PREFACE

Cardinal Mazarin's librarian had a low opinion of history books. In what time he could spare from his master's collection of 40,000 volumes--opened to the public in 1644, Thursdays only! (Clarke 1970)--Gabriel Naude wrote some brief tracts himself. One was the first-ever book on library science (Naude 1627/1644). Another, less well-known today, was a cry of outrage against historians (Naude 1625/1657).

His specific accusation was that they'd maligned the growing band of automata makers as dangerous dabblers in illicit magic, instead of recognizing them as brave pioneers of the mathematical arts. His general complaint was that the authors of history books are "a sort of people seldome or never representing things truly and naturally, but shadowing them and making them according as they would have them appear" (1625/1657: 9).

That was a spot too fierce. ("Seldome"? "Never"??). After all, Naude believed that his own history of automata making was close to the truth. But his basic point was correct. Every history is a narrative told for particular purposes, from a particular background, and with a particular point of view.

Someone who knows what those are is in a better position to understand the story being told. This Preface, then, says what this history aims to do, and outlines the background and viewpoint from which it was written.

i: The Book

This is a historical essay, not an encyclopaedia: it expresses one person's view of cognitive science as a whole. It's driven by my conviction that cognitive science today--and, for that matter, tomorrow--can't be properly understood without a historical perspective. In that sense, then, my account describes the field as it is *now*. It does this in a second sense too, for it features various examples of state-of-the-art research, all placed in their historical context.

Another way of describing it is to say that it shows how cognitive scientists have tried to answer myriad puzzling questions about minds and mental capacities. These questions are very familiar, for one doesn't need a professional licence to raise them. One just has to be intellectually alive. So although this story will be most easily read by cognitive scientists, I hope it will also interest others.

These puzzles are listed at the opening of Chapter 1. They aren't all about "cognition," or knowledge. Some concern freewill, for instance. What is it? Do we have it, or do we merely appear to have it? Under hypnosis, do we lose it? Does any other species have it? If not, why not? What is it about

dogs' or crickets' minds, or brains, which denies them freedom? Above all, how is human free choice possible? *What type of system,* whether on Earth or Mars, is capable of freewill?

My account is focussed on ideas, not anecdotes: it's not about who said what to whom over the coffee cups. Nevertheless, the occasional coffee cup does feature. Sometimes, a pithy personal reminiscence can speak volumes about what was going on at a certain time, and how different groups were reacting to it.

Nor does it explore sociopolitical influences at any length, although some are briefly mentioned--for instance, the seventeenth-century respect due to the word of a "gentleman", the twentieth-century role of military funding, and the post-1960 counter culture. In addition, I've said a little about how various aspects of cognitive science reached--or didn't reach--the general public, and how it was received by them. What's printed in the newspapers, accurate or (more usually) not, has influenced the field indirectly in a number of ways--and it has influenced our culture, too.

Mainly, however, I've tried to show how the central ideas arose--and how they came together. To grasp what cognitive science is trying to do, one needs to understand how the multidisciplinary warp and weft were interwoven in the one interdisciplinary field.

My text, too, holds together much as a woven fabric does. It's best read entire, as an integrated whole--not dipped into, as though it were a work of reference. Indeed, I can't resist quoting the King of Hearts' advice to the White Rabbit: "Begin at the beginning, and go on till you come to the end: then stop."

I realize, however, that many readers won't want to do that--though I hope they'll read the whole of Chapter 1 before starting on any of the others. Moreover, even reading a single chapter from beginning to end will typically leave lots of loose ends still hanging. Most of the important topics can't be properly understood without consulting *several* chapters. Freewill, for example, is addressed in more than one place (7.i.g-h, 14.x.b, and 15.vii). Similarly, nativism--alias the nature/nurture debate—is discussed in the context of:

* psychology: Chapters 5.ii.c and 7.vi;
* anthropology: 8.ii.c-d and iv-vi;
* linguistics: 9.ii-iv and vii.c-d;
* connectionism : 12.viii.c-e and x.d-e;
* neuroscience: 14.ix.c-d;
* and philosophy: 2.vi.a and 16.iv.c.

