In A. Brenner & J. Gayon, eds., *French Studies in the Philosophy of Science: Contemporary Research in France,* Springer, 2009. pp. 255-300

PHILOSOPHY OF COGNITIVE SCIENCE

Daniel Andler

The rise of cognitive science in the last half-century has been accompanied by a considerable amount of philosophical activity. No other area within analytic philosophy in the second half of that period has attracted more attention or produced more publications. Philosophical work relevant to cognitive science has become a sprawling field (extending beyond analytic philosophy) which no one can fully master, although some try and keep abreast of the philosophical literature and of the essential scientific developments. Due to the particular nature of its subject, it touches on a multitude of distinct special branches in philosophy and in science. It has also become quite a difficult, complicated and technical field, to the point of being nearly impenetrable for philosophers or scientists coming from other fields or traditions. Finally, it is contentious: Cognitive science is far from having reached stability, it is still widely regarded with suspicion, philosophers working within its confine have sharp disagreements amongst themselves, and philosophers standing outside, especially (but not only) of non-analytic persuasion, are often inclined to see both cognitive science and its accompanying philosophy as more or less confused or even deeply flawed.

The sensible way to go under the circumstances, or so one might judge, would be to pick a sample of salient topics, in the present case, philosophical discussions of some central foundational issues, in the hope thereby of giving the reader a sense of what the field is about. This however is not the path I propose to take. There are two reasons for choosing another tack. The negative reason is that there is now available a plethora of excellent expositions, of any length one might desire, from one-page summaries to chapter- or volume-length introductions, of central topics in philosophy of mind (which constitutes in turn the core of what most philosophers think of as philosophy of cognitive science: more on this in a moment)[1]. Producing one more such exposition seems hardly worth the effort. The positive reason is that philosophy of science in general has a number of goals not all of which consist in elucidating foundational issues; for example, there are issues of methodology; there are conceptual problems linked to empirical issues which seem not yet ripe for direct scientific

[1] Including abundant web resources: Stanford Encyclopedia of Philosophy - http://plato.stanford.edu/; Dictionary of Philosophy of Mind - http://philosophy.uwaterloo.ca/MindDict/; Internet Encyclopedia of Philosophy -- http://www.iep.utm.edu/; Field Guide in the Philosophy of Mind (Nani, M. & Marraffa, M., eds.) - http://host.uniroma3.it/progetti/kant/field; / Chalmers, D., Contemporary Philosophy of Mind: An Annotated Bibliography - http://consc.net/biblio.html

resolution by available means. But there is also the more general concern of providing a perspective on the structure and dynamics of a field, its relations to other areas of inquiry, its purported limitations or misconceptions, its future directions. This applies to science as a whole as it does to the specific fields and disciplines, however broad or narrow. And it would seem to apply all the more to fields which have emerged recently, and which therefore give rise to questions about where they fit in the overall scheme of inquiry, why they didn't appear sooner and whether they are here to last. In short, philosophy of cognitive science can also, and perhaps should, be thought of as a division within the philosophy of science, on par with philosophy of biology, philosophy of economics, etc. This may seem obvious, but it is not how it is usually treated.

This at any rate is how the present chapter proposes to view the topic. In broadening its scope, it will necessarily sacrifice depth, but will still stop short of providing answers to all the concerns just listed. I will begin with a discussion of metatheoretical issues, which will prepare us for the sequel. I will be not attempt to draw the contours of cognitive science. This would go without saying in the case of any other field: who would expect a 'definition' of mathematics in a chapter devoted to contemporary philosophy of mathematics? Cognitive science is different: its image is blurred. It will be less so, one may hope, by the time the chapter closes. Meanwhile however, it might be helpful to some readers to have a working definition. Let's content ourselves with the following: Cognitive science corrals a variety of disciplines and approaches with the aim of providing an integrated scientific account of the mind, its states, processes and functions. If 'mind' is thought to command a premature commitment to a dubious ontology, one can for the time being substitute 'behavior', withholding any strong preconception about what counts as behavior and what may or may not enter in the sought-after accounts. Finally, the reader who has trouble seeing what distinguishes cognitive science thus characterized from (scientific) psychology is asked to only accept the following amendment: cognitive science is, as it were, psychology pursued by novel means; it draws on any potentially relevant discipline (the main contenders being neuroscience, computer science and related modeling techniques from physics and mathematics, linguistics, philosophy and parts of social science), and its detailed agenda is thereby broadened far beyond the ones pursued in previous epochs of psychology.

1. METATHEORETICAL ISSUES

a. The French Dimension

Philosophers in France working on cognitive science are for the most part wedded to an internationalist view of scholarship. For them, the very idea of a "French cognitive science" or a "French philosophy of cognitive science" is reminiscent of sorry episodes in the history of science (German physics, Soviet genetics,...) and, more to the point, of a recent period where French academia, especially in the humanities and sciences of man, was isolated from the international community and entered a phase of parochialism.

However commendable, these internationalist convictions should not lead one to take it as a necessary truth that there is nothing to be said about cognitive science, or about philosophy of cognitive science *in* France. It certainly makes sense that there would exist local schools in those areas, as there are in every enduring academic field[2]. It is neither absurd, nor a political fault, to inquire whether there are, or have been, specific traditions which originated or were developed in France. The answer turns out to be mixed, and cannot be expounded in any detail here[3]. As for cognitive science proper, the prominent American psychologist David Premack was fond of saying that it was 'invented' by the French Nobel laureate Jacques Monod, who founded the Centre Royaumont pour une science de l'homme, a think-tank *cum* conference center which hosted in 1975 a memorable encounter between Noam Chomsky and Jean Piaget[4]. Even interpreted with a grain of salt, this view can be taken as a conclusive refutation of the popular conception of cognitive science as a US import[5]. France had strong traditions in the neurosciences, in linguistics, in anthropology, in mathematics, in cybernetics. There were also a small number of scientifically-minded psychologists spread over France and neighboring French-speaking countries[6]. These resources could be pooled to constitute small informal groups, within which a culture emerged which we can retrospectively identify as cognitive-scientific, and on which the present generation was able to build in its successful efforts to set up cognitive science in the contemporary French academic scene.

As for the philosophy of cognitive science, its fate has been closely connected to the development of analytic philosophy, which was all but barred from France after World War Two. Only in the late 1960s were a small number of young philosophers able to cultivate it, and it took another 20 years or so for analytic philosophy to gain acceptance. On that front, then, there is no denying that the now well-established group of analytic philosophers of cognitive science working in France, some of whom enjoy an international reputation, are true cosmopolitans. But next to them, there are a small number of philosophers who lean on Kantianism or on phenomenology (mostly in Husserl's and Merleau-Ponty's traditions) to approach cognition from an angle unfamiliar to most analytic philosophers. The mathematician René Thom's visionary ideas in natural philosophy and in theoretical biology have also had some influence, and so have some successful non-technical books written by prominent neuroscientists[7]. These strands are more clearly French, but limitations of space would force me to give their ideas short shrift, or else not present the international context in which they are deployed, something they would disapprove of for the reasons stated above.[8]

The upshot is that for the most part this chapter will dwell on the philosophy of cognitive science *simpliciter*, without an eye on a specifically French source or style.

b. Philosophy of/and/as/in *Cognitive Science: A Logical Geography*

[2] In a highly enlightening and influential essay entitled "The logical geography of computational approaches: a view from the east pole" (1986), Dennett contrasted two radically different conceptions of cognitive science, one cultivated at and near MIT, on the east coast of the United States, the other in California.
[3] See Andler, 2006a.
[4] See Piatelli-Palmarini, 1979.
[5] Even if one were to discount the work done in Continental Europe (which is not acceptable even as an idealization), the UK was very strong from the start. This would make cognitive science and the associated subfield of philosophy a US-UK import on the Continent. Although mistaken, that view is less so than the 'all US' theory.
[6] See Parot et Richelle, 1992; Berthoz 1999.
[7] Changeux, 1983, 2003; Berthoz, 1997; Jeannerod, 1997; Imbert, 2006.
[8] Thom, 1988; Petitot *et al.,* 1999, Some of these ideas started converging in the 1990s with the connectionist and dynamicist views developed in the US and elsewhere (see below).

Up to now I have been using 'philosophy of cognitive science' as the most general term embracing all forms of philosophical activity connected in one or another way to cognitive science, but I indicated that there exist under that general rubric some rather different forms of inquiry. Indeed, there is no agreement on names for these various forms, nor on the significance of the differences. For some, there is no reason to draw sharp distinctions, or any distinctions at all, between philosophy of psychology, philosophical psychology, philosophy of mind, philosophy of cognitive science or cognitive philosophy[9]. But the fact that there are indeed no crisp boundaries, that some issues can be seen as taking part in different projects, and that there exist multiple connections between topics belonging to different areas, does not imply that distinctions cannot be usefully drawn. In fact, I claim that such distinctions are an integral part of the preliminary agenda of the philosophy of cognitive science in the wide sense, one which hasn't received sufficient attention.

There are, as I see it, two dimensions of contrast to consider. Along the first axis, one can plot proximity to (cognitive) science. Proximity involves either collaboration, or sympathy, or both. Near one end of the line, one finds research programs in which philosophers and scientists from one or another discipline, or sometimes several, attempt to provide a solution to some specific problem concerning a cognitive phenomenon. This effort, when successful, leads to a scientific achievement which both owes to, and rewards, the philosophical investment which went into it. This is the sense in which philosophy is one of the basic components of the cognitive federation: the relation is one of inclusion, philosophy *in* cognitive science (or again, philosophy *as* cognitive science). Examples abound in such areas of study as visual perception, reasoning, linguistic communication, numerical knowledge, voluntary movement, 'mind-reading', social skills, to pick just a few examples among dozens. Such shoulder-to-shoulder activity makes sense only at the cost of renouncing the traditional view of the philosopher as a respectful witness, an expositor or again an appraiser, of science. It also implies the maximal degree of sympathy, *viz.* direct involvement.

Near the other end of the line, one finds the standard situation where philosophy, allied with history, examines and appraises cognitive science as an enterprise situated at some distance, somewhat like philosophy of art stands (for the most part) outside art, or general philosophy of science stands outside science (by necessity, as there is no such thing as 'general science'), or philosophy of chemistry, in all but exceptional cases, stands outside chemistry. Philosophers who operate in such a framework typically do not attempt to directly contribute to the enterprise they are appraising. Sympathy may be present, but with a certain professional distance; and it may also be altogether absent, when the philosopher finds the science (or the art form) she is examining flawed and develops an (informed) critique. This

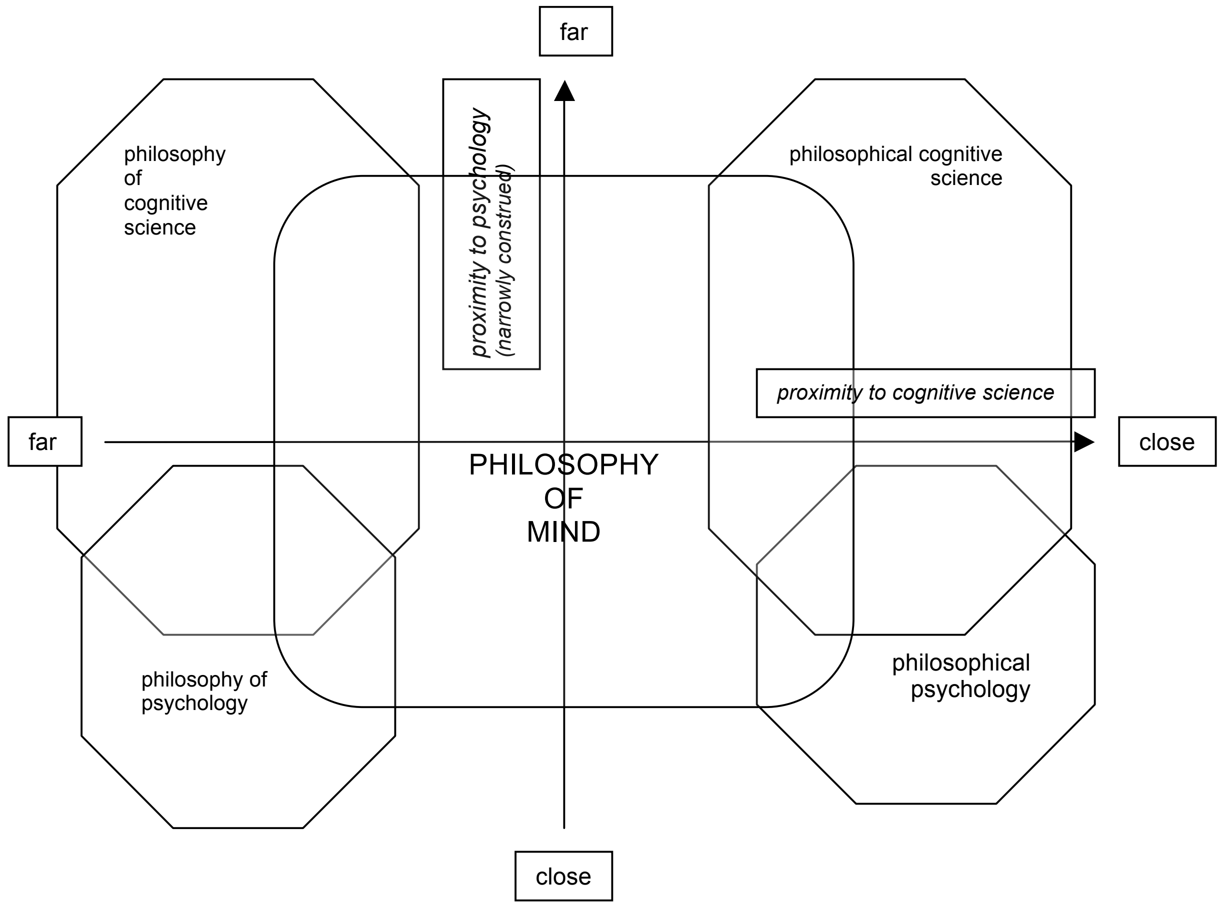
[9] For example, the opening sentence of F. Jackson's and G. Rey's article "Mind, philosophy of" in Craig (1998) reads as follows: "'Philosophy of mind', and 'philosophy of psychology' are two terms for the same general area of philosophical inquiry: the nature of mental phenomena and their connection with behaviour and, in more recent discussions, the brain". M. Davies' "Cognitive science" in Jackson & Smith (2005) is one chapter of a part entitled "Philosophy of mind and action" (next to "Consciousness", "Intentionality" and "Action"), and consists almost entirely in a discussion of central topics in philosophy of mind. Note that this distribution of titles makes philosophy of cognitive science a part of philosophy of mind, while others might draw the inclusion sign in the opposite direction, and I prefer to think of the two as distinct areas having a broad intersection. F. Jackson, again, had previously co-authored a book with D. Braddon-Mitchell whose title seems to reflect yet another way of cutting the pie: *Philosophy of Mind and Cognition*.

type of inquiry is, in the case at hand, philosophy *of* cognitive science in the traditional, restricted sense.

There is a continuum, rather than two discrete positions, along this proximity/sympathy axis, and this is important. As one leaves the high-proximity end (philosophy in/as cognitive science), the problems become increasingly conceptual, the philosophical component takes precedence and one moves away from the day-to-day scientific work. One gets nearer the traditional position of appraisal and concomitant philosophical speculation, and one enters the area of 'ontological problems'.

The second axis measures distance from psychology in the traditional sense. At the near end, there is psychology itself; at the far end, cognitive science in its widest sense, where psychology as traditionally construed no longer occupies a privileged, central position. Some readers may be surprised to learn of the existence of such a conception, but the history of cognitive science began with claims by artificial intelligence (AI) to subsume and thus supersede psychology, and has now come to a point where similar claims are made on behalf of cognitive neuroscience. One of the main duties of philosophy of cognitive science is to critically examine and compare these opposing views of the essential nature of the field. But notice again that there is a continuum: on some views, psychology retains a distinguished position without constituting the very heart, let alone the entirety, of the field.

