natural-kind status of emotion, suggests that nothing less than a monograph would be adequate to pose, much less answer, such a question. Once that question is answered, however, an account of *what cognition is* and *what cognition does* is still required.

Natural kinds, stripped to fighting weight, are categories. One of the major functions of categorization—the human kind, at any rate—"is to allow extrapolation from observed to unobserved instances" (Griffiths 1997:187). Scientific discovery, which presupposes regularities in Nature, rests on the ability to reliably extrapolate from samples of a category to a whole category (Griffiths 2004). Pylyshyn (1999) may be right that some (many?) of the concepts grounded in commonsense psychology that are currently used in the cognitive sciences will be abandoned as the sciences develop. I am betting that cognition is not one of them.

Now, to work.