So besides the Subject Index, I've provided many explicit cross-references, to

encourage readers to follow a single topic from one disciplinary chapter to another. Peppering the text with pointers saying "see Chapter *x*" isn't elegant, I'm afraid. But I hope it's useful, as the best I could do to emulate links in hypertext. (The King of Hearts, of course, hadn't heard of that.)

These pointers are intended as advice about what to look at next. They're helpful not least because I may have chosen to discuss a certain topic in a chapter *other than* the one in which you might expect to find it. (The theory of concepts as "prototypes", for example, is discussed in the anthropology chapter, not the psychology one.) My placements have been decided partly in order to emphasize the myriad interdisciplinary links. So no chapter that's dedicated to one discipline avoids mention of several others.

History, it has been said, is "just one damn thing after another". Were that true, this account would be hardly worth the writing. In fact, any history is a constructed narrative, with a plot--or, at least, a reasonably coherent theme.

The plot can always be disputed (hence some of Naude's scorn), and in any case usually wasn't obvious to the *dramatis personae* concerned. Several examples of work experienced at the time as thrilling new beginnings are described here, and with hindsight it's clear that some of them actually *were.* But I'll also describe examples where it looked as though the end had already come--or anyway, where it wasn't known whether/when there'd be a revival. As for future episodes of the story, no-one can know now just what they'll be. I'll indicate some hunches (17.ii-iii), but with fingers firmly crossed.

In the case of cognitive science, *theme* is as problematic as *plot.* The field covers so many different topics that a single theme may not be immediately obvious. At a cursory glance, it can seem to be a hotch-potch of disparate items, more properly ascribed to quite distinct disciplines. Indeed, some people prefer to speak of "the cognitive *sciences*", accordingly (see 1.ii.a).

The key approaches are psychology, neuroscience, linguistics, philosophy, anthropology, AI (artificial intelligence), and A-Life (artificial life)--to each of which, I've devoted at least one chapter. Control engineering is relevant too, for it provides one of the two theoretical 'footpaths' across the many disciplinary meadows of cognitive science (see Chapters 1.ii.a, 4.v-ix, 10.i.g, 12.vii, 14.viii-ix, and 15.viii.c.).

Ignorance of the field's history reinforces this ragbag impression. So does a specialist fascination with particular details. But my aim, here, is to see the wood as well as the trees. I want to help readers understand what cognitive science as a whole is trying to do, and what hope there is of its actually doing it.

Each discipline, in its own way, discusses the mind--asking what it is, what it does, how it works, how it evolved, and how it's even possible. Or, if you prefer to put it this way, each discipline asks about *mental processes* and/or about how the *mind/brain* works. (That doesn't prevent them asking also whether the emphasis on the mind/brain is too great: some say we should consider the mind--or rather, the person—as *embodied,* too. And some add that we should focus on *minds,* not on *mind*: that is, we should remember the essentially social dimension of humanity.)

Moreover, each discipline, insofar as it's relevant to cognitive science, focusses on computational and/or informational answers--whether to recommend them or to criticize them (see Chapter 1.ii).

These questions, and these answers, unify the field. In my view, the best way to think about it is as *the study of mind as machine.* As explained in Chapter 1.ii.a, however, more than one type of machine is relevant here. In a nutshell: some for digital computing, some for cybernetic self-organization or dynamical control. Much of the theoretical—and historical--interest in the field lies in the tension that follows from that fact.

In short, I've tried to give a coherent overview, showing how the several disciplines together address questions that most thinking people ask themselves, at some time in their lives.

Many trees would need to be felled for a fully detailed history of cognitive science, for every discipline would require at least one large volume. The prospect is daunting, the forests are already too empty, and life is too short. This account has a more modest aim: despite its length, it's a thumbnail sketch rather than a comprehensive record.

That means that decisions have to be made about what to mention and what to omit. So my story is unavoidably selective, not only in deciding what research to include but also in deciding which particular aspects of it to highlight.

Some of my selections may surprise you. On the one hand, you may find topics that you hadn't expected. For instance, the psychological themes include emotion, personality, social communication, and the brain's control of movement. (In other words, cognitive science *isn't* just the science of cognition: see Chapter 1.ii.) Other perhaps-unexpected themes include evolutionary robotics, the mating calls of crickets, and the development of shape in embryos. However, all those topics are relevant if one wants to understand the nature of mind.