We are now in a position to draw a map of the area (see fig.1). We start by pretending there is a discrete 2-dimensional logical space with 4 positions. Horizontally, we have FAR / CLOSE with respect to (cognitive) science. Vertically we have CLOSE / FAR with respect to traditional psychology. Then in the lower left position we find philosophy of psychology; above, philosophy of cognitive science (in the strict sense); in the lower right there is philosophical psychology; in the upper right, philosophical cognitive science (or philosophy as cognitive science).


far
proximity to psychology
(narrowly construed)
philosophy
of
cognitive
science
philosophical cognitive
science
proximity to cognitive science
far
close
PHILOSOPHY
OF
MIND
philosophy of
psychology
philosophical
psychology
close

Finally, we need to introduce two complications in order to get a realistic picture with familiar labels. The first is to replace the discrete positions by a continuum along both dimensions. This creates a middle zone straddling all four positions. The only available label for this middle zone is philosophy of mind. It intersects with philosophical psychology and philosophical cognitive science, on the right, and with philosophy of psychology and philosophy of cognitive science, on the left. However, philosophy of mind extends beyond the entire space, as some philosophers working on the mind raise metaphysical issues quite independently of any science of the mind, past, present or future. Thus there is a part of philosophy of mind which resolutely straddles both cognitive science and its philosophy (this part is sometimes called cognitive philosophy, or again philosophy of cognition), and a (admittedly smaller and less visible) part which is light-years away from it all, with all degrees in between. Similarly, one may be tempted to take philosophy of psychology to be a proper part of philosophy of cognitive science, and philosophical psychology a proper part of philosophical cognitive science. But that would be to ignore or preclude the possibility of a non-cognitive form of psychology, together with a philosophy of non-cognitive psychology, or rather, a philosophical examination of the claim that psychology is not, and should never become, entirely immersed in cognitive science.

Some readers, especially among cognitive philosophers[10], might object to this taxonomy, finding it otiose or at least unnecessarily complicated (or worse yet: making a fuss over labels). Yet, on the one hand, it seems important to emphasize the co-existence of rather different research programs involving philosophy and cognitive science. On the other hand, it seems no less crucial to allow for philosophical or scientific enterprises which are both concerned with the mind and free of any analytical connection to cognitive science, although clearly, as their distance from cognitive science grows, the relevance of these enterprises for philosophy of cognitive science in the wide sense vanishes to zero.

c. Philosophical Styles and the Place of History

As hinted above, philosophy of cognitive science (lato sensu) and especially philosophy of mind (in the restricted sense of cognitive philosophy) have been so deeply linked to the development of analytic philosophy in the last quarter century that up until recently there have been few contributions from other traditions in philosophy, such as phenomenology. There were exceptions: AI, an early avatar of cognitive science, was critically examined in the 1970s with the help of tools drawn mostly from existential phenomenology or hermeneutics[11]. This enterprise was however conducted in North America and in the framework and style of analytic philosophy. There is now a growing body of work which blends the analytic style with phenomenological themes[12]. On the other hand, philosophers trained outside the analytic tradition, which compose, for example, the vast majority of French philosophy, have by and large remained unconcerned by or very dubious about both cognitive science and the associated philosophical inquiries[13]. Perhaps these differences will become less relevant, as

[10] Note however that this chapter is not particularly aimed at them, but rather at practitioners of other branches of philosophy of science.
[11] Dreyfus, 1972; Haugeland, 1998 (original paper 1981), Taylor, 2005 (original paper 1971).
[12] Andler, 2006b, Smith and Thomasson, 2005.
[13] An exception was mentioned above, see note 3.

parts of analytic philosophy become more permeable to 'continental' influence, and vice-versa: opinions vary widely on the probability and on the desirability of such a *rapprochement*. But it would take us too far afield to discuss the infamous 'analytic-continental divide' and its repercussions on the topic at hand.

Let us turn our attention, instead, to a more restricted issue, that of history of science. Continental, and more particularly French, philosophy of science, is wedded to history, and analytic philosophy of science today, or what is sometimes called 'post-positivist' philosophy of science, has espoused the view that philosophy of science cannot dispense with a historical perspective. Thus philosophy of cognitive science (stricto sensu) would seem to go hand in hand with history of cognitive science. The trouble is that this latter field remains to this day quite underdeveloped, leaving the former somewhat handicapped. In the present chapter, some minimal historical landmarks for cognitive science will be provided in passing, yet nothing will be said about the history of philosophy of mind. Although this conforms to the usual treatment of the topic, and to accepted practice within analytic philosophy, it must be acknowledged as a deficiency. First, there is an inconsistency in providing some historical perspective, however scant, for the scientific part, narrowly construed, of cognitive science, while denying the more centrally philosophical part a similar treatment, given that the multiple connections between the two parts is a key feature of the field. Second, it is quite likely that an account of the genesis of the main ideas and concepts in philosophy of mind, both currently in the mainstream and heterodox or less fashionable, would throw some light on the present debates and provide conjectures on the underlying dynamics; it would assuredly also considerably help the non-specialists find their way in the thicket of the existing literature. Unfortunately, not only do I lack the expertise to provide this much needed historical perspective; but it does not seem readily available[14]. To be sure, there are collections which include a few classical texts, from Plato to Russell[15]. However, they suffer from two unavoidable shortcomings: the excerpts are perforce taken out of context, and invite a whiggish interpretation by the historically innocent; and they carefully avoid providing a narrative, without which, however subjective and incomplete, not much can be gained, beyond encyclopedic knowledge which is of little use in achieving a true grasp of the current situation. In addition, these collections tend to favor the distant past (sometimes going back to Plato, Aristotle, more ofter starting with Descartes, carefully avoiding Kant and reaching for Brentano), and then jump to contemporary authors, leaving not much room for recent authors (*e.g.* Wilfrid Sellars, to name just one recently resurrected figure, Gilbert Ryle or Herbert Feigl, or again the Pragmatists), who are likely to have directly inspired some of the more esoteric proposals under scrutiny today, and almost completely neglecting non-naturalist thinkers, followers of Husserl, Merleau-Ponty or Wittgenstein, not to mention Ernst Cassirer or Suzanne Langer.

There are several well-known retorts to these historical scruples. The first is that they are misplaced, either because philosophy of mind is an essentially new enterprise which is no more interestingly related to its distant ancestry as say contemporary physics is to

[14] Alas, I discover only now, as I finalize this chapter, that Paul Livingston's 2002 dissertation, and his subsequent 2004 book, are devoted to precisely this enterprise, and it seems that his motivation was exactly the worry I express here.
[15] *E.g.* Rosenthal (1991), Morton (1996), Chalmers (2002), Heil (2004).

Archimedes; or because the genesis of a philosophical problematic is at best of tangential help in our attempts to clarify and solve the problems as they are set up today. The second is that the dream of a systematic philosophy, which seems to underlie this perhaps unreasonable longing for history, belongs to the past, or else must been thought of as a blessing rather than a goal: one cannot aim for it, one can only hope that it be realized, at certain times and within certain communities or single thinkers, as a form of culture or wisdom. The third is that historical studies directly relevant to present concerns are not truly feasible, be it for conceptual or institutional reasons.[16] But whatever the merits of these responses, they are ineffective as cures against my scruples: I still deplore the absence of an informative historical frame for philosophy of mind.

We are finally ready to examine some substantive issues. The rest of the paper comprises two sections. Section 2 is devoted to the central task which philosophy of mind has set itself, *viz.* to provide cognitive science with a conceptual framework. Section 3 concerns the fit between the framework and cognitive science, and includes discussions of a sample of issues internal to the field.

2. A CONCEPTUAL FRAMEWORK FOR COGNITIVE SCIENCE

Cognitive science is often simply defined as the science of the mind, while philosophy of mind can be seen, first and foremost, as the exploration of the ontology of the mind. As we limit ourselves, from now on, to that part of philosophy of mind which is in direct contact with cognitive science (cognitive philosophy, or philosophy of cognition), we can view its general aim as providing a foundation to cognitive science. Such a foundation is to be sought neither in pure a priori, conceptual analysis, nor in some kind of inductive generalization from the practice of cognitive science. It consists rather in searching for a reflective equilibrium between the ontological principles suggested by philosophical inquiry and what may be called the ontological practices, or perhaps the implicit ontological commitments of cognitive science. Paraphrasing a famous title[17], we are asking, as it were, What is a Mind, that Cognitive Science May Know It, and Cognitive Science, that It May Know a Mind? While the second part of the question falls squarely in the province of the philosophy of cognitive science, the first states in the most general way the purpose of the philosophy of mind. Thus combined, they are seen as mutually dependent. One must however choose a starting point, and I shall begin with the conceptual framework which philosophy of mind has proposed for cognitive science. The focus will be on what is often called the 'classical' paradigm, but some mention will be made of possible deviations, and what will be offered is a liberalized form of the paradigm, which I believe to be at this point in time the inevitable point of departure for any inquiry into cognitive science (I will return to this point in section 3).

[16] There are exceptions, of course. For example, there are monographs on Kant as a precursor of some key ideas in contemporary cognitive psychology and philosophy of mind: Kitcher, 1990; Brooke, 1994. The limits of such works is that they tend to rehabilitate only one particular line of ancestry (the Kantian one being admittedly of unusual importance).

[17] Warren S. McCulloch, "What Is a Number, that a Man May Know It, and a Man, that He May Know a Number?", in McCulloch (1965) (original paper 1961).

a. The Mind-Body Problem, Physicalism, Functionalism

A survey of the mind-body problem, however compressed, cannot be attempted here, for reasons of space[18]. Fortunately, in the somewhat limited perspective I have just proposed, such a survey is not strictly necessary. What we need to focus on are the points of contact between the philosophical analysis of the problem and the scientific practice.

The first thing to notice is that the connection is not as strong as one might think.

Although there are many variants of the mind-body problem, they share a core, which is a longing for an understanding of the way in which the mental realm, which appears to float free of the physical realm, might fit, or, alternatively, of the reasons why it could not conceivably fit, in an overall picture which accommodates everything that, and nothing but what exists in the natural world. Thus one might think that the analysis of the mind-body problem which philosophy of mind is supposed to deliver is crucial to cognitive science, in the sense of being an 'enabling' factor. Or, conversely, that cognitive science provides philosophy of mind with essential empirical, or more broadly, scientific ingredients of, or constraints on a solution to its central problem.

The starting point for this way of thinking is the core belief, accepted by many philosophers, and scientists, that cognitive science will inevitably provide in due course a thoroughly naturalistic account of the mind, on par with our current understanding of lightning, eclipses, tropical storms or illness which were once held to elude explanation by natural causes. This core belief in turn is supported by two lines of thought. One, the 'fast track', enlists, severally or jointly, two arguments. The first takes as premiss the perfect record of physics (construed in a sufficiently broad sense as to include chemistry and all the 'special' natural sciences) in accounting for any and all aspects of the real world which it has attacked, and proceeds inductively to the conclusion that the mental realm is likely to succumb as well. The second takes as premiss the success of evolutionary theory in accounting in strictly natural terms for the presence of complex functional systems in the living world, and proceeds inductively to the conclusion that the mind, a complex functional system if there ever was one, can and eventually will be seen as nothing over and above an evolved, and hence natural part of the living world.

The second line of thought (the 'slow track') starts with a consideration of the ongoing work in cognitive science, which began, in recognizable form, in the 1950s, and is generally regarded as flourishing[19], unimpeded by any trace of a major crisis which, as happened for the first phase of artificial intelligence, might spell the end of the enterprise. To the contrary, the rise of cognitive neuroscience, fuelled by functional imagery, seems to both considerably enrich the toolkit of cognitive science and reinforce its character as a *natural* science. Thus cognitive science is in the process of showing the natural character of the mind by actually proceeding, step by step, to an effective naturalization of the mental realm.

[18] See however Ludwig (2003) for a particularly clear and thorough exposition; Warner & Szubka (1994) is a good collection of papers.
[19] For a less sanguine view, see Johnson & Erneling, 1997.

Both tracks converge on the *philosophical* thesis of naturalism, which may or may not be further specified as physicalism. The thesis, in both its liberal and strict (physicalist) versions, generates in turn a budget of philosophical puzzles –such as the apparent impossibility of mental causation due to the alleged causal closure of the physical– which constitute an important part of the agenda of philosophy of mind.

However, the general arguments briefly sketched above in favor of naturalism, or more stringently, physicalism, are notoriously inconclusive in the eyes of those not already persuaded, and in some sense they seem indeed to beg the question. In fact, even the keenest defender of philosophical naturalism can see that a full naturalization of the mind delivered by cognitive science remains a distant prospect. But the discussion lies for the most part outside the immediate agenda of the philosophy of cognitive science, except for the methodological examination of the naturalization programs deployed in the field, which will not be undertaken here due to space limitations. The important point here is that cognitive science does not require a prior belief in the inevitability of a complete success of naturalization. At most it needs reassurance that there is no incontestable proof or overwhelming evidence arising from other areas of science, including the formal sciences[20], or from non-scientific sources, which would establish beyond reasonable doubt that the mind (or at least some essential dimension of it) is not accessible to science.

This being said, cognitive science and philosophy of mind do lend one another considerable support, as indicated, and for the general reason given at the outset. The apparent conflict arises from the mind-body problem. How can the philosophical discussions of this issue both be and not be relevant for cognitive science, and conversely how can the general orientations and results of cognitive science both be and not be relevant for the resolution of the philosophical mind-body problem? The short answer is that philosophy suggests, and cognitive science supports, precisely the idea that a solution of the philosophical problem, at least in the traditional terms of dualism *versus* monism, is not required for cognitive science. The detailed answer consists in an exposition of the doctrine of functionalism.

Functionalism means somewhat different things for different authors at different times[21], but the core of the doctrine is that mental entities such as beliefs and desires, pains and rememberings, regrets and fears, are functionally defined kinds of inner states which can be entirely individuated by the role they play in the dynamics of the cognitive system, with sensory stimulations and motor responses constituting a set of observable boundary conditions. So that believing that the sun is setting is individuated by the relations which obtain between that belief, other beliefs, desires, perceptual states and motor responses: the belief that the sun exists, the perception of the sun nearing the horizon, the knowledge that the air cools at sunset, the desire to keep warm, the memory of where the sweaters are stored, the motor commands leading to the cupboard, etc. The crucial point is that there need not be anything further to know, or perhaps even to be known, about desires and other mental states.

[20] There is in fact an entire tradition of attempts to infer such an impossibility from Gödel's incompleteness theorems, originating in Lucas, 1961.

[21] I am not referring to the many varieties of functionalism outside psychology (linguistics, anthropology, etc.), although there presumably is an even more abstract core common to all forms of functionalism.

We can remain blissfully ignorant of their ontological status, of the stuff they are cut out from, whether material or ethereal, and go about uncovering the laws of thought, or in other words, conduct the business of empirical psychology[22].

This core idea calls for specification along a number of dimensions. The more fully worked-out theories divide up in three main types. *Commensense* (or *analytic*) *functionalism* holds that mental states are actually *defined* by the set of 'platitudes' or commonsense regularities in which they are unreflectively seen to enter: the *meaning* of 'belief (B) that aspirin relieves headaches' is *exhaustively* provided (albeit implicitly) by the myriad generalizations regarding headache-relieving aspirin episodes (such as: Having a headache and having belief B tends to induce, *ceteris paribus,* aspirin-absorbing behavior). These are platitudes insofar as they are analytic truths, holding in virtue of the meaning of the mental state terms involved. Someone unable to immediately see them as true is not a psychological idiot, but simply fails to grasp the meaning of these terms.

Psychofunctionalism (or *empirical functionalism*) rests on a distinction between the *meaning* and the *reference* of mental state terms. What we need to know, regarding, say, the generic notion of belief, or one particular belief such as B above, is for science to discover. Following recent discussions in philosophy of language, many philosophers see the meaning of a natural-kind term like water as fixed by a combination of everyday linguistic and other social practices, while the reference of water (stuff made of H_2O molecules) is for chemistry to determine. Psychofunctionalists extend this view to mental state terms. Psychology, embedded in cognitive science, will ideally determine the causal nexus characteristic of any given mental state (or process)[23].