Moreover, quite a few of the people I discuss aren't in the mainstream. Some have been unjustly forgotten, while others hold views that are (currently) distinctly off-message.

Indeed, some aren't even in a sidestream, since they deny the possibility Of *any* scientific explanation of mind. And some, such as Johann von Goethe, are highly unfashionable to boot. Other authors recounting the history of the field might not mention any of them. Nevertheless, I try to show that they're all relevant, in one way or another. Sometimes, admittedly, it's largely a question of *Know your enemy!* (see the *Aperitif* to Chapter 16). But even one's intellectual enemies usually have things of value to say.

On the other hand, I deliberately ignore some themes and names which you might have expected to encounter. In discussing linguistics, for example, I say almost nothing about phonetics, or about automatic speech processing. These aren't irrelevant, and they figure prominently in more specialist volumes. But the general points I want to stress can be better made by addressing other aspects of language.

Similarly, in my account of cybernetics only a few people feature strongly: Norbert Wiener, John von Neumann, Warren McCulloch, Gordon Pask, W. Grey Walter, W. Ross Ashby, and Kenneth Craik. Others (such as Gregory Bateson and Stafford Beer) are only briefly visible, but might have been featured at greater length. And some bit players don't appear in my pages at all. In a comprehensive volume devoted solely to cybernetics, one could try to mention all of them (see Heims 1991). In a history spanning cognitive science as a whole, one can't.

That space constraint applies in all areas, of course--so please forgive me if I haven't mentioned Squoggins! Indeed, please forgive me if I haven't mentioned someone *much* more famous than Squoggins: the characters in my narrative are numerous enough as it is.

Even those who do appear could have been discussed more fully, so as to do justice to the rich network of formative influences behind any individual's ideas. With respect to the origins of A-Life, for example, I mention the coffee-house conversations of von Neumann and Stanislaw Ulam (Chapter 15.v.a). But just how much credit should be given to Ulam? To answer that question--which I don't try to do--would require many more pages, including a discussion about how sowing an intellectual seed should be weighed against nurturing the developing plant. In short, to detail *every* researcher of any historical importance would be impossible.

Still less could one specify all *current* work. For such details, there are the numerous specialist textbooks--and, better still, the professional journals and conference proceedings. However, I've mentioned a range of up-to-date examples, in order to indicate how much--or, in some cases, how little--has changed since the early days.

Sir Herbert Read once said that whereas the art historian deals with the dead the art critic deals with the living, an even more risky thing to do. Although I've written this book primarily as a historian (which of course involves a critical dimension), I've dipped my toes into the riskier waters of contemporary criticism too. That's implicit in my choices of what recent work to mention, and what to ignore. And in the final chapter, I've said which instances of current work I regard as especially promising. However, those choices are made from what's already a highly selective sample: contemporary cognitive science contains many more strands than I've had space to indicate.

So the bad news is that some things which merit discussion don't get discussed. The good news is that if you find the recent examples I've selected intriguing, you can be sure that there are more. Tomorrow, of course, there will be more still.

ii: The Background

One of the founders of cognitive science expressed Naude's insight in less disgusted terms. As Jerome Bruner put it: “The Past (with a capital letter) is a construction: how one constructs it depends on your perspective toward the past, and equally on the kind of future you are trying to legitimize" (Bruner 1997: 279).

The future I'm trying to legitimize here is one in which interdisciplinarity is valued and alternative theoretical approaches respected--and, so far as possible, integrated.

As for my perspective on the past, this springs from my own experience of the field over the past fifty years. Indeed, it's even longer than that if one includes my reasons for being drawn to it in the first place. For I was already puzzling over some of its central questions in my early-teenage years.

(I was born in 1936. I mention that, and give other researchers' years-of-birth whenever I could discover them, less to record the appearance of particular individuals on this planet than to indicate the passage of intellectual *generations.)*

I first encountered cognitive science in 1957, at the University of Cambridge. I'd just completed the degree in medical sciences there, during which time I'd been especially interested in neurophysiology and embryology.