Machine functionalism is Hilary Putnam's initial version of functionalism[24], and it equates mental states with the states of a Turing machine or more broadly a computational system. Despite the fact that it has been all but abandoned in its original form, also by its proponent himself[25], machine functionalism is by far the most relevant for cognitive science as what provides it with a (partial and provisional) foundation, for it is strongly connected to the 'computer model of the mind', or more accurately, to the computational theory of mind.[26]

b. The Computational Theory of Mind (CTM)

[22] As Hatfield, 1995 makes clear, some major figures in 18th century psychology had already grasped the possibility of combining a naturalistic approach with an agnostic attitude toward ontological monism (or even an acceptance of dualism).

[23] I do not discuss at this point the difficulties of functionalism or of any one of its varieties. However it is hard not to notice one glaring threat on psychofunctionalism. If the meaning of B is determined by the set (P) of platitudes of everyday psychology, but the scientific inquiry leads one to the conclusion that B does not obey any approximation of (P), what is one to do? The 'Eddington' move (accepting both the commonsense representation of the table as hard, etc., and the physical representation as a cloud of particles) is not available, as we have no grip on a given belief comparable to the one we have on the table in front of us. We seem forced to eliminate B from our ontology. But if this happens for all, or even for many mental state terms, hasn't the topic vanished altogether? One response is to rule out as highly improbable, or at any rate, unmotivated, this worst-case scenario, and to stress the plausibility of scientific psychology rectifying commonsense platitudes without massively contradicting them. This is the bare beginning of a discussion which has been raging for three decades.

[24] Putnam, 1975 (the first original paper is dated 1960).

[25] Putnam, 1988.

[26] So much so, in fact, that in many contexts (outside 'advanced' philosophy of mind or philosophy of psychology), functionalism just means CTM. (Of course, functionalism means many other things in the context of other disciplines).

A natural and quick way to introduce the CTM is to first consider another reason why machine functionalism is so important in this context: it provides a vivid, accessible illustration of the main conceptual features of all forms of functionalism.

The most central of these is that functionalism allows for *multiple realizability*. Chairs, functionally defined as pieces of furniture providing support for sitting, can have all sorts of shapes, be constructed in all sorts of ways, out of all sorts of materials. Similarly, a lever, a pulley, a wheel, a carburetor, a corkscrew are what they are not by virtue of their *specific* or *intrinsic* properties, but by virtue of their ability to fill a certain role in the framework of a larger system. Functionalism as a doctrine of the mind likewise views mental states and processes as roles variously filled by (human) brain states and processes (different brains, or even the same brain at different moments, filling the roles in different ways), and also conceivably by certain non-human brains, and even by artefacts such as computers or other complex devices. Computers, or their notional paradigm the Turing machine, provide a perfectly straightforward example of multiple realizability: 'believing' that 3 times 7 is 21 is 'realized' in an indefinite variety of ways according to the program (or machine table) which executes an algorithm for multiplication; and what makes it the 'belief' which it is is entirely, and non-mysteriously, a matter of the causal links between that state and other states and processes. The very same 'belief' can be present in computers with very different logical and material structures.

The computational theory of mind originates in this intuition, to which Putnam's papers in the mid-70s only gave a philosophically arresting form, as it had been adumbrated by Turing himself in his 1950 *Mind* paper, and developed by the fathers of artificial intelligence from the mid 1950s onward.

There is another crucial dimension to Turing's idea: computation as he (re)defined it is mechanical, hence poised for inclusion in the physical realm. The mind can be seen as a set of processes operating on a set of states. The processes are natural to the extent that they are causal, and if computation is mechanical (in a conceptual sense), Turing's work shows, and modern computers prove, that the mind is actually mechanizable, *i.e.* realized by a concrete mechanical system. What appears to some as the beauty of Turing's proposal (and to others as a weakness) is that it allows, yet does not force, a materialist solution.

But now we clearly see that the computational intuition is nonetheless not sufficient to generate a theory of the mind[27]. This it achieves only when grafted onto a much older idea, which originates in the philosophical psychology of the 17th century and permeates scientific psychology until the advent of CTM's predecessor, behaviorism. This venerable idea is that of a (mental) *representation* (or *idea* in the language of Descartes or Locke). Computation operates on *data*, in the original sense, or *inputs.* Now data *need* not represent (stand for) anything beyond themselves as material formal entities. The essence of computation is made (almost) entirely manifest by an example such as: concatenating XXX with XX yields XXXXX, where each one of the five Xs, as a material token, provides the computational device all it needs to proceed[28]. However, the most salient attribute of the mind is its ability to

[27] This was not clear, I think, to the ancesters of CTM, *viz*. the cyberneticians who started working in the 1940s on the question of how a brain could think: see Andler, 1992. For Turing, on the other hand, it was clear, only too clear: he failed to see that representation posed no less a challenge than computation: see Andler, 1998.

[28] There is, to be sure, a lot more to say regarding the status of X as a token of the X-type: the device needs to 'know' that despite inevitable differences between the X tokens it runs into, they all count as interchangeable

deal with entities lying clearly outside itself and forming an open-ended collection. In simple terms, it is constitutive of a mind, under nearly any construal, to be engaged, via the various perceptual modalities, in continuous interaction with the world (including other mind-bearing organisms). To determine what distinguishes minds in perceptual contact with the world among systems causally affected by the world is a problem of the highest difficulty, and the concept of a mental representation is offered as the starting point, indeed as the lynchpin of a possible solution. When an asteroid hits the Earth, the result is a crater. When a rock impinges on my visual system, a representation of the rock is formed in or by my mind. Although perception is presumably not the only process by which representations are generated, it is enough to make *some* notion of representation very nearly impossible to dispense with.

To summarize: mental states are representations, appropriately labelled according to their status as propositional attitudes (beliefs, desires, fears, regrets...), and mental processes driving the mental dynamics, the transitions from state to state, are computational.

Computation and representation can now be assembled into a skeletal model of the mind. The connection is located at one precise point: the formal tokens on which the computation operates now represent aspects of the outside world; in fact, they can represent anything at all, including facts and events concerning the model itself, and including non-facts and non-events involving real or unreal entities. The representational content of a token X is dubbed *information*: it is, in roughly the usual sense of the word, information *about* what X is a representation of. This is why the CTM is more properly called by some authors the 'computational-representational', or 'computational-informational' theory of the mind (and, for the sake of completeness, let me repeat that sometimes 'functionalism' is used as yet another synonym).

As I have introduced it, CTM requires quite a bit of building up to start looking like something more than a ghost of a theory, a 'we-know-not-what'[29]. This is the purpose of the next two subsections.

c. Rationality, the Systematic Mind, and the Language of Thought Hypothesis

The intellectual landscape in which the CTM emerged was shaped by the accomplishments of logic in the 1930s. Of prime importance for CTM is the notion of a formal system. Formal systems were adumbrated, though not fully constructed, by Frege, who was after a 'good' language for mathematics and science (this was of course, an old philosophical goal, but Frege was aiming this time for a scientific solution). Formal systems are exactly what CTM requires as a medium of computation *and* of representation. Their generative structure endow them with representative power. As Leibniz had dreamt, and Frege shown possible, they potentially contain (*ie* they can generate) symbols for any number of entities and states of affairs (under an essential proviso, which will be discussed presently).

---

tokens of the same type. Thus in a philosophically interesting sense, any X token *represents* type X. Proceeding along this line might yield a resolution of an old controversy between Jerry Fodor and Hilary Putnam, the former holding that 'no representation, no computation', on the grounds that a computation necessarilty implies a computational *domain*, and the latter arguing that the theoretical foundation of computation was historically developed, and can be entirely presented, without any reference whatsoever to a notion of representation..

[29] This is the expression which Putnam 1994 applies to functionalism (or CTM) even *after* it has been beefed up. I will return shortly to the incompleteness of CTM.

And their compositional nature makes them suitable for computation. This in turn involves two crucial properties: mechanical semantics and mechanical inference. Let us spell this out.

Leibniz's dream and Frege's aim was a scientific language in which the pursuit of truth could be conducted in the fashion of elementary arithmetic: 'Let us calculate'. Frege turned logic, which for Aristotle was the canon of valid inference, into a language, without sacrificing its original function. Aristotle's notion of formality contains the modern idea of syntax achieving the right semantic effect. In the syllogism "All As are Bs, some Cs are As, therefore some Cs are Bs", the mind remains on the path of truth by merely noticing the correct formal identities (as we would say, correctly identifying the types of the various tokens): it need not know what A, B, C stand for, let alone investigate the C population to ascertain whether some are also Bs. In other words, rationality (construed as truth-following) is achieved by formal or syntactic means only. But there is no *operation* involved: the inference rests on a *noticing*. In the modern (Hilbertian) perspective, an inference is an operation, notionally involving the manipulation of marks on paper. The second modern ingredient missing in ancient logic is the ability to encode or represent relations: only monadic predicates have a formal expression. Frege (and Peirce) revolutionized logic by introducing binary predicates and the ontological operators, quantifiers, which they require. They paved the way for the definition, in the 1920s, of a formal language in which all of the features of a situation in a given domain can be expressed, and where the syntax (the set of formal operations) 'mimics' the semantics. Gödel and Turing took the last essential step by showing that the formal operations are computable, and thus mechanizable.

The entire paragraph above, it is important to notice, has nothing directly to do with psychology, a theory of the mind. It has to do with thought, which is what a mind does. There is no simple abduction from what the mind does to how the mind is constituted: extra work, and considerable scientific imagination, is required. In fact, for centuries it seemed that the fact that the mind could apparently do just about anything meant that it could not have much structure.

The language of thought hypothesis (LoTH) is precisely the conjecture that the mind itself, regardless of what it is doing, in the familiar sense, at any given time, operates like an automated formal system, *i.e.* applies inference rules to formulas of a formal language, called 'mentalese' or 'language of thought' (LoT). LoTH presupposes CTM, and is often in fact taken to be a way of specifying it.

This way of putting LoTH is however less than fully satisfactory for two reasons at least. The first is that having the mind 'apply inference rules' seems to force one into either an infinite regress or a counterintuitive psychology. For if we mean to say that the mind applies rules in the same sense as we would say, for example, that a child adds 12 and 17, or a shopkeeper returns change, then we want to know again how the mind accomplishes this task (aren't we after an account of how a child adds and a shopkeeper returns change?). Alternatively, if what we mean is that, contrary to what appears on casual introspection, whatever we are accomplishing mentally (thinking of how to begin a lecture, remembering a deceased parent, spelling our name on a form, making our way to our airport gate, breaking up with a lover, grasping a missing link in a story, etc.) *we* are actually applying rules to formulas, then we have to face severe objections of an empirical or phenomenological kind. But *neither one or the other is what is meant or implied by LoTH.* Rather, the suggestion is that when we accomplish a typical mental task, there are components of our cognitive apparatus, *i.e.* our central nervous system, whose trajectory is conspicuously described as the application of some rules to some formulas of an inner language. In a nutshell, when *I* accomplish any one among the myriad mental tasks people go through as a matter of course, the *cognitive system* responsible for my having a mind, and the mind I have, applies rules to formulas. The fate of LoTH, and probably of CTM as well, hangs on the possibility of making this suggestion fully intelligible.

The second reason why the simple formulation above is unsatisfactory is that it runs together two kinds of computational process. One is composition and regards *representations*, which are typically names for things or for states of affairs. The name, in mentalese, for Nicole's grandmother (call it X) is made up of the mentalese name N for Nicole, the mentalese word G for grandmother, and the mentalese word P indicating the relation of possession; and the process by which the mind goes from the set of symbols {N, G, P} to X is computational. Likewise, the fact that the cat is on the mat is represented by a formula of mentalese –something like $On_mCat_mMat_m$– which is the result of a computational process acting on its components On_m, Cat_m, Mat_m. The other kind of computational process concerns *inferences*, which are the transformations undergone by sets of representations. For example, from the facts that (a) the cat is on the mat and (b) the mat is in the back porch, there follows (c) the cat is in the back porch. How *I* get to believe (c) results from *my cognitive system* having applied a rule of spatial logic to the mentalese counterparts of (a) and (b). (For readers suspicious of spatial logic, the example can be recast as (a), (b') If something is on the mat, it is in the back porch, (c)). Now it is just as important that the process of getting from the representation of some entities or relations to the representation of either a composite entity or a state of affairs be computational as the property of getting from some premisses to a valid conclusion be computational, for the mind requires both kinds of moves to get on with its business.

The point of going over the familiar genesis of the modern idea of a formal system is to stress (i) the non-trivial character of this invention; (ii) the non-obvious (indeed, controversial) inference from the notion of a formal system as a medium *on which* the (logical) mind operates to the hypothesis of a formal system as a medium *by which* the mind, engaged in whatever task, operates; and therefore (iii) the unavailability of LoTH and hence of a reasonably worked-out form of CTM to thinkers working prior to the development of modern logic. I think all three of these points, however unoriginal, deserve to be made, as they are often not taken in, a cause for unnecessarily long expositions of objections which miss the target[30].

Finally, two arguments in favor of LoTH must be mentioned. The first counts more generally in favor of CTM. I said earlier that a prime feature of the mind is that it seems able to carry out just about anything, from determining the 111th prime number to imagining a purple cow with wings, buying bread or writing a haiku, and that this argued against any account of the mind which would give it any determinate structure. Now if the mind is somewhat like a computer, which is what LoTH implies, then it could be an approximation to

[30] They also argue against putting too much weight on past authors' alleged prescience of the CTM (Ockham, Hobbes, Leibniz, Kant, Boole, Babbage et al.)

a *universal* Turing machine, *i.e.* a device which can effect *any* computation.[31] Together with the representational potential of formal languages, this universality plausibly confers a high amount of versatility to the system.

The second argument, developed at length by Fodor and Pylyshyn, consists in showing that LoTH is the best explanation (in fact the only available explanation) for some central features of the set of thinkable thoughts, *viz.* its productivity (the fact that for any finite set of thinkable thoughts, there is a thinkable thought which doesn't belong to it) and its systematicity (in the sense given by Fodor: it leaves no gaps, so that if 'John hates Peter' is thinkable, so is 'Peter hates John', 'Peter hates someone', etc.). *If* thinking consists in moving between structured symbolic representations, then the productivity and systematicity of the set of thinkable thoughts is nicely accounted for by LoTH[32].

d. Intentionality

The general problem of intentionality is often seen as the central problem in the philosophy of mind, and it consists in accounting in ontologically acceptable terms for the *aboutness* relation which holds between a mental event such as a thought or a belief and what that thought or belief is about. There seems to exist nothing else in the world which has this essentially directional character. It does not do to say that a thought of mine is about the Eiffel tower just because it is associated, in however strong or privileged a way, to the Eiffel tower, for then the Eiffel tower, by the same token, would be about my thought. It does not do to say that the way in which smoke is about fire is a paradigm for the way in which my Eiffel tower thought it about the Eiffel tower, because fire always causes smoke (in the usual sense of fire and the usual circumstances), while the Eiffel tower only very rarely causes an Eiffel-tower thought in me. It does not do to say that my thought about the Eiffel tower is related to the Eiffel tower in somewhat the way in which the French locution 'la tour Eiffel' is related to the Eiffel tower, because the latter relation presupposes the first: only a creature capable of understanding the French locution, hence of having Eiffel-tower thoughts, can sustain the relation between the French words and the metal construction (a point sometimes made in terms of *derivative vs original*, *primitive*, or *intrinsic* intentionality).

In the context of CTM, the problem becomes that of understanding how the formal symbols in the system get to carry information about the Eiffel tower and other external entities, events or matters of fact. This is because for the system to have a thought about the Eiffel tower consists in the system 'activating' a (possibly complex) symbol which carries the appropriate information about the tower, and labelling it with the appropriate propositional attitude: belief, desire, regret, fear, etc.

There are actually two problems. One is to make intelligible the very idea of a representation in naturalistic terms. The other is to give an account of how a given internal (mental or brain) state gets to represent a particular thing (object, relation, event, fact,...). There is a functionalist attempt at solving the first problem, which goes roughly as follows. For the internal state S to represent X simply is to play the appropriate role in a causal nexus

[31] For an appraisal of the significance of the universal Turing machine, see Herken, 1988.