The medical course was almost uniformly fascinating (although the biochemistry was fairly low on my list of priorities). I remember being intrigued by Lord Adrian's work on spinal reflexes and action potentials, and spellbound by Andrew Huxley's hot-off-the-press lecture on muscle contraction--which had earned him a standing ovation from the usually *blase*

medical students (2.viii.e). Likewise, I'd been amazed by Alan Turing's paper on morphogenesis, and entranced by D'Arcy Thompson's writings on mathematical biology (15.iii-iv).

I now had one year to spare before going--or so I thought--to St. Thomas's Hospital in London. There, I would do my clinical training, as a prelude to a career in psychiatry.

My College expected me to spend the year specializing in neurophysiology, which indeed I found absorbing. And Cambridge was a superb place to do it. Besides the awe-inspiring Adrian-Huxley tradition, exciting new work was being done by Horace Barlow: 14.iii.b. (He was one of my physiology demonstrators: many's the time he helped me to coax a frog's leg to move in a physiology practical.)

But that would have meant doing lengthy experiments on cats, and the comatose rabbits pinned out in my pharmacology practicals had been troubling enough. The neurophysiological experiments that could then be done were fairly broad-brush, since single-cell research had only just begun (2.viii.f). But I don't know that a unit-recording approach would have made much difference to the way I felt. (For a description of what this involves today, see Anderson, Pellionisz and Rosenfeld 1990: 215.) My qualms were largely irrational, of course: not only would I not have felt quite the same about rats, but the cats would be anaesthetized, or decorticate, or even decerebrate. Nevertheless, I hesitated.

As for psychology as an alternative, I'd originally planned to do this in my third year--but the course at Cambridge had turned out to be too rat-oriented, and too optics-based, for my taste. I'd already gate-crashed all the psychopathology lectures, and for six weeks worked as a resident nursing assistant at Fulbourn mental hospital nearby. But mental illness, the psychological topic which interested me most, figured hardly at all in the curriculum. Perhaps that was because precious little could be done to help. (Psychotropic drugs were still a rarity: largactil, a.k.a. chlorpromazine, was being given to schizophrenics on the ward I nursed on at Fulbourn, but that was because the hospital's director was exceptionally forward-looking.)

Moreover, I now knew something I hadn't realized until after my arrival at Cambridge, namely that universities offered degrees in philosophy. This was a revelation. (Without it, I'd probably have ignored my qualms and turned to the cats.)

I'd discovered philosophy while I was still at high school, and found it deeply engaging. I remember reading Bertrand Russell with excitement, cross-legged on the floor in the second-hand bookshops on London's Charing Cross Road. I also remember plaguing several of my schoolteachers with

questions that were philosophical in intent. But I had no idea that one could study philosophy at university.

Now, some five years later, I'd discovered that it was an option available in the final year at Cambridge, after completing the exams in medical sciences. I hadn't lost my love for philosophy, and this seemed to be my one and only chance to do something about it.

So I decided, against all (and I do mean *all*) advice, to spend my interim year studying what was then called Moral Sciences--a label that elicited relentless teasing from my fellow medics. I planned to concentrate as far as possible on the philosophy of mind and of science. And despite opposition from an unimaginative Director of Studies, I insisted on being taught by Margaret Masterman--who was neither a Fellow of Newnham nor a University faculty member, and who was far too original and eccentric to be popular with the College authorities.

I found my philosophical studies so exciting that the "one" year turned into two. Meanwhile, my medical contemporaries and I received our degrees in 1958 from Lord Adrian himself, who was Vice-Chancellor at the time. (We each knelt down with our two hands between his, transfixed--in my case, anyway--by the University's huge golden seal-ring on one of his long, slender fingers.)

During those two years, and alongside some (very different!) supervisions with the logician Casimir Lewy, Masterman taught me weekly at the Cambridge Language Research Unit, or CLRU. This had been founded in 1954--one year before I arrived in Cambridge (and two years before artificial intelligence was named).

The Unit wasn't an official part of the University but an independent, and distinctly maverick, research group directed by Masterman. Most of its funding came from military agencies in the USA (11.i.a). Its home was a small brick building tucked away on 'the other side' of the river. There were apple trees in the garden, and Buddhist gods carved on the big wooden doors. ("The place is full of gods", Masterman had said to me when I first phoned her to ask for directions. I couldn't imagine what she might mean.)

It was an exciting place, and not just because of the gods. Nor even because several members, seeking to combine science and religion, had founded the Epiphany Philosophers. This was a small community for worship and discussion, who met sometimes in a chapel hidden behind a wall upstairs and sometimes in a fenland mill. It was later widely taken as the inspiration for Iris Murdoch's novel *The Bell*. (Murdoch had studied philosophy in Cambridge in 1947-8.) The Epiphany Philosophers notwithstanding, what was most exciting about CLRU was its intellectual diversity and originality.

Masterman's research group in 1957 included a number of people

specializing in the study of language:

* Karen Sparck Jones, now a distinguished researcher in information and language processing (Sparck Jones 1988);
* Richard Richens, a pioneer of machine translation who was by then a senior figure in the Commonwealth Abstracts Bureau (Richens 1958);
* Robin Mackinnon Wood;
* and Frederick Parker-Rhodes, who could read proficiently in twenty-three languages and who (like Masterman) saw metaphorical, not literal, language as primary (Parker-Rhodes 1978).
* Several members of CLRU were then working on automatic Chinese-English translation (Parker-Rhodes 1956; Masterman 1953), helped by Michael Halliday, who became involved with CLRU while Lecturer in Chinese at Cambridge.

Yorick Wilks and Martin Kay, now professors of artificial intelligence and computational linguistics at Sheffield and Stanford universities, joined them very soon after I left.

Another member of the language group at that time was Roger Needham. He was working at the still-new Computer Laboratory at Cambridge, where Maurice Wilkes had built the first relatively easy-to-use computer only a few years before, in 1948/49. Much later, he succeeded Wilkes as its Head, and recently directed Microsoft's UK research laboratory (sadly, he died in 2003). He and his wife Sparck Jones immediately aroused my admiration, for building their house with their own four hands. They were living on-site in a caravan surrounded by mud--hence their well-worn Wellington boots--while also doing high-level intellectual work.

Among the others I encountered in the Unit was physicist Ted Bastin. He and Parker-Rhodes were developing a highly maverick account of quantum theory, with quanta as self-organizing entities. This is now (so I'm told: I can't make head or tail of quantum physics) a standard alternative view, with several web sites devoted to Parker-Rhodes.

In addition, the exceptionally original cybernetician Pask--today, the object of even more numerous web sites--was literally a back-room boy. He was usually hunched over his DIY computer, which he'd cobbled together out of biscuit tins and string.

Last but not least, philosophers Richard Braithwaite--Masterman's husband--and Dorothy Emmet were frequently around. I shared their interest in the philosophy of religion (a subject I later taught for many years) and, above all, in a scientifically grounded philosophy of mind.

Braithwaite--whom I saw more often--was a leading philosopher of science, and also held the Knightsbridge Chair of Moral Philosophy at Cambridge. Much

concerned to integrate science with other areas of life, he'd recently recommended the theory of games as a tool for the moral philosopher (Braithwaite 1955). And he combined a broadly positivistic philosophy of science with Christian beliefs (Braithwaite, 1971)--or rather, with a practical commitment to the moral principles illustrated by Christian stories. (Rumour has it that when called upon to recite the Creed at his public baptism service, his full-voiced "I believe ... " was preceded *sotto voce* by "I will behave in all ways as if ... ".)

Emmet, who'd very recently (1950-53) given the Stanton Lectures in the Philosophy of Religion at Cambridge, held the Chair of Philosophy at the University of Manchester. She knew of the growing excitement about the potential of computing, for the prototype of the world's first stored-program electronic computer had been operational in Manchester since 1948. Indeed, Turing--who wrote some of the first programs for the full version of the Manchester machine--had worked there also. As early as 1949, Emmet's philosophy seminar had discussed "The Mind and the Computing Machine", with Turing present as one of the discussants (16.ii.a).

That's not to say that she was a devoted Turing fan. True, her department had developed an electrical machine for teaching symbolic logic, already in use for some years. Designed by Wolfe Mays and Dietrich Prinz (who was closely involved in the design of the Manchester computer), it had been exhibited at the annual British philosophy conference (Mays and Prinz 1950; Mays, Henry and Hansel 1951). But Mays' device wasn't based on Turing's ideas. Rather, it was inspired by the keyboard-and-rods "Logical Piano", originated in 1869 by Stanley Jevons to illustrate the formal principles of validity--see 2.ix.a. (Jevons had been Professor of Logic at Owens College, the forerunner of the University of Manchester.)