[32] See the papers by Fodor, Fodor and Pylyshyn, and Smolensky in Horgan and Tienson, 1996 or Macdonald and Macdonald, 1995.

which terminates at the perceptual and motor interfaces. I think it is fair to say that this attempt has failed to elicit a broad consensus. More progress has been made on the second problem, yet even on that front there is hardly a consensus that we are nearing a solution. There is also little agreement on how important it is to solve the problem, on what would count as a solution, or even on whether it is well-posed.

There are, it would seem, at least four broad categories of mental representations. The first represent objects, the second represent n-ary relations (n = 1, 2,...), the third represent logical constants, and the fourth represent states of affairs. The discussion is initially restricted to concrete entities: this particular object, the particular predicate 'is a cherry', or 'is red', the singular fact 'this is a red cherry'. And it leaves to the side the issue of nonconceptual content, which concerns representations, if they exist, which are not propositional (more about this in the next subsection).

Surprisingly perhaps, what gets most attention in current discussions in philosophy of mind is the seemingly most complicated kind, *viz* representations of states of affairs. Several theories are at hand which purport to account for the relation between a mental representation and the state of affairs which it represents (for example: 'There is a predator closing in on me', or 'This is a horse'); or to put it in causal or genetic terms, an account is sought of how a given material state or event gets to carry the information that there is a predator in front of me, or this is a horse. One strategy is to start from clear biological, psychological or engineering cases in which there is no reasonable doubt about a component, a state or a process *carrying* some information regarding the external world. The aim then is, on the one hand, to generalize from these unproblematic examples to an overarching informational theory of mental content or intentionality (or semanticity as one sometimes says); and on the other, to give a principled characterization of the relation which discharges a number of crucial obligations, the most central of which is to make intelligible the possibility of misrepresentation. This is the proper domain of informational (Dretske) and teleo- (Millikan, Papineau) semantics.[33] The intuition common to these theories is that some situations must be identified by a creature if it is to survive or at least to behave adaptively, and that those situations get to be represented mentally. The possession of a disposition to form a representation of one such situation is a biological trait of the creature which results from either ontogenetic learning or phylogenetic selection. Whatever the merits of these proposals, which cannot be discussed here, it is far from clear, at this point, whether they generalize beyond a set of very simple concrete, ecologically significant situations, which are indeed likely to be detected by evolved creatures.

But how do n-ary relations (n = 1 or 2 will do) get to be represented? Another way of asking the question is, How does a mind acquire a concept? One answer is, by learning from examples and counterexamples (this is what the field of machine learning is about). This answer again, whatever its merits, remains incomplete, for no learning is possible without the possession of more basic concepts. It is so difficult to imagine how basic concepts could be acquired that some authors have come to believe that they are not, that they are innate. But this makes the first problem of intentionality more pressing: having some idea of how a particular thing gets to be represented mentally may lead one to begin to grasp what the

[33] Dretske, 1981; Millikan, 1984; Papineau, 1987.

representing relation consists in; but in cases where there is *no* imaginable process by which a particular thing (in the case at hand, a concept such as 'is a cherry' or 'is a solid, permanent, rigid object', or a relation such as 'is taller than' or 'is hidden behind' or 'is a part of', or 'is caused by') *gets* to be represented mentally, then one has no initial grip on the problem of what it is for a mental state to represent the concept 'is a cherry' or 'is caused by'.

Leaving untouched the issue of logical constants, and the problem of abstract and of non-existent entities, I conclude this section by returning to the matter of what a language of thought can represent. I said that, with an important proviso, it could conceivably represent anything at all. The proviso is: anything which is representable as a finite combination[34] of its primitive symbols. Seen semantically, this means that only those entities, relations or facts which are expressible within the bounds of the conceptual apparatus of the creature under consideration, can be represented by that creature. This seemingly rigid limitation is seen by some critics to run against the evidence of an unbounded capability of humans to create new concepts and generate new thoughts (a capability which precisely goes beyond the creativity which LoTH accepts, and on the basis of which it actually is defended).

Yet our discussion of intentionality is not complete. It is time to introduce an essential new concept.

e. Tacit Knowledge, Subpersonal Processes, non-conscious Mentality

I have up to this point left implicit the assumption, encouraged by the toy examples provided, that what is represented in the mind is the sort of entity which typically populates our conscious, or even more specifically our verbalizable thoughts, and that the representations themselves are in fact either conscious or potentially so. That this has been thought, in certain contexts, to be a reasonable assumption is illustrated by the first phase of artificial intelligence. And that it can be essentially invalidated by experience is perhaps illustrated by the failures of this phase of AI[35]. But AI paid a dear price for ignoring the lessons of the prehistory of cognitive science. From Leibniz to William Hamilton, Helmholtz, Bain[36] or Lashley, there are numerous rediscoveries of the idea that our conscious mental states do not form a connected flow, whether in a causal, a rational or a temporal sense of connectedness, and that there must be intermediate or subjacent states without which no account of our conscious mental lives will be forthcoming. As Bain puts it: "Outward expression, however close and consecutive, is still hop, skip and jump. It does not supply the full sequence of mental movements".[37]

In the rise of cognitive science, this insight played a decisive role and is generally credited to Noam Chomsky, who postulated that linguistic ability is based on the speaker's

---

[34] This is non-technical shorthand for a notion of logic which depends on the logical system one believes to be the correct one for LoT.

[35] This may seem unfair and unwarranted: some might dispute the failure of early AI, and, even if one grants it, to incriminate this particular assumption among the many initial theoretical and technological conditions which presided over the genesis of AI requires some justification. Hofstadter, 1985: chapter 26, pp. 631-665 is a detailed attempt to do just that.

[36] I owe the references to Hamilton and to Bain to the above-mentioned chapter by Martin Davies (Davies 2005).

[37] Bain, 1893, p. 48, quoted in Davies, 2005.

tacit knowledge of syntactic rules[38]. Daniel Dennett developed, in his influential first book[39], a distinction between 'personal' and 'subpersonal' states and processes. Stephen Stich (1978; see also 1983) coined the expression 'subdoxastic' to characterize mental states which lead to genuine doxastic states such as full-fledged conscious beliefs. Douglas Hofstadter developed the notion of 'subcognition' in a paper dated 1982[40]. The rise of the connectionist approach, in the wake of a famous book subtitled *The Microstructure of Cognition*[41], led Paul Smolensky to draw the contours of a 'subsymbolic' level of description[42].

Throwing all of these notions into one basket seems to beg for trouble, as they are not just different presentations of the same idea. They do however have three crucial features in common. First, they are both related to and different from regular propositional attitudes as traditionally construed. Second, they are assumed to play an essential role in the genesis, attributes and dynamics of these regular mental states, as well as many if not all other aspects of our mental life. Third, they are proffered as genuine *cognitive* entities, and not as merely *physical* (neurobiological) entities.

As to the first point: These states or processes differ from regular propositional attitudes in one or more of the following dimensions. They are non-conscious; they may be inaccessible in principle to consciousness; they may defy description in ordinary language; they may be devoid of a logical structure. As to the second point: these states and processes are assumed to be either causally and/or explanatorily essential to the nature and regularities of our conscious mental life, particularly its more rational province populated with stateable beliefs, desires, fears, regrets and so forth. They are assumed to fill the gaps noted by Bain, but even more importantly they are thought to provide a *uniform* account of cases which seem to only require standard rational explanations *and* of cases which resist the treatment. The 'microcognitive' approach (using this neologism as shorthand for the various departures from 'fully' cognitive in the traditional sense) is also applicable beyond the realm of propositional attitudes, to perceptual or motor states for example. In fact, the discussion of the 1980s is continuing now in the context of mental states with *non-conceptual* (typically but not exclusively, perceptual) content.[43]

The third and final feature of these entities is that they are not *simply* brain processes or neuroscientific constructs. There is always a difficulty in stating precisely what this is supposed to mean, for, as methodological naturalists, cognitive scientists, like naturalistically inclined philosophers of mind, do not doubt that in the last resort, mental processes of any sort *are* brain processes. The point is that, just as is assumed about *cognitive* states, there is an essential feature which microcognitive states possess and which is not captured in the extensional vocabulary of biology as it is now conceived. (This is *not* as mysterious as it might seem: a 100-euro bill is, beyond doubt, a piece of paper, but it is also endowed with a function which paper science does not capture). They are *also* endowed with a function which, although accomplished by strictly neurobiological means, is *cognitive* in an extended

[38] Chomsky, 1965.
[39] Dennett, 1969.
[40] Hofstadter, 1985: chapter 26.
[41] Rumelhart & McClelland, 1986.
[42] Smolensky, 1988 (reprinted in refs given in note 31). For a more recent presentation of his view, see Smolensky & Legendre, 2006, vol. 2, chap. 23.
[43] Bermudez, 2003.

sense. What this amounts to is a contentious matter. Friends of cognitive science (as we know it today) appeal to a generalized sense of 'intentionality', which extends the usual relation to cases where the target or the content of the 'vehicle' (the entity inside the brain which is the first argument of the relation) can be something which does not belong to the usual ontology, *i.e.* is not picked by words, or concepts, or percepts to which we have unaided conscious access. They believe that microcognitive states can be *informational,* that they carry information which can be manipulated and combined so as to yield the desired conscious mental states or observable behavior, as LoTH postulated of cognitive states.

Skeptics such as Hubert Dreyfus or John Searle[44] see no hope of making sense of a general notion of information which would not appeal to a conscious human agent. In other words, they deny the reality, or the usefulness, of a level floating mid-way between full-fledged mentality and brain events (or, perhaps, electronic or other physical events). These skeptics about information *ipso facto* harbor doubts about cognitive science as deployed today, insofar as it may be surmised that any attempt to account for full-fledged, personal-level, conscious mentality in terms of something else which is not simply brain facts, will meet with their disapproval. And it can be argued that cognitive science as we know it parts ways with all previous epochs of psychology, as well as with commonsense psychology, despite numerous historical and conceptual affinities, precisely by giving pride of place to this intermediate level which is still mental but not conscious, nor translateable into behavioral dispositions, nor (as in Freudianism) systematically correlated to states which could intelligibly be interpreted as conscious.

This disagreement, seen in a broad perspective, pits two camps against one another. There are, on one side, those who see the aim of psychology as explaining mental phenomena as they appear to the untutored eye, embedded in a 'folk theory' nourished by introspection, by familiarity with people's sayings and behaviors, and by first and third-person reports, whether purportedly real or fictitious, in essentially different terms. On the other side, there are those who expect psychology to account for regularities in the mental lives and behaviors of human beings in terms commensurate with folk psychology, *i.e.* by adverting to fully intentional mental states and events. To put it bluntly, there are on the one hand the reductionists who believe that psychology *must* seek accounts of mental phenomena in the usual sense in terms of something different, on pains of being sterile, and on the other, the anti-reductionists who see this strategy as self-defeating, providing at best explanations or descriptions of something *else;* in a nutshell, as *changing the topic.* However this is *too* blunt: many in the first camp regard themselves as *anti*-reductionists. Reduction is always a relative matter.

f. Consciousness

Cognitive science, and philosophy of mind as we know it today, owe their existence, to a large extent, to the rejection of behaviorism's rejection of mental states. Yet cognitive science is also the heir to behaviorism, and one telltale 'hereditary' sign is that for a very long time it did not regard consciousness as part of its agenda. Philosophy of mind, by contrast, did, but the strategy, following Dennett's recommendation[45], was to tackle intentionality first, in the

[44] Dreyfus, 1972; Searle, 1992.
[45] Dennett, 1969.

hope that consciousness would fall out of the solution to that problem. Consciousness did play a marginal role in some experimental paradigms, but as a phenomenon to be studied it was reduced to attention, sometimes referred to as the 'gateway' to consciousness.

There was a rather abrupt change in the years 1975-1980, when almost simultaneously consciousness appeared, on the one hand, as presenting some deeply puzzling, strongly counterintuitive features, and hence as requiring scientific attention, and on the other hand, as unsuited for scientific study. So just as evidence of a 'cognitive unconscious' was mounting, some spectacular manifestations of which are priming (unconscious perception), blindsight, or unconscious memory[46], it dawned on some philosophers that for broadly logical reasons, conscious experience, the first-person phenomenon *par excellence,* could no more be studied from the third-person viewpoint which is constitutive of science than teeth can bite themselves.[47] Ned Block proposed to distinguish two 'kinds' of consciousness, the first of which, 'access consciousness', he assumed to be open to third-person investigation, the second, 'phenomenal consciousness', containing (in both senses of the word) the seemingly irreducibly first-person component[48]. Whether this is a valid distinction or a valid strategy remains hotly disputed. David Chalmers defends the view that there are 'easy' and 'hard' problems of consciousness, the latter forcing one into a position which he calls 'naturalistic dualism'.[49]

There is now a quasi-autonomous field of 'consciousness studies' where the disciplines which are involved in cognitive science play a leading but not an exclusive role[50]. Two of the new players are quantum physics, and that part of philosophy of mind which has no privileged relation to cognitive science. This autonomy however seems to reflect historical, institutional and other social-psychological factors, rather than a natural division of labor. This is not to say that consciousness does not constitute a special area within cognitive science, somewhat in the way cosmology, although part of physics, is a special part of it. Nor should we assume that consciousness is for cognitive science a 'normal' problem which is bound to yield to systematic efforts; the arguments to the effect that it might not are not easily deflected[51]. But intentionality, as we saw, gives rise to a similar, if perhaps less radical doubt, and so does, according to Fodor, 'belief fixation' in his sense. In other words, cognitive science accommodates problems *and* mysteries, which is why philosophy makes such an important contribution.

On the other hand, there remains a tension between those who see consciousness as a problem to be solved (or dissolved) by cognitive science with resources commensurate with those which it deploys at present (including mainstream naturalistic philosophy of mind), and those who believe that the lack of progress in accounting for the subjective, phenomenal character of conscious experience presages a conceptual revolution, in the wake of which the

[46] On priming: Marcel, 1983; on blindsight: Weiskrantz, 1986; on amnesia: Warrington & Weizkrantz, 1968. For more details see *e.g.* Frith and Rees in Velmans and Schneider, 2007 (chap.1).
[47] The *locus classicus* is Nagel's famous 1974 paper; Levine, 1983 coined the felicitous expression 'explanatory gap' to refer to what separates the target of any conceivable scientific account of consciousness from actual conscious experience.
[48] Block, 1995.
[49] Chalmers, 1996, and in Velmans and Schneider, 2007 (chaps.17 and 28).
[50] See Velmans and Schneider, 2007, in particular the list of institutions on pp. 727-8.
[51] See *e.g.* Levine in Velmans and Schneider, 2007 (chap.29).

entire edifice of cognitive science would be so thoroughly revamped as to become nearly unrecognizable.

3. ASSESSING THE FRAMEWORK IN THE LIGHT OF COGNITIVE SCIENCE AND *VICE-VERSA*

The entire discussion of section 2 provides a partial characterization of a family of related views about the mind, centered around the CTM and the LoTH, but allowing possibly important departures from them. This family of views, or this basic approach, provide what I will refer to as the liberalized classical framework (LCF) for cognitive science. There are two questions which must now be asked. One concerns the adequacy of the framework with respect to cognitive science as it 'really' is: How faithful, how informative, how useful? The second concerns the plausibility of the framework: Is it internally consistent, is it orienting the field in the right direction, *i.e.* does a true picture of the mind plausibly lay in the direction towards which it is pointing? The two questions –descriptive and prescriptive– are not independent. Cognitive science cannot be grasped as it 'really' is without the help of some principle of interpretation, which cannot be re-invented entirely anew for the purpose of assessing the framework under scrutiny: there is the familiar problem of the Archimedean point. As for the probability of being on the right track, it cannot be assessed independently of an evaluation of the results and failures of cognitive science as it is today: the proof of the pie is in part in the eating. To illustrate the predicament: An essential feature of the framework is that the mind is to be understood as an informational-computational system. How can we evaluate this proposal against what is being actually shown in cognitive science, when the criterion of a successful piece of research in the field is that it throws light on some mental phenomenon by revealing its informational-computational nature, where 'information' and 'computation' are to be understood in the way fixed by the conceptual framework?