Nor did Emmet and her Manchester colleagues agree with Turing that there was no good reason to deny that some conceivable digital computer could *think.* In the departmental seminar he attended, she'd objected that a machine could not be conscious. Michael Polanyi had added that whereas a machine is fully specifiable, a mind is not. And Mays had argued trenchantly that computers are, as John Searle (1980) would later put it, all syntax and no semantics (n.a. 1949).

In the Cambridge apple orchard, however, Turing's influence was strong (see Chapters 4.i and 16.ii). He'd died in 1954, only one year before my arrival in Cambridge. And he'd been close to Braithwaite. Soon after the publication of Turing's seminal paper in *Mind* (Turing 1950), Braithwaite had chaired a BBC radio debate, in which Turing participated, on the possibility of machine intelligence. (The transcript is in Copeland 1999: 445-476.) Some years before that, they'd been fellow-Fellows at King's College. Indeed, Braithwaite was one of the only two people to have requested an offprint of Turing's "Computable

Numbers" paper written in 1936 (Hodges 1983: 123f.). And, so he told me later, it was he who'd pointed out to Turing its relationship to Godel's work (letter from R.B.B. to M.A.B., 21 Oct. 1982).

By the time of my becoming a once-to-thrice-weekly visitor to CLRU in 1957, Turing's vision was rarely discussed there in general terms. When it was, the emphasis was more on his technological predictions than on his philosophical views. Believing those predictions to be well-grounded, the denizens of CLRU focussed rather on the exciting challenges involved in bringing them to fruition.

In other words, the interdisciplinary community amidst the apple trees was making early attempts in the mechanization of thought. In particular, they were trying to identify, and formalize, some of the structural principles informing learning and language.

Pask, for example, was doing pioneering work on adaptive machines, using a wide variety of devices he'd built himself (Pask 1961). Some of his ideas may be viewed as early attempts in AI and A-Life, but he saw them as research in cybernetics (see Chapter 4.v.e). He was largely inspired by Ashby's self-equilibrating "Homeostat" of the 1940s (Ashby 1948). And he received strong encouragement from McCulloch, one of the founders of the cybernetics movement in the USA--whom I was to meet six years later (Pask 1961: 8).

Four years earlier, in 1953, Pask (with Mackinnon Wood) had constructed "Musicolour", an array of lights that adapted to a musician's performance. It had toured various theatres, ending up in a Mecca dance hall. Being a devotee of Mecca dance halls at the time, I much regretted never having encountered it. (I didn't know that it had acquired a reputation for bursting into flames--Mallen 2005: 86.) And in 1958 he started building self-organizing chemical systems that "learned", "evolved", and grew their own "sensors" (sound detectors)--Pask 1961: 105-8 (see 4.v.e and 15.vi.d).

But his main interest at that time was in adaptive teaching machines (Pask 1961, chap. 6). Rejecting the easy notion that one size fits all, he was trying to make his machines respond to individual differences between people's thought patterns, or cognitive styles (4.v.e). He'd been designing adaptive teaching machines since 1952, and his SAKI (Self-Adaptive Keyboard Instructor) of 1956, which taught people how to do key punching efficiently, was the first such system to go into commercial production (Pask 1958; 1961: 96ff.).

Unfortunately, I saw only very little of Pask in his back room at CLRU. A few years after leaving Cambridge, however, I would visit his makeshift office-laboratory in Richmond, where he was exploring yet more ambitious automatic teaching aids (Pask 1975a).

Bastin, too, was interested in cybernetics. Much of his spare time went into building a self-equilibrating machine (Bastin 1960). This was inspired by Grey Walter's electromechanical "tortoises" (Walter 1950a,b), which I'd seen exhibited at the Festival of Britain a few years earlier, in 1951 (4.viii.a). But it also involved ideas about hierarchy, which he was applying to quantum physics as well as to life (Bastin 1969).