There are however ways of breaking these two entangled circles, by some familiar back and forth maneuvers. Cognitive science provides explicit and implicit arguments pro and contra the framework, which in turn is needed to give meaning to the concepts which these arguments employ; the framework is thus thrown slightly off balance, yielding a slightly different view of cognitive science, and so on. The second circle is more problematic, as illustrated by the fate of previous epochs in psychology, such as behaviorism, which flourished until it all but vanished (or so it seemed) from the academic scene. However the issue raised is abundantly discussed in the debate opened by Kuhn's 1962 *Structure of Scientific Revolutions*, and this is not the place to explore it. Perhaps the history of cognitive science, as seen by future generations, will provide an interesting new case study[52].

a. The Incompleteness of LCF

[52] Unfortunately space does not allow for even a preliminary discussion of the distortions affecting many historical accounts of scientific psychology. The issue raised in the text depends of course on getting one's history reasonably straight. The reader is referred to Gary Hatfield's pioneering revisionary studies (Hatfield, 1995, 1997, 2002).

Perhaps the single most important fact regarding the status of cognitive science is that it is not simply the unfolding of LCF. First there are cognitive scientists who explicitly reject LCF. But second, and more importantly, despite its being quite substantive, and thus open in principle and also in fact to destructive attacks, LCF leaves a considerable amount of freedom to anyone wishing to develop an empirical science of the mind. Although more than a ghost of a theory, LCF provides at best a skeleton, and there are indefinitely many ways of building the full creature around it.

A simple way of making apparent the abyss between LCF and a worked-out research program for cognitive science is to ask the following schematic question: How does one go about finding out about X, where X = memory, learning, linguistic and communicative skills, musical ability, concept acquisition, language learning, reasoning, problem-solving, social competence, navigation, vision, audition, motor control, face recognition, special skills such as driving, playing chess or baseball, etc., and where 'finding out' can mean providing an explanatory account, a model, a functional analysis, a evolutionary rationale, a neural basis, a clinical manual? LCF provides precious little guidance.

LCF does provide an important constraint on what counts as a valid contribution to any of these undertakings. In particular, what counts as a 'model' is under pervasive influence of LCF. A model can mean a computational model in the strict sense, *i.e.* a piece of software or a dedicated digital device or an (artificial) neural net; but in general 'model' can take on an indefinite variety of other meanings, and the LCF is *prima facie* all-important in placing limits on this variety. Acceptable models meet two broad specifications: they treat the ability to be modeled as a particular way of processing *information,* and this processing must be either shown either directly or indirectly to be mechanizable by well-identified or at least conceivable means. Indeed, the first label of what can be retrospectively seen as the beginnings of cognitive psychology was 'information-processing psychology'. However, the sense of mechanism is quite broad, leaving a lot of elbow-room to the modeller. First, there is no requirement that some machine be proposed which would realize or implement the X at hand; it suffices to show that the performance of a case of X can be broken down into sub-tasks or processes Y_1, .., Y_n, each of which is shown, or assumed to be, itself mechanizable in the same sense (with a general clause forbidding a non-ending decomposition). Second, perhaps paradoxically, while the strict computational framework demands that the mechanizable processes be computations in the technical, Turing sense, the models accepted by LCF are just called computational without their proponents necessarily having any clear idea of what a computation is in the technical sense, or finding it at all relevant. To them, computational just means mechanizable in principle in the foregoing sense. An important case in point is the neuroscientific or brain models of cognitive functions: they are informational alright, but they are hardly ever computational in any sense recognizable by a logician or a computer scientist. To some neuroscientists, the further obligation of showing that the brain mechanisms involved in the model are computational in some well-defined sense[53] is important and must eventually be discharged by theoretical or computational neuroscientists, on pain of rejection of the proposed model. But others don't think this is necessarily called

[53] I am leaving aside the difficult matter of deciding whether there is a useful notion of computation which goes beyond the classical theory of recursive (Turing-computable) functions *and* is relevant for cognitive scientific models.

for, and would be loth to reject a model on the sole basis of a failure, on the part of computational neuroscientists, to come up with a plausible computational account.

To sum up, the constraints which LCF imposes on what counts as an acceptable model raise issues of interpretation. These constraints remain quite important nonetheless, as they fix a basic level of theoretical description (the informational level) while maintaining a demand for physically realizable mechanisms. But beyond that, LCF provides not the slightest cue when it comes to attacking any particular mental function, ability or process. The job of the philosopher of cognitive science cannot stop at discussing LCF and its limitations: it must examine the ways in which cognitive scientists, and philosophers of mind working with or among them, seek actual scientific results. This does not imply an exclusive focus on the most local issues (say, non-nutritional sucking *vs* gaze duration as a method of eliciting neonates' perceptual and cognitive abilities, or prototype *vs* exemplar theories of concepts, or ventral vs dorsal visual pathways, or phenomenal *vs* neurophysiological accounts of pain, or weak central control *vs* metarepresentational theories of autism); all scales and levels of generality cry out for examination. The point is that the highest level, where LCF is discussed, is by no means the only one, and as far as descriptive philosophy of cognitive science goes, may not even be the most important one.

In the remainder of this chapter I will try and compress some examples of theoretical and methodological issues at various scales which contribute both to operationalize LCF and to assess its role and legitimacy as a foundation for the field.

b. Theoretical Issues. Some Examples: Modularity, Innateness, Reasoning and Rationality

The single most important and controversial theoretical issue at a level of generality sitting close to, yet below LCF bears on the 'architecture' of the mind. As we saw, one argument in favor of CTM is that there exists a universal Turing machine, an abstract device which can be regarded *post facto* as an idealized model of a programmable computer. One such machine can simulate any other Turing machine, if it is provided with its table (in computer jargon: a programmable computer can compute the results which any given dedicated computer grinds out, once the wiring pattern of the latter is translated into a program fed to the former). Similarly, the universal Turing machine might be regarded as a model for the mind as an information-processing system, which would account for the apparent ability of the mind to perform any imaginable operation, regardless of subject-matter; in other words, the mind, on that view, would be comparable in some respect to a 'general-purpose' computer. This is but the first ingredient in the elaboration of a proposed architecture of the mind which early AI sought to substantiate, and which is also akin to folk-theoretical or traditional views. The mind, under this proposal, is completely impartial as to the tasks which befalls it: whatever learning, memorizing, retrieving, inferring, comparing, etc. are required to perform a task, whether visual, arithmetic, motor, linguistic etc., whether it involves people, cars, animals, abstract entities, etc., they are just the general processes of learning, memorizing, retrieving... which the mind is equipped with and which it deploys as the need arises. In simplistic terms, the mind stands to cognitive tasks somewhat like a pick-up truck stands to moving medium-sized objects: it gets them loaded in one place and carries them in another, and neither the nature of the objects nor the points of arrival and departure make any theoretical difference[54].

It has been argued that this architectural proposal is deeply mistaken. Chomsky has defended a notion of a 'language faculty' clearly individuated within the cognitive system, somewhat like a bladder or an arm are clearly delineated organs of the body. He has argued, more generally, leaning on both logical and empirical considerations, against Piaget and Putnam among others[55], that no purpose-general learning procedure can lead a baby to acquire her mother tongue, and has suggested more generally that the very idea of purpose-general learning is a flawed concept, or one with limited use in the study of basic cognitive abilities. Fodor, in an epoch-making though slim book[56], had proposed to distinguish two broad kinds of cognitive processes. The first, which he called 'input systems', are akin to Franz Gall's faculties, and are accessible to scientific study. The second are 'central processes' which may well be forever hidden from the scientific eye. The distinguishing trait of input systems is their 'modularity'. Fodor drew a list of features which tend to cluster and constitute partial criteria of modularity. The discussion of this list and of various possible understandings of what a module truly is (in particular, how important to the proposal is the possibility, or the demand that a module be *physically* locatable in the brain) has generated an immense literature, which continues to this day.

Characteristically, 'lower' faculties which we share with other animals are modular: the various perceptual modalities and motor control are modular; in addition, linguistic competence (or at least certain dimensions or stages of it) are modular. By contrast, 'higher' faculties which Fodor groups under the heading 'belief fixation' are non-modular. It is worth noting that this division corresponds to a large extent to the personal/subpersonal distinction discussed above (§ 2.e), and also to the boundaries of the folk-psychological domain: as much as we seem to have an untutored grip on 'belief fixation' (how we come to believe what we do believe, in the usual explicit, verbalizable, introspectible sense), we have no clue on how the visual system delivers, on the basis of the retinal array, intelligible scenes such as an unknown human face perched on top of a human bust reclining in an armchair, or the bus arriving at the stop; or on how, from a string of air vibrations, we instantly infer that we are asked to pass the salt, and so on. Thus Fodor's position in the two-camp situation described at the end of §2.e is this: psychology is free to try both the subpersonal reductive and the personal non-reductive approaches, but only the first seems to work, and then only for topics where there is no personal-level processual phenomenology, just personal-level end products. (In Bain's vocabulary, there is just one big "hop, skip and jump" between the 'pure' stimuli, inaccessible from the first-person viewpoint, and the conscious perception or understanding).

---

[54] The metaphor respects relatively obvious features of the original: a pick-up truck cannot carry mount Everest or the Gange, it cannot cross oceans, etc. The mind balks at calculating the 10000th decimal of pi, stops short of parsing a sentence a million words long, and cannot deploy a winning strategy for chess, despite the fact that it exists. Technically, this sort of limitation is abundantly studied under the rubric of complexity or feasability. I personally remain unconvinced that it is worth the effort, but mine is definitely a minority view. We will return to this theme in §3.b below.

[55] Piatelli-Palmarini, 1979.

[56] Fodor, 1983.

This pessimistic outlook provoked a reaction from some thinkers engaged in the study of higher processes, and they developed arguments in favor of ‘massive modularity’[57]. They conjecture that modularity is a massive characteristic of the human cognitive system, not one restricted to peripheral and other automatic processes. Their defense generally combines evidence from developmental studies, showing that very small children, and indeed other animal species, exhibit high capabilities in certain specific domains clearly related to higher processes in adults and clearly unlearnt in any normal sense of the word; neuropsychological evidence, showing that certain higher processes are regularly selectively impaired, psychological studies of patterns of performance in normal adults, which tend to show ‘informational encapsulation’ (segregation of bodies of information, preventing a module from exploiting knowledge not belonging to its proprietary ‘data base’); evolutionary considerations; anthropological studies of deep invariants and ethological evidence and speculations about basic cognitive needs of the emerging *Homo sapiens;* and finally conceptual considerations from complexity (the need for the system to overcome the combinatorial explosion).

What emerges as the most important characteristic of massive modularity is ‘domain-specificity’, the conjecture that the human mind applies distinct sub-systems to handle different regions of its environment (different ‘task domains’), from one-one interactions with other people and social behavior to handling middle-sized objects, categorizing living beings, or counting. This raises the following difficult question: Is the mind entirely modular? If not, then aren’t we left, for the non-modular part, with essentially the same problems which have beset the study of higher functions since the beginning of psychology and led Fodor to his pessimistic conclusions? And if the mind is entirely modular, how can one account for its flexibility and its ability to bring together vastly different realms?[58]

The discussion of the central thesis and these related issues is inconclusive at this point[59], but it is a goldmine of questions which are of independent interest. This is an example of an area in which the shoulder-to-shoulder collaboration between philosophers and empirical scientists is most clearly productive.

Massive modularity clearly implicates the issue of nativism, which has a much longer tradition of its own. The question has undergone a quiet revolution as a result of recent work in evolutionary theory, from biologists as well as philosophers of science, advances in developmental psychology, evidence from anthropology and linguistics, progress in neuroscience and in genetics. Perhaps the single most striking fact to emerge from this body of work is that, on the one hand, any simple notion of innateness which might be proposed is inadequate, and that instead several distinct notions may be required –a fact which critics will see as counting against the entire nativist undertaking–, while particular research programs broadly consonant with nativism are thriving[60]. Naturally, nativism is not thereby vindicated, and indeed it runs into harsh attacks[61], in particular from the ‘constructivist’ camp[62], but even

[57] D. Sperber *in* Dupoux, 2001; Hirschfeld & Gelman, 1994; Barkow, Tooby & Cosmides, 1992; Carruthers, Laurence, Stich, 2005.
[58] Sperber, 2005 proposes an answer.
[59] See *e.g.* the first section of Stainton, 2006.
[60] See Carruthers, Laurence, Stich, 2005.
[61] See *e.g.* B. Scholz & G.Pullum, *in* Stainton, 2006, chap.4; Cowie, 1999.
[62] Elman *et.al.,* 1996; Quartz & Sejnowski, 1997.

then the conclusion that there is little worth retaining from the research would be unwarranted: controversy is the norm in science, rarely the sign of a mortal crisis. Unfortunately, this point, though obvious from the history, and present state of mature sciences, is not always understood when it comes to cognitive science, the result being an exaggerated suspicion towards the field[63].

One of the least likely faculties to succumb to domain specificity would seem to be reasoning: the very idea of reasoning seems to imply domain-generality. As a final example of a theoretical issue which concerns to an equal degree psychologists and other cognitive scientists and philosophers, I will say a few words about the psychology of reasoning and the rationality debate. Decades of experimental research have shown that people fail at certain simple reasoning tasks, both deductive and involving uncertainty, despite being put in ideal conditions. Further, insofar as these failures affect the ability to reach good decisions, in real-life conditions, they are sometimes thought to establish our 'irrationality': we are unable to optimally adjust our means to our ends. Finally, it has been more recently shown that by manipulating the context, or cognitive setting, in which the tasks are performed, so as to make them less general and world-detached than the supposedly simpler, leaner original setting, then the performance can rise dramatically. (This summary ignores the considerable differences between deductive and 'inductive' or uncertain reasoning).

A bird's-eye view of the field reveals three major issues of immediate interest to us. Closest to the topic of the present section is domain-specificity. It has been proposed that humans are equipped with a special-purpose 'cheater-detecting' mechanism which allows them to perform quite well on certain deductive tasks, provided the context of the task makes salient a particular cheating condition[64]. This mechanism would have evolved in order to enable stable cooperative arrangements which gave our species a selective advantage in the 'EEA' (environment of evolutionary adaptedness). If this highly speculative, and much disputed hypothesis were confirmed and if other broadly consonant hypotheses, covering other aspects of reasoning, were formulated and supported, one would be in a position to argue that a domain-general ability such as (abstract) reasoning is parasitic on a domain-specific, evolved competence; that the latter is fast and reliable, while the former is slow and prone to failure; and that in some situations a transfer mechanism can allow the fast, reliable, specialized mechanism to trump the slow, mistake-prone general method and procure the needed solution in an area different from the one for which it evolved.

This conjecture is related to two more general ones which have received broader attention. The first is that contrary to what logicians and rationalist philosophers have taught us, reasoning is not based on formal logic, but rather on content-processing manipulations. 'Reasoning without logic' is the slogan brandished by Philip Johnson-Laird, arguably the most influential psychologist of reasoning and the creator of the theory of 'mental models'.[65] A critical minority[66] argues that this is neither shown by the experimental results nor conceptually coherent. However, if we cross the domain-general / domain-specific opposition with the formal / content-driven (or syntactic / semantic) contrast, we get a two-by-two

[63] Compare with the disputes which the notion of gene has occasioned in the last quarter century: they have not resulted in throwing the entire domain of the life sciences, or even of genetics, in disrepute.
[64] Cosmides, 1989.
[65] Johnson-Laird, 1983.
[66] Rips, 1986, Bonatti, 1994; Andler, 1995.

logical space in which the mental models theory (and related proposals) occupy an interesting position, combining domain-general with content-driven. Now LoTH being a syntactic theory of mind, it would appear incompatible with Johnson-Laird's views, but this is hardly the case: Johnson-Laird has written a forceful textbook devoted entirely to a view of psychology based on the assumption that LoTH is true[67]. Similarly, many defenders of massive modularity accept LoTH, which one might not have expected as LoTH sits so well with domain-generality. This only serves to illustrate the point made earlier regarding the intended meaning of LoTH, and the problems cognitive scientists run into when they try to bring it into consonance with their own practice. This is true a fortiori of LCF, which leaves unspecified some degrees of freedom with respect to LoTH. Cognitive scientists are thus justified in leaving it in a semi-interpreted state, expecting its full meaning, and its true worth, to emerge from future developments in cognitive science.