The main efforts at CLRU, however, were in the study of language (9.x.a and d). Masterman's group was doing research on what's now called Natural Language Processing, or NLP (Wilks in press). They ranged widely over topics later claimed for AI and cognitive science. These included machine translation, the representation of knowledge for information retrieval, and the nature and process of classification. Although their theory of classification was never described in print as computational "learning", it dealt with issues later so described by AI (10.iii.d and 13.iii.f).

Masterman was one of the first people in the world to attempt machine translation, and she made semantics, not syntax, the driving force. She was deeply influenced by certain aspects of Ludwig Wittgenstein's later philosophy of language. Despite her gender--Wittgenstein was notorious for his mysogyny--she'd been one of his favourite students, to whom he'd dictated the lectures later known as *The Blue Book* (Monk 1990: 336). Indeed, she described herself to me on our first meeting as "the only person in England who really understands Wittgenstein". (Modesty wasn't one of her virtues.)

Accordingly, she handled translation by way of a computational thesaurus (Masterman 1957, 1962). More subtle than word-for-word dictionary lookup, her approach enabled word ambiguities to be resolved by inspecting the penumbra of concepts associated with neighbouring words in the text. Or rather, it made this possible in principle. In practice, the method was far from infallible: she delighted in telling people that Virgil's sentence *agricola incurvo terram dimovit aratro* had come out as *ploughman crooked ground plough plough.* This couldn't have happened without the thesaurus, because only *aratro* has a root likely to be listed against *plough* in a dictionary.

The work was practical as well as theoretical, asking how concepts and their semantic interrelations could be implemented in computers. --"Could be", rather than "were": computing facilities in 1957 were primitive (see 3.v.b). Using CLRU's data, Needham did some classification experiments on the EDSAC-2 in the Computer Laboratory. But this machine (in use until 1964) was far too small to handle a comprehensive thesaurus like Roget's. Moreover, no machine-readable thesaurus existed.

Some genuinely computational, though very primitive, work was done at the CLRU in the 1950s, using a Hollerith punched-card sorter. It wasn't until

1964, five years after I left, that the Unit received its first electronic computer: an ICL 1202, with 200 registers on a drum (Sparcke-Jones p.c.).

Because of these practical difficulties, the language team often had to do pseudo-computational tests. That is, they often worked 'mechanically' with paper lists, in the way required for the procedures using punched-card apparatus then being devised at the Unit (Masterman et al. 1957/86: 2). (Perhaps the Buddhist gods were witnessing the first instantiation of Searle's Chinese Room?--16.v.c.)

Masterman was a stimulating, if often infuriating, presence. Her conversation and teaching were peppered with provocative, sometimes deeply insightful, remarks. She encouraged my interest in the philosophy of mind. At her urging, I sent an early essay on 'free will' (i.e. the nature of intentions) to Gilbert Ryle, who published it in *Mind* eighteen months later, in April 1959 (Boden 1959). And her computational thesaurus was highly intriguing: how *could* one get the farmer to plough his ground in English, as well as in Latin?

However, it seemed to me, as an occasional looker-on, to be a technological project, not a psychological one. It clearly rested on intuitions about how people understand language. But I never heard it described as an exercise in the psychology of language--still less, as part of a general project aiming to understand all mental processes in computational terms.

Nor did I have the wit to recognize that possibility for myself--although if I'd interacted more often with Pask, I probably would have done. Masterman's research emphasized (semantic) *structure* rather than *process,* and didn't immediately suggest a way of conceptualizing mental processes as such.

Although I felt that it must somehow be connected with the puzzle of how thought of any kind is possible in a basically material universe, I couldn't see how to generalize it to the mind as a whole. I found her work interesting. But--or so I thought at the time--it wasn't relevant to the issues that most concerned me, and which had fascinated me as a schoolgirl even before becoming a medical student.

These were the nature and evolution of mind, the mind-body problem in general, and free will and psychopathology in particular. I was intrigued by paranoia, multiple personality, automatisms, and hypnosis. And I was especially puzzled by psychosomatic phenomena, such as hysterical paralyses and anaesthesias.