The second important idea to which the cheater-detection hypothesis is linked is that of real-world, as opposed to ideal rationality. While the latter consists in making optimal decisions in all circumstances, including the most difficult from the theoretical standpoint, the former is geared to the set of ecologically likely situations, and produces in probability the best compromise between feasibility and efficacy. The feasibility condition was first proposed by Herbert Simon under the label 'bounded rationality'[68], where it went hand-in-hand with domain-generality[69]. In the framework of a modular architecture of the mind, an assemblage of domain-specific, evolved capacities, efficacy is relativized to particular sets of events or choices which are likely in the creature's ecological niche, and may well not include theoretically worst-case scenarios. It has been shown[70] that very simple ('fast and frugal', or again 'quick and dirty') heuristics can do as well or better than elaborate and cognition-intensive reasoning strategies in certain families of real-world situations. Whether these heuristics are less brittle than the AI-inspired expert systems which made the headlines in the 1980s remains to be shown, but the general principle of taking into account ecological priorities in the overall rating of a given notion of instrumental rationality is worth exploring. It is connected with two yet wider perspectives. One, defended most vigorously by Stephen Stich[71], concerns philosophy and rejects the traditional method of conceptual analysis as a valid way of fixing the meaning or reference of such terms as 'truth' or 'rationality': empirical work is needed to discover what humans, across the globe, take truth or rationality to be, and cognitive-scientific methods, including evolutionary considerations, can help us interpret the results to yield a naturalistic description of such 'thick' normative terms. The second concerns psychology: the rationality debate has focused on the 'bleak implications' interpretation of the empirical data regarding systematic failure on the part of subjects to respond to problems in inductive reasoning (especially, although deductive aberrations are also implicated) according to the norms set by probability calculus and logic. The bleak implications view holds human rational aptitudes to be severely deficient. It contributes to a generally harsh assessment of human capacities, also based in part on a negative evaluation of the role of emotions. The purpose of 'positive psychology' is to discredit the general aim of

[67] Johnson-Laird, 1988.
[68] Simon, 1957.
[69] Cozic, 2005 for a careful analysis of the notion of bounded rationality.
[70] Gigerenzer & Selten, 2002.
[71] Stich,1996 and later publications. See also Nisbett, 2003.

finding faults in human cognitive performance[72]. Humans should be judged, according to this new orientation, not on their performance in laboratory and pen-and-pencil tests, but on their ability to cope with situations and problems arising in the real world, in their daily lives; and this performance again is not to be assessed against the standards of abstract academic thinking, but by reference to success and failure, as perceived by the agents and their peers. Cognitive science is thus invited to step down from its normative pedestal and take a more descriptive stance, somewhat in the way in which philosophy of science, one generation back, took its historical-descriptive turn.

c. Methodological Issues: Paradigms, Levels and Reduction

The third broad family of issues which keeps philosophers of cognitive science and at times cognitive scientists busy concerns the comparative assessment of competing and seemingly incompatible conceptions of what counts as a theory, a model or an explanation in cognitive science. These issues can be framed, and often are framed, as ontological questions, and are discussed alongside such central issues in philosophy of mind as representations, information, intentionality, consciousness, externalism, with which in fact they overlap. From that perspective, one asks, for example, whether the mind really is functionally equivalent to a programmable computer, or rather to a neural net, or again to a dynamical system; or whether the brain really is a biological computer, or some other kind of system. However it seems more profitable to view such questions as methodological. There is indeed no such thing as a biological computer, or an information-processing system, or for that matter a system, whether complex, dynamical or whatnot, independently of a theoretically-driven description. Thus one might as well go straight for questions which are typically meaningful for philosophy of science: How do the various available proposals for describing the mind, for explaining and predicting its dynamics, compare? What are their fundamental tenets, what are their goals, to what extent are they mutually exclusive?

The main contenders in the last twenty years have been the so-called symbolic and connectionist ‘paradigms’. The symbolic, also sometimes called the ‘classical’ paradigm (briefly: ‘classicism’), is best represented by LoTH. The connectionist paradigm is defined by reference to neural nets, to which it stands in roughly the same relation as the Turing machine stands to classicism. The most detailed exposition of connectionism, which comes in many versions, is given in the two-volume collective work entitled *Parallel Distributed Processing: The Microstructure of Cognition,* already mentioned, and more particularly in the seminal paper “On the Proper Treatment of Connectionism”, and subsequent writings of Paul Smolensky[73]; I will use PTC, after the title of his paper, to refer to Smolensky’s formulation and use it as a ‘representative’ of the connectionist paradigm on par with LoTH on the other side[74]. Both are complex mixes of descriptive and prescriptive views of what cognitive science is or should be doing, both are clusters of doctrinal theses concerning the true nature of cognition, its typical operations, the ‘architecture’ of the mind, the proper basic level of

[72] Snyder & Lopez, 2002.
[73] Smolensky, 1988; Smolensky and Legendre, 2005. Smolensky was a member of the PDP Group.
[74] For another, equally important and somewhat different version of connectionism, see Amit, 1989. An illuminating philosophical account of connectionism is provided in Clark, 1989.

explanation, and how it relates to other possibly relevant levels. Both can be extended by making additional choices, such as between 'empiricism', a view often thought to sit better with connectionism, and 'rationalism' (here a form of nativism), better accommodated, it is claimed, by classicism.

PTC is in agreement with LoTH in conceiving of cognition as a material phenomenon best characterized as information processing, where processing is a kind of mechanical computation transforming content-bearing material entities. But whereas LoTH is committed to inference rules operating sequentially on language-like data structures, PTC locates the causal transactions at a level where discrete operations are but limiting cases of continuous operations consisting in registering and modulating activation levels. The system, prompted by some input, runs these operations in a massively parallel fashion, with no central command; it reaches a temporary equilibrium, which is interpreted as the result of the system's processing of the input. In an important family of cases, the entire transaction can be understood as the application of a negotiating procedure, extended over time, between competing 'soft' rules (rules which can be violated under the pressure of contrary demands). Or to use a conceptual scheme drawn from thermodynamics, the system 'seeks' a final state which maximizes the overall 'harmony' between the built-in soft rules embodying the system's knowledge, given the initial exogenous impulse.

The architecture of a PTC cognitive system is modeled after a network of so-called formal neurons, which are akin to the simple threshold automata introduced by McCulloch and Pitts in a paper which launched the connectionist movement in 1943[75]. Devoid of central coordination, and operating in parallel, such networks come in many kinds of configurations, with sometimes important differences between them, but they all differ more radically from Turing machines or their real-world counterparts von Neumann machines (programmable digital computers as we know them).

The third contrast concerns representations, in both their syntactic and semantic nature. First, only LoTH actually uses syntax, in the sense of formal systems, which coincides with the appropriate level of causal analysis of the system's dynamics, while PTC retains only the notion that causal regularities ('form'), without syntax, can be analytically distinguished from 'content'. In PTC, the form is a combination of architectural constraints and numerical parameters, either semi-permanent or transient. Content is distributed over semi-permanent parameters ('weights', playing the role of synaptic weights) and transient parameters ('activations'). In LoTH, content is typically expressible in linguistic terms, although as we have seen LoTH can accommodate 'subdoxastic' content which corresponds to no ordinary concepts or words. In PTC, content is typically 'subsymbolic' or 'subpersonal', and 'personal' content, *i.e.* thinking material used in everyday speech and conscious trains of thought, emerges at the global level of functioning of the entire system.

Fourth, for PTC the typical cognitive process is a form of literal or generalized perception, the identification of a previously registered pattern or the classification of a new pattern as akin to a stored one (in particular, a perceived pattern can be understood as a partial view and the processing consist in its completion). For LoTH, the typical cognitive process is a formal inference from complex symbols which are truth-evaluable (the 'sentences' of the

[75] "A logical calculus of ideas immanent in nervous activity", in McCulloch, 1965; see Anderson and Rosenfeld, 1988; Andler 1992.

language of thought). Thus, PTC seems to support a broadly associative psychology and LoTH a strictly inferential-logical psychology.

Related to the last, one more major difference deserves to be mentioned here. It concerns the place of learning. Neural nets naturally learn from exposure to experience, and store the knowledge thus acquired in the semi-permanent parameters (weights) regulating the mutual influence of the nodes upon one another. This learning process is an integral part of PTC, as shown by the fact that one could not even conceive of a neural net accomplishing some given cognitive task on the basis of 'instructions' received from outside: it has, or so it seems, to learn 'for itself'. This is in stark contrast with LoTH, which countenances models whose cognitive functionality is due to 'knowledge' directly provided by the cognitive scientist, programmer or expert. To be sure, a classical system can be equipped with a learning device (the field of machine learning was mentioned earlier), but the point is that a classical system which has been directly fed the necessary knowledge is indistinguishable from the same system which owes its knowledge to its learning component. A classical system can be understood quite independently from the process by which it has acquired its competence; in fact, the possibility that it might not have literally acquired it, except in the indirect sense of the distal causation of evolution, can be made sense of.

It would thus appear that LoTH and PTC advocate such widely different views of cognition and of what counts as a valid piece of research in cognitive science that at most one can be on the right track.

This has in fact been the understanding of scores of authors, who have defended their preferred approach by attempting to find a fatal flaw in the other, hoping thereby to win by default. Thus friends of PTC (and other brands of neural-net approaches) have found a whole series of lethal faults in LoTH, ranging from its rigidity and brittleness, its separate treatment of 'normal' and 'exceptional' inputs, its inability to naturally deal with context-sensitivity, to biological (both neuro-physiological and neuro-pathological) implausibility. Friends of LoTH have focused on problems affecting connectionist models such as scalability, applicability outside carefully crafted situations, to their purported inability to account for the essentially symbolic, systematic and generative character of thought.[76]

However difficult and often technical these debates have been, they pale in comparison with a more basic issue: Is it the case that (at least) one or the other approach must be wrong? In other words, are they truly competing or are they rather partial views, from different standpoints, at different levels of description, and even possibly of different regions, of one large and complex realm of phenomena, so that they could both conceivably be at least partly right? The second alternative has by now almost become the received view, although for a wide variety of reasons.

It has been asked, first, whether connectionist models are not simply redescriptions, at a lower level of aggregation, of classical models. After all, a computer running a program with a proven cognitive functionality can be described at a level where symbols, central control, sequentiality all disappear. Maybe connectionist nets are mere 'implementations', material realizations seen at an intermediate level of description, of classical models. This view is sometimes thought to be buttressed by formal results to the effect that neural nets can

[76] McDonald & McDonald, 1995 or Horgan & Tienson, 1996.

approximate Turing machines and that conversely, any net can, and in fact usually is simulated on a von Neumann computer. These implementational or reductive interpretations of PTC are ruled out, I believe, by the fact that contents cannot be redescribed in a way compatible with the redescription of the computation processes, but this would require a long and careful examination.[77] In a similar vein, Smolensky[78] has proposed a well worked-out emergentist view in which, roughly speaking, classical models are descriptions in the limit of emerging behaviors of connectionist models. The levels issue is far from settled (and more will be said about it shortly).

Second, for many cognitive scientists, connectionism and classicism are simply providers of heuristics for constructing models, and models are not true or false, but better or worse in the sense of providing a more or less perspicuous mechanical account of the cognitive phenomenon at hand. Many believe that as a general rule, hybrid models are the proper way to go: most cognitive functions, in their view, are carried out by complex assemblages of connectionist and classical subsystems.

But this in turn raises a general issue: How relevant are models for cognitive science? Of course, given the nearly vacuous content of the concept ‘model’ in contemporary scientific parlance, the question thus posed makes little sense. So let us return for a moment to a bygone era, that of artificial intelligence (AI), in the years from 1956 (its official birth date) to the late 1970s. AI then had a dual goal. It wanted to produce a general theory of human cognition, on the one hand, and on the other a model for each and every cognitive function, taken one by one. These goals were interlocked: the general theory inspired the construction of the specific models, and the models confirmed, and helped amend, the general theory. This proposal, reasonable as it may perhaps sound, nevertheless took on a highly unusual character due to the concept of model to which it appealed: a model of a cognitive function (playing checkers, or recognizing edges in a visual array, or proving the Pythagorean theorem, or translating from German to Russian...) was construed as a program, in the computer sense of the word, which allowed a machine to roughly equal the human performance on the given function or, in AI jargon, ‘task’. Thus was born, and philosophically defended, the idea of ‘programs as theories’.[79] Herbert Simon thought of AI and cognitive psychology as essentially identical, and another pioneer, Roger Shank, was fond of saying that AI was nothing but the study of human intelligence.

This view was quickly abandoned, though, for two reasons: first, in most cases, the programs proposed by AI were unintelligible (or provided no additional intelligibility beyond what had gone into their making in the first place); second, they did not perform well. A distinction was thus drawn between information-theoretic models for psychology or linguistics, on the one hand, and AI models for advanced computing, on the other. The general theory was elaborated (by philosophers) to become LoTH, and AI all but relinquished its early ambition of providing the ‘glue’ between machine simulation and psychological explanation: it chose to become an engineering discipline, period. And so cognitive science was left with a blander notion of a model: a psychological explanation of how elementary mechanisms, known or strongly conjectured to be a part of the basic equipment of the mind,

[77] See refs. in note 76. Andler, 1990.
[78] See refs. in note 73.
[79] Simon, 1979.

can be strung together so as to produce a given cognitive function. Interestingly, this did not amount to an outright rejection of AI's initial proposal, which instead was trivialized. Mainstream cognitive psychology subscribes at least to LCF (if not to more stringent versions of functionalism), and thus it agrees with AI about the general nature of the explanation sought, *viz.* a string of computations on content-bearing (informational) units. But while AI was betting that the essential part of the explanation lay in the complexity of the program (the structure of the particular computation), cognitive psychological models hardly refer to that structure, which is usually about as rich and surprising as a bunch of connected watering hoses in an average backyard. The bulk of the explanation lies in the identification of the elementary mechanisms, and the precise delineation of the explanandum; computation is not mentioned at all, except sometimes as a label appended to the model, amounting to no more than 'mechanistically acceptable in view of our present state of knowledge'; finally, *no* empirical support is drawn from the examination of computer simulations.

Connectionism has undergone a differentiation process partly reminiscent of the split between engineering AI and cognitive psychology. The engineering branch, now fully integrated into AI, uses artificial neural nets to simulate a variety of cognitive functions. The psychological branch, best exemplified by PTC, holds on to a strong notion of model which purports to offer both a standard of causal explanation in psychology (a general theory of cognition) and a method for constructing models of specific functions, just like classical AI, and attempts to resist trivialization by deploying, as we saw, a complex strategy resting on the postulation of a distinct level of description, lying between the 'wetware' and the symbolic level, that populated by typically introspectible and/or linguistically expressible thoughts and transitions. Finally, there is a 'neuroscientific' branch of connectionism, which sees itself as providing models of neural states and processes in the style of mathematical physics; in other words, connectionism in this sense is a style of modeling within theoretical neuroscience.

Whether these enterprises are essentially identical, complementary or largely independent is open to debate. However, the lesson which can be drawn from the earlier phase of cognitive science, where classical AI and LoTH held sway, is that a large part of the actual work on the various cognitive processes and faculties, at whatever level, is unaffected by the outcome: engineers will spend most of their energy on statistical analyses of data and of network performance and learning algorithms; psychologists will work on subjects' performance, employing the tools of empirical psychology, developmental psychology and neuropsychology; and theoretical neuroscientists will be looking at the fine details of certain neural processes and applying advanced modeling methods from statistical physics.

However, cognitive science has entered a new phase in the last decade, which can be seen as the third, coming after the pioneering phase dominated by the classical paradigm and the theoretical role of AI (mid 1950s to late 1970s), and the developing phase where psychology, allied to philosophy, played the leading role, while connectionism competed with AI as the dominant modeling approach. In much the same way as AI during Phase I, cognitive neuroscience in Phase III claims the central role, offering as main source of empirical evidence its proprietary method of functional brain imagery. The 'programs as theories' view of a cognitive-scientific model has been replaced by a 'dynamic map of the brain' conception, whereby once the areas involved in the production of a given (overt of covert) behavior have been identified, together with the flow chart of their activations, one has reached the desired understanding. Thus stated, the view raises several objections. First, there is an 'explanatory gap' between a brain map, whether static or dynamic, and the cognitive process or function which is concomitant with the imaged brain activity, just as there was an explanatory gap between an AI program and the mental function it was simulating. The gap can only be bridged by psychology and same-level disciplines. Second, the brain mapping comes in once the basic structure of the explanandum has been uncovered by psychological, linguistic, or other empirical work, buttressed by conceptual analysis, just as an AI program could be written (or should have been written) only after psychology and others has cleared the ground. In short, functional imagery is neither the beginning nor the end of the process of discovery.

This should all seem straightforward[80], and most cognitive neuroscientists are at pains to present functional imagery as a 'tool', albeit one as crucial to cognitive science as the telescope is to astronomy. Cognitive neuroscience in fact is often defined as the blending of psychology and functional neuroscience. But this leaves room for a range of reductionist positions which are in fact incompatible with the view that psychology and other disciplines concerned by cognition can continue to develop according to their own criteria and dynamics. There are three main sources feeding the neuro-reductionist current. The first is eliminativism, the belief that psychological concepts arise from our untutored intuitions and that they should be discarded wholesale in favor of a respectable conceptual structure drawn entirely from natural science[81]. The second is fundamentalism, the belief that there is a unique ultimate level of description which captures the causal order and is thus the only literal truth-bearer, and to which neurobiology is closer than psychology. And the third is mechanicism, understood as a conception of the aim of inquiry (at least in the 'special sciences') as to provide mechanical models (in the contemporary sense of mechanism). There are a moderate and a radical form of neuro-reductionism. The moderate form recognizes as basic for cognitive science an integrative or emerging level which is relatively insulated from the cellular, molecular and lower levels, just as classical functionalism sees the informational level as relatively insulated from the physical level. The radical form[82] sees no reason to set a principled limit to the search for ever more basic explanations, and makes the (controversial) case that cellular-molecular models are already at hand for some basic cognitive functions, showing the way to a reduction of the entire domain of cognition.

Whatever the merits of these views as visions of tomorrow's cognitive science, as interpretations of the current state of the field they run against two uncontrovertible facts. The first is that there are few, if any, elucidations of a cognitive phenomenon which can be stated as a self-contained neuroscientific theory or 'model'. The second is that there are many important phenomena which we simply would not know how to attack as neuroscientific problems. This does not tell against the importance of neuroscience for cognitive science, or of imagery as an essential new resource; what is at issue here is the idea that neuroscience is wedded to *the* natural basic level of cognition and that it is therefore fated to produce a complete, self-standing theory of the mind. And this idea in turn is connected to the dubious

[80] Andler 2005; Ravenscroft, 1998, or for an all-out attack on the explanatory pretensions of cognitive neuroscience, Bennett and Hacker, 2003.
[81] Note that naturalism need not subscribe to eliminativism.
[82] Defended in particular in Bickle, 2003.

assumption that providing a model is all we need to explain a given cognitive process or function.

Much depends, of course, on what cognitive neuroscience is meant to encompass. It has traditionally been thought of as combining neuroanatomical, neurophysiological, neuropsychological dimensions, as well as the relevant constraints from cellular and molecular neuroscience. Should it also include the relevant part of evolutionary theory? The considerable revival and extension of evolutionary thinking in cognitive science seems to show that, regarding the brain also, the Mayrian duality of proximal and distal causes must be countenanced. But most of the recent work in evolutionary psychology and allied disciplines has little to do with brain science, although of course there are systematic attempts (not all theoretically motivated) to connect the two areas.

The unavoidable conclusion, it would seem, is that cognitive science is for now, and possibly forever, wedded to a form of explanatory and causal pluralism.

d. How Can the LCF Survive in a Hostile World?

Pluralism is a peace-making device, but it comes at a risk. First, of course, the unity of the field is under threat. Perhaps in an age where 'disunity' is seen by many as a fact of scientific life, this should not worry us. But second, the coherence of the entire enterprise is under suspicion: for what grounds are there for hoping that cognitive science will keep holding together conceptually disciplines and approaches which have a long tradition of separatism, if it is forced to renounce its monistic principles?

Cognitive science up until now has overcome these dangers. This it has achieved by containing the pluralistic pull within a monistic framework, while at the same time developing a federalist practice and ethos. The framework was initially provided by the CTM, which was then gradually liberalized to LCF so as to make room, on the one hand, for fairly precise alternatives to classicism (as explicated in LoTH) such as PTC and other varieties of connectionism, and on the other hand for the countless research programs sharing a family resemblance with one another and with CTM, yet having no clear theoretical connection with it but only a shared understanding of the nature of the problems at hand and of what counts as progress in their resolution. In fact, this is how I propose to characterize, functionally so to speak, what I have been referring to as LCF. LCF views cognition as based on, if not necessarily strictly identical with a set of functionalities of a complex biological system (the central nervous system) within the human organism and by extension within other organisms. Cognitive science is thus seen as a branch of biology, giving rise to a non-threatening structure / function dichotomy, more recently doubled by a distal / proximal duality which is the hallmark of evolutionary thinking. LCF proposes to express regularities at the functional level in terms of information, where information is a place-holder (much like what 'force' was for a long time in physics, or 'gene' in biology) to be gradually filled as cognitive science develops. Whatever scientific concept will eventually be regarded as meeting the specifications, the immense resources of commonsense disciplined by philosophy, logic, linguistics, economics and other sciences of man can meanwhile be tapped to provide partial descriptions and explanatory schemas regarding cognitive dynamics, for they all variously circumscribe areas of informational transactions.

LTC thus affords both a unifying framework and the leeway required by the many enterprises which tend to one or the other of the countless manifestations of cognitive activity without committing themselves to any particular theoretical option regarding cognition *in toto.* An imperfect historical parallel is provided by the physical sciences in the 18th century, which all accepted the notional umbrella of Newtonianism, yet felt free to develop their proprietary ontologies and practices without feeling the need to actually connect with Newton's dynamics[83]. The parallel is imperfect because chemistry, electricity, magnetism, heat, light were a small number of well-structured complex domains, with relatively weak motivation to achieve unity, while the various areas within cognitive science are many and are much more aware of the need for integration within a respectable field.

So how is the integration supposed to work? The answer lies in the interdisciplinary practices developed over the half-century of cognitive science's existence: practices rather than overarching explicit principles, and interdisciplinary in the sense that any proposal regarding one particular phenomenon under one particular description is offered for discussion to neighboring subfields, and under pressure to achieve compatibility and if possible full articulation with the other descriptions and the other phenomena implicated[84]. In a sense made familiar by Kuhn and his followers, this is as much a matter of tradition and tacit understanding of what counts as a solution, as an articulation, etc. as one of explicit methodology. But to anyone familiar with the best and most characteristic cognitive science, there is a clear sense of a particular and fruitful search for consilience at work.

However, as we know from democracy, what works with small communities in the early phases of a process may not extend to larger populations over the long term. Two related trends have become apparent in cognitive science in the last decade. One is the exponential growth not only of the field as a whole but of subfields and even subcultures within it, some with strong connections very far from the original core. Thus we have witnessed the appearance of many topics which were not part of the original agenda of cognitive science (consciousness, emotions. self, culture, norms...), and at a higher level of aggregation the rise of the 'affective sciences', 'consciousness studies', 'action theory', and other branches with no fixed name as yet, dealing with rationality, esthetics, ethics, human sociality, culture.... As these new clusters grow, they reach the size and complexity which was that of cognitive science as a whole a mere quarter century back, and necessarily move away from one another to find the breathing space needed for their development. It simply ceases to be feasible, or perhaps even useful, for the individual scientist specializing in one or the other of these new areas to maintain a strong connection with cognitive science as a whole. Perhaps cognitive science is on the verge of becoming an idle structure, like the Roman Empire toward the end, ripe for elimination.

The second trend is the multiplication of proposals which go against LCF, either directly or by implication. In fact, many philosophers of cognitive science these days are occupied with either perfecting, or evaluating the claims of a critical proposal of that sort. Readers

[83] Harman, 1982.
[84] This is shown inter alia by the fact that one does not often find in the cognitive science literature immunizing clauses such as 'X *as understood in the field I work in*', which prevent anyone looking at X from another angle to object. This is in stark contrast with a large majority of research traditions in the human and social sciences. Note however that generative linguistics and its subfields make extensive uses of such clauses. This is not the place to ask why.

familiar with the field may in fact ask why so much of the present chapter is devoted to presenting and discussing LCF, in terms which are not very different from those which would have been appropriate twenty years ago, given that it has been so severely criticized, leaving it, one would think, no other fate than rejection or obsolescence.

It might appear that the only goal worth pursuing at this juncture is to evaluate those attacks and to reach a considered judgment on whether they succeed or fail. Although I agree that it is a goal worth pursuing, I don't believe it is the only one, and in fact I doubt whether it really is a goal rather than a horizon. My aim in these concluding lines is rather to assess the situation resulting from the historical fact that LCF is battered from all sides yet has not fallen into oblivion nor been replaced.

From the start, LCF came under fire: Dreyfus thought of it as the ultimate expression of the rationalist tradition in philosophy, and thus both highly plausible in the contemporary context and deeply wrong[85]; Searle has tried to show that it is absurd and based on a massive mistake. A little later, Putnam, who put it on the map of analytic philosophy, rejected it, and by now has spent more time and ink criticizing it than he did setting it up; Chomsky, whose contribution is seen by many as having been decisive for LCF, is critical of the way it is construed today[86]. Just as dangerous, if not more, for LCF are the indirect threats posed by research programs predicated on assumptions which, if true, would render LCF either false or irrelevant; examples of such programs abound, ranging from the more general (dynamicism, neo-Gestalt and phenomenological approaches, constructivism, neuronalism, radical externalism, and other proposals which either reject or profoundly alter the key LCF concept of mental representation) to the more local (having to do, for example, with emotions, motivation, action, perception, consciousness, memory, norms, the body,...).

Yet LCF still plays a role in the debate. It is a fact that no discussion of the foundations and structure of cognitive science can proceed without first setting the stage, and LCF does exactly that. Nor does it serve as a mere historical preliminary, for if it were the case, LCF could be replaced by its current successor, and there is no successor. Simon and Newell proposed a version of LCF, the 'physical symbol system hypothesis'[87] and claimed for it the status of a 'law of qualitative structure', comparable to Pasteur's germ theory of disease. But if this is right and the LCF is wrong, then the entire field collapses: if illnesses were not typically caused by germs, then the germ hypothesis would have been rejected and there would remain no such thing as the medicine of infectious diseases; instead, we would have other research programs in medicine, based on our current best theory of the causes of illness.

We have something of a puzzle on our hands, one which may remind philosophers of science of the one posed by the status of 'the Legend', Kitcher's name for a view of science which was by and large adopted by the great thinkers who set up philosophy of science as an academic field (at least in the United States), and has been under such systematic attack for the last 40 years or so that it can no longer pass as an acceptable first shot, to be straightened out by suitable amendments[88].

---

[85] Interestingly, Dreyfus's critique of the radical version of rationalism offered by AI and the then nascent cognitive psychology echoes Neurath's attack on 'hyperrationalism' (Neurath, 1983; original papers from 1922).
[86] Dreyfus, 1972; Searle, 1992; Putnam, 1988; Chomsky, 2000, 2002.
[87] Newell and Simon, 1976.
[88] Kitcher, 1995.

In the case of cognitive science and the structuring or foundational role of LCF, there seem to be three main ways of dealing with the puzzle.

The first is to accept Simon and Newell's idea of the physical symbols system hypothesis as a structural hypothesis, but to grant that it has since co-opted many of its opponents' ideas and is now considerably more complex, and looser, than it was. The worry here is: How much structuring can be achieved by an ever looser LCF, one which accepts just about any amendment, caveat and heretical thesis which one can think of? Some defenders of LCF will insist on the other hand that LCF is far less diluted by the amendments and far less threatened by the attacks than an excited lot would like to think.

The second option is to deny that LCF plays any role today, other than loosely delimiting the subject-matter of cognitive science, and that an alternative foundation is gestating. For a long time, I defended the view that LCF was something like a rocket used to launch a satellite (cognitive science), only to be discarded once a certain altitude has been reached. But if this were right, we could talk about cognitive science without referring to LCF just as one can talk, as in fact one does, about contemporary physics without any mention of Descartes' mechanical philosophy, although the latter was instrumental in putting physics on its present track. Note that it may soon appear to be the case, but the necessary recasting of the field, if it is possible, remains to be done.

The third way out is to give up entirely on the need for a structural hypothesis. The usual argument is to compare cognitive science with biology. Biology thrives without the help of a structural hypothesis, and it thrives despite being highly diverse and without any prospect of a grand unification; biology just is the systematic study of a million different phenomena, loosely connected by transverse themes such as the cell, the gene, evolution and the basic features of most living systems (metabolism, growth, homeostasis, etc.). This line however raises two objections. One is that this view of biology as non-unified and devoid of structural principles, however faithful it may be to the present state of the field, is not necessarily a desirable or permanent feature. The other is that while there is little reason to fear that biology will blow up in many fragments and disappear as a conceptual and institutional entity, this is precisely what may be in store for cognitive science, an outcome which motivates in the first place the inquiry on LCF and its possible replacement. Nothing rules out the possibility that the 21st century will feel no more concerned about the project of *a* science of cognition than the 20th was about the conditions of possibility of *a* science of man. We have been living, some happily, others less so, with an institutional label almost entirely devoid of theoretical content, the sciences of man. One sees with increasing frequency references to the cognitive sciences, or the sciences of cognition, the plural indicating an indifference to, or distancing from any pretension to unity in whatever sense. It may turn out that LCF will have been no more than a philosophers' rational reconstruction of an early phase (or perhaps even just a phase *tout court*) in the history of psychology, linguistics and related disciplines.

This question, and the issues leading into it which were barely scratched here, are a large part of the present agenda of philosophy of cognitive science. By dwelling at some length on the formulation of the problem, I hope to have given a sense of what my colleagues are working on now and which will presumably occupy them for years to come. Their conclusions, based in part on, and determining to some extent what path cognitive science will end up following, will be of paramount importance to the sciences of man.

BIBLIOGRAPHY

Amit, Daniel (1989), , *Modeling Brain Function — The world of attractor neural networks*, Cambridge, Cambridge University Press, 1989.

Anderson, James A. & Rosenfeld, Edward (eds.) (1988), *Neurocomputing. Foundations of Research,* Cambridge, MA, MIT Press.

Andler, Daniel (1990), "Connexionnisme et cognition: à la recherche des bonnes questions", *Revue de Synthèse,* série générale **CXI**, nº 1-2, janv.-juin 1990, p. 95-127.

— (1992), "From paleo to neo-connectionism ", *in* G. van der Vijver (ed.), *New Perspectives on Cybernetics,* Kluwer, Dordrecht, p. 125-146.

— (1995), "Logique, raisonnement et psychologie", *in* J. Dubucs & P. Lepage (eds.), *Méthodes logiques pour les sciences cognitives,* Paris, Hermes, p. 25-75.

— (1998), "Turing: Pensée du calcul, calcul de la pensée", *in* F. Nef & D. Vernant, *Les années 1930: réaffirmation du formalisme,* Paris, Vrin.

— (ed.) (2004), *Introduction aux sciences cognitives,* 2nd ed., Paris, Gallimard.

— (2005), "Les neurosciences cognitives: une nouvelle 'nouvelle science de l'esprit'?", *PSN – Psychiatrie, sciences humaines, neurosciences*, 3, 12, p. 74-87.

— (2006a), "Cognitive Science", in L. Kritzman (ed.), *The Columbia History of Twentieth Century French Thought,* New York, Columbia University Press, p. 175-181.

— (2006b) "Phenomenology in cognitive science and artificial intelligence", *in* H. Dreyfus & M. Wrathall (eds.), *A Companion to Phenomenology and Existentialism,* Oxford, Blackwell, p. 377-393.

Bain, A. (1893), "The respective spheres and mutual helps of introspection and psychophysical experiment in psychology", *Mind*, 2, p. 42–53.

Barkow, Jerome, Cosmides, Leda, Tooby, John (eds.) (1992), *The Adapted Mind. Evolutionary Psychology and the Generation of Culture,* New York, Oxford University Press.

Bechtel, W, & Graham, G. (eds.) (1999), *A Companion to Cognitive Science*, Oxford, Blackwell.

Bennett, M.R., Hacker, P. M. S. (2003), *Philosophical Foundations of Neuroscience,* Oxford, Blackwell.

Bermúdez, J. L. (2003), *Thinking Without Words*, Oxford, Oxford University Press.

— Marcel, A., Eilan, N. (eds.) (1995), *The Body and the Self,* Cambridge, MA, MIT Press.

Berthoz, Alain (1997), *Le sens du mouvement,* Paris, Odile Jacob. *The Brain's Sense of Movement,* Cambridge, MA, Harvard University Press, 2002.

Berthoz, Alain (ed.) (1999), *Leçons sur le corps, le cerveau et l'esprit,* Paris, Odile Jacob.

Bickle, J. (2003), *Philosophy and Neuroscience. A Ruthlessly Reductive Account,* Dordrecht, Kluwer.

Block, Ned (1995), "On a Confusion about a Function of Consciousness", *The Behavioral and Brain Sciences,* 18, 2, p. 1995, p. 227-247. Reprinted in Block, Flanagan and Güzeldere, op.cit; and in Block (2007)

— (2007), *Consciousness, Function, and Representation. Collected papers, vol. 1,* Cambridge, MA, MIT Press.

Flanagan, O., Güzeldere, G. (eds.) (1997), *The Nature of Consciousness*. Cambridge, MA: MIT Press.

Bonatti, Luca (1994), "*Propositional reasoning by model?", Psychological Review,* 101, 4, p. 725-733.

Braddon-Mitchell, D. & Jackson, F. (1996), *Philosophy of Mind and Cognition,* Oxford, Blackwell.

Brook, Andrew (1994), *Kant and the Mind,* Cambridge, Cambridge University Press.

Carruthers, Peter, Laurence, Stephen and Stich, Stephen (eds.) (2005), *The Innate Mind,1: Structure snf Content,* New York, Oxford University Press.

Chalmers, David (1996), *The Conscious Mind,* New York, Oxford University Press.

— (ed.) (2002), *Philosophy of Mind: Classical and Contemporary Readings,* Oxford, Oxford University Press

Changeux, Jean-Pierre (2003), *L'homme de vérité,* Paris, Odile Jacob,

— (1983), *L'homme neuronal,* Paris, Fayard. *Neuronal Man: The Biology of Mind*, translated by Laurence Garey, Princeton, Princeton University Press, 1997.

Chomsky, Noam (1965), *Aspects of the Theory of Syntax,* Cambridge, MA: MIT Press.

— (2000), *New Horizons in the Study of Mind and Language,* Cambridge, Cambridge University Press.

— (2002), *On Nature and Language,* Cambridge, Cambridge University Press,

Clark, Andy (1989) *Microcognition. Philosophy, Cognitive Science, and Parallel Distributed Processing*, Cambridge, Mass., MIT Press.

Cosmides, L. (1989), "The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task", *Cognition,* 31, p. 187-276.

Cowie, F. (1999). *What's within? Nativism reconsidered,* Oxford, New York, Oxford University Press.

Cozic, Mikaël (2005), "Fondements cognitifs du choix en rationalité limitée", PhD dissertation, Université de Paris-Sorbonne (Paris IV)

Craig, E., ed. (1998), *Routledge Encyclopedia of Philosophy*, London, Routledge.

Davidson, R.J., Scherer, K.R., Goldsmith, H.H. (eds.) (2003), *Handbook of Affective Sciences,* Oxford, Oxford University Press.

Davies, Martin (2005), "Cognitive science", in F. Jackson and M. Smith (eds.), 2005, p.358-394.

Dennett, Daniel C. (1969), *Content and Consciousness,* London, Routledge and Kegan Paul.

— (1978), *Brainstorms*, Montgomery, VT, Bradford Books.

— (1986), "The Logical Geography of Computational Approaches", in M. Brand and R. M. Harnish, (eds.), *The Representation of Knowledge and Belief*, p. 59-79. Tucson, University of Arizona Press.

Dretske, Fred (1981), *Knowledge and the Flow of Information* , Cambridge, MA, MIT Press.

Dreyfus, Hubert L. (1972), *What Computers Can't Do,* New York, Harper & Row; augm. edition: *What Computers* Still *Can't Do*, Cambridge, MA, MIT Press, 1993.

— Hall, Harrison (eds.) (1982), *Husserl, Intentionality, and Cognitive Science,* Cambridge, MA, MIT Press.

Dupoux, Emmanuel (ed.) (2001), *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler,* Cambridge, MA, MIT Press.

Elman, J., Bates L.E., Johnson, M.H., Karmiloff-Smith, A., Parisi, D. & Plunkett, D. (1996), *Rethinking Innateness. A Connectionist Perspective on Development,* Cambridge, MA, MIT Press.

Enfield, N.J. and Levinson, Stephen C. (eds.) (2006), *Roots of Human Sociality. Culture, Cognition and Interaction,* Oxford, Berg.

Fodor, Jerry (2000), *The mind doesn't work that way. The scope and limits of computational psychology,* Cambridge, MA, MIT Press.

— (1983), *The Modularity of Mind,* Cambridge, MA, MIT Press.

— (1975), *The Language of Thought,* New York, Thos. Crowell; repr. Cambridge, MA, Harvard University Press.

Gardner, H. (1985), *The Mind's New Science*, New York, Basic Books.

Gazzaniga, M. (2000), *The New Cognitive Neurosciences,* 2nd ed., Cambridge, MA, MIT Press.

Gigerenzer, Gerd & Selten, Reinhard, eds. (2002), *Bounded Rationality: The Adaptive Toolbox,* Cambridge, MA, MIT Press.

Guttenplan, Samuel, ed. (1994), *A Companion to the Philosophy of Mind,* Oxford & Cambridge, MA, Blackwell.

Harman, P.M. (1982), *Energy, Force, and Matter,* Cambridge, Cambridge University Press.

Hatfield, Gary (1995), "Remaking the science of the mind. Psychology as natural science", in C. Fox, R. Porter, R. Wokler (eds.), *Inventing Human Science. Eighteeth-Century Domains,* Berkeley & Los Angeles, University of California Press.

— (1997), "Wundt and psychology as science: Disciplinary transformations", *Perspectives on Science,* 5 3, p. 349-382.

— (2002), "Psychology, philosophy, and cognitive science: Reflections on the history and philosophy of experimental psychology", *Mind and Language,* 17 3, p. 207-232.

Haugeland, John (1998), *Having Thought,* Cambridge, MA, Harvard University Press.

Heil, John (ed.) (2004), *Philosophy of Mind: A Guide and Anthology,* New York, Oxford, Oxford University Press.

Herken, R. (ed.) (1988), *The Universal Turing Machine. A Half-Century Survey,* Oxford, Oxford University Press.

Hirschfeld, Lawrence A. & Susan A. Gelman, (eds.) (1994), *Mapping the Mind. Domain Specificity in Cognition and Culture,* Cambridge: Cambridge University Press,

Hofstadter, Douglas (1985), *Metamagical Themas,* New York, Basic Books.

Horgan, T. and Tienson, J. (eds.) (1996), *Connectionnism and the philosophy of psychology,* Cambridge, MA, MIT Press.

Imbert, Michel (2006), *Traité du cerveau,* Paris, Odile Jacob.

Jackson, F. & Smith, M., eds. (2005), *The Oxford Handbook of Contemporary Philosophy,* Oxford, Oxford University Press.

Jacob, Pierre (1997), *What Minds Can Do,* Cambridge, Cambridge University Press.

Jeannerod, Marc (1997), *The Cognitive Neuroscience of Action,* Cambridge, Cambridge University Press.

Johnson, David and Erneling, Christina (eds.) (1997), *The Future of the Cognitive Revolution,* New York, Oxford, Oxford University Press.

Johnson-Laird, Philip N. (1983), *Mental models,* Cambridge, Cambridge University Press.

— (1988), *The Computer and the Mind. An Introduction to Cognitive Science*, Cambridge, MA, Harvard University Press.

Kahneman, D., Slovic, P. & Tversky, A. (1982), *Judgment Under Uncertainty: Heuristics and Biases,* Cambridge, Cambridge University Press.
Kitcher, Patricia (1990), *Kant's Transcendental Psychology,* New York, Oxford, Oxford University Press.
Kitcher, Philip (1995), *The Advancement of Science: Science Without Legend, Objectivity Without Illusions,* New York, Oxford, Oxford University Press.
Levine, J., (1983), “Materialism and qualia: The explanatory gap”, *Pacific Philosophical Quarterly,* 64, p. 354–61.
Livingston, Paul (2004), *Philosophical History and the Problem of Consciousness,* Cambridge, Cambridge University Press.
Lucas, J.R. (1961), "Minds, machines and Gödel", *Philosophy,* 36, p. 112-127; repr. in Alan R. Anderson (ed.), *Minds and Machines,* Prentice Hall, 1964.
Ludwig, K., "The Mind-Body Problem: An Overview," in Warfield & Stich (2003), p. 1-46.
Macdonald, Cynthia and Macdonald, Graham (eds.) (1995), *Connectionism,* London, Blackwell.
McCulloch, Warren S., (1965), *Embodiments of Mind,* Cambridge, MA, MIT Press.
Marcel, A. J. (1983) “Conscious and unconscious perception: Experiments on visual masking and word recognition”, *Cognitive Psychology*, 15, p. 197-237.
Millikan, Ruth (1984), *Language, Thought and Other Biological categories*, Cambridge, MA, MIT Press.
Morton, P. (ed.) (1996), *A Historical Introduction to the Philosophy of Mind: Readings with Commentary,* Calgary, Alberta, Broadview Press.
Nadel, L, gen. ed., (2003), *Encyclopaedia of Cognitive Science*, London, New York & Tokyo, Nature Publishing Group
Nagel, T. (1974) “What Is It Like To Be a Bat?”, *Philosophical Review*, 83 4, p. 435-450
Neurath, Otto (1983), *Philosophical papers, 1913-1946,* ed. and transl. by Robert S. Cohen and Marie Neurath, Dordrecht, D. Reidel.
Newell, Allen & Simon, Herbert (1976), “Computer Science as Empirical Inquiry: Symbols and Search”, *Communications of the ACM,* 19 3, p. 113-126.
Nisbett, Richard E. (2003), *The Geography of Thought,* New York, The Free Press.
Papineau, David (1987), *Reality and Representation*, Oxford, Basil Blackwell.
Parot, Françoise and Richelle, Marc (eds.) (1992), *Psychologues de langue française,* Paris, PUF.
Petitot, J., Varela, F.J., Pachoud, B., Roy, J.-M. (eds.) (1999), *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science,* Stanford, Stanford University Press.
Piattelli-Palmarini, Massimo (ed.) (1979), *Théories du Langage, Théories de l'Apprentissage, le Débat Chomsky/Piaget,* Paris, Seuil; *Language and Learning: The debate between Jean Piaget and Noam Chomsky,* London, Taylor & Francis, 1983.
Port, R.F., & van Gelder, T. (eds.) (1995), *Mind as Motion: Explorations in the Dynamics of Cognition,* Cambridge, MA, MIT Press.
Putnam, Hilary (1960), “Minds and Machines”, reprinted in Putnam (1975).
— (1975), *Mind, Language, and Reality,* Cambridge, Cambridge University Press.
— (1988), *Representation and Reality*, Cambridge, MA, MIT Press.
— (1994), Entry “Putnam, Hilary” *in* Guttenplan (1994).
Quartz, Steven R. and Sejnowski, Terrence J. (1997), “The neural basis of cognitive development: A constructivist manifesto”, *Behavioral and Brain Sciences*, 20, 4, p. 537-596.
Ravenscroft, Ian (1998), “Neuroscience and the mind”, *Mind and Language,* 13, p. 132-137.
Rips, L.J. (1986), “Mental Muddles”, in M. Brand & M. Harnish, eds., *The Representation of Knowledge and Belief*, Tucson, University of Arizona Press, p. 258-86.
Rosenthal, David M. (ed.) (1991), *The Nature of Mind,* New York, Oxford University Press.
Rowlands, Mark (2003), *Externalism,* Montreal & Kingston, McGill-Queens University Press.
Rumelhart, D., McClelland, J. & the PDP Research Group (1986), *Parallel Distributed Processing. The microstructure of cognition,* Cambridge, MA, MIT Press.
Searle, John R. (1992), *The Rediscovery of Mind,* Cambridge, MA, MIT Press.
Simon, Herbert (1957), "A Behavioral Model of Rational Choice", in *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*, New York, Wiley.
Simon, T. W. (1979), “Philosophical objections to programs as theories”, in M. Ringle, (ed.), *Philosophical Perspectives in AI*, Atlantic City, NJ, Humanities Press.
Smith, David W. and Thomasson, Amie L. (eds.) (2005), *Phenomenology and Philosophy of Mind,* New York, Oxford University Press.
Smolensky, Paul (1988), 'On the proper treatment of connectionism', *Behavioral and Brain Sciences,* 11, p. 1-74.

— & Legendre, Géraldine (2005), *The harmonic mind: From neural computation to optimality-theoretic grammar,* vol. 1: *Cognitive Architecture*; vol. 2: *Linguistic and Philosophical Implications,* Cambridge, MA, MIT Press.

Snyder, Charles R., Lopez, Shane J. (eds.) (2002), *Handbook of Positive Psychology,* New York, Oxford University Press.

Sperber, Dan (1996), *La Contagion des idées,* Paris, Odile Jacob; *Explaining Culture: A naturalistic approach,* Oxford, Blackwell, 1996.

— (1994), "The modularity of thought and the epidemiology of representations", *in* L. Hirschfeld & S. Gelman, eds, *op. cit.*

— (2005), "Modularity and relevance: How can a massively modular mind be flexible and context-sensitive?", in Carruthers, Laurence, & Stich, *op. cit.*

Stainton, Robert J. (ed.) (2006), *Comtemporary Debates in Cognitive Science,* Oxford, Blackwell.

Stich, Stephen (1996), *Deconstructing the Mind*, New York, Oxford University Press.

Taylor, Charles (2005), *Philosophical Papers,* Cambridge, Cambridge University Press.*,*

Thom, René (1988), *Esquisse d'une sémiophysique,* Paris, Interéditions.

Tomasello, Michael (2001), *The Cultural Origins of Human Cognition,* Cambridge, MA, Harvard University Press.

Turing, Alan M. (1950), "Computing machinery and intelligence", *Mind*, LIX, 236, p. 433-460,

Velmans, Max and Schneider, Susan (eds.) (2007), *The Blackwell Companion to Consciousness,* Oxford, Blackwell.Warfield, T. and S. Stich, S. (eds.) (2003), *The Blackwell Guide to the Philosophy of Mind*, Oxford, Blackwell.Warner, Richard and Szubka, Tadeusz (eds.) (1994), *The Mind-Body Problem,* Oxford, Blackwell.

Warrington, E.K., Weiskrantz, L. (1968), "New method of testing long-term retention with special reference to amnesic patients", *Nature*, 217, p. 972-974.

Weiskrantz, L. (1986), *Blindsight: A Case Study and Implications,* Oxford, Oxford University Press.

Wilson, R. A. & Keil, F.C. (eds.) (1999), *The MIT Encyclopedia of the Cognitive Sciences,* Cambridge, MA, MIT Press.