In these cases, there's no bodily damage: under hypnosis, the 'paralyzed' arm moves normally, and the 'anaesthetized' skin is sensitive. Still more

puzzling, the bodily limits of the clinical syndrome are inconsistent with the gross neuromuscular anatomy, and seem to be determined instead by what the layman-patient *thinks of* as an 'arm' or a 'leg.' For example, the movements that the 'paralysed' patient is unable to make don't correspond to any specifiable set of spinal nerves. They can be described only by using the non-anatomist's concept of an arm, thought of as bounded by the line of a sleeveless shirt. In other words, the mind appears not only to be influencing the body, as in normal voluntary action, but even overcoming it. How could this be?

Machine translation didn't help me to answer such questions. The most promising avenues, I thought, lay elsewhere: in the philosophy of mind and psychology, and in psychiatric medicine.

My intention at that time was to become a psychiatrist. The foray into philosophy was merely temporary. But in May 1959, when I was revising for my Moral Sciences finals and looking forward to going on afterwards to St. Thomas's Hospital, I was unexpectedly invited (at Braithwaite's suggestion) to apply for a philosophy lectureship at the University of Birmingham. This was "unexpected" in more senses than one. I'd never considered such a possibility for a moment. Nor was there much time to think about it, for the interviews were to take place only three days later.

Since I wasn't sure that I wanted the job, and didn't think I'd get it anyway, I was totally relaxed on the day. To my amazement, they offered it to me at interview. (It turned out that the little piece in *Mind* had helped.) But I asked for forty-eight hours to think it over: *medicine or philosophy?* was a difficult decision. Masterman's very strong support (she showed me her written reference, when she found that I was dithering) was one of many factors that influenced me to accept the offer.

(I also sought advice from my former pathology supervisor, today a distinguished Emeritus Professor of Pathology. He observed that having a wife with a medical degree, like his--whom I could see hanging out the washing in the garden with pegs between her teeth, while he smoked his pipe in his armchair--would always help a family to get a second mortgage if needed. This remark, in that all-too-familiar domestic context, was less persuasive than he'd intended.)

It was a strong department (Peter Geach was one of the luminaries), and I was very happy there. However, I soon got bored. For Birmingham's Chinese walls between disciplines impeded my interests in the philosophy of psychology and biology. I considered returning to medicine (St. Thomas's said "Yes, come!"), but having forfeited my state studentship to earn my own living I could no longer afford to do so.

Instead, I followed the suggestion of my old Cambridge friend Charlie

Gross (who a few years later would discover the 'monkey's hand' neurones in the monkey's brain: 14.iv.b). He said "There's this man Bruner at Harvard, who's been doing some work I think you'd find interesting--and there are scholarships you can apply for to go to the States". So I applied for a Harkness Fellowship, which enabled me, in the autumn of 1962, to cross the pond to study cognitive and social psychology with Bruner. (When I first met him, he was chatting with George Miller. "Here's our double-first from Oxford", he said to him. "Cambridge!" I protested--and "Welcome to Yale!" came quick as a flash from Miller.)

By the time I left for the USA, I'd already decided to go to the just-initiated University of Sussex when I got back to England. This was because, most unusually, it was conceived from the start as an interdisciplinary institution. I was already committed to interdisciplinarity, of course. But, sailing happily through the storms on the magnificent Queen Mary (the roughest voyage for twelve years), I never imagined that my colleagues and I would eventually found Sussex's Cognitive Studies Programme (later the School and now the Centre for Cognitive Science), which in 1973 world-pioneered degrees integrating AI, philosophy, psychology, and linguistics. The idea couldn't even have occurred to me.

Barely a week after docking at Manhattan, however, it might have done. The conceptual leap from computation to psychology, and to the mind-body problem, happened (for me) a mere two days after arriving in the other Cambridge.

The occasion was my first sight of the remarkable book *Plans and the Structure of Behavior*, by Miller, Eugene Galanter, and Karl Pribram (1960). I picked it up while browsing in a second-hand bookshop on Massachusetts Avenue. Why I did so, I'll never know. It was a hideous object: bound in a roughly textured cloth, a dull rust in colour (my least-favourite hue), horribly coffee-stained, and defaced by heavy underlining on almost every page. But it changed my life.

Nor was I the only one, for it was highly influential (see Chapter 6.iv.c). I soon discovered that it was on Bruner's reading lists at Harvard's new Center for Cognitive Studies, founded only a few months before--and not just because Miller was the co-founder! It was recommended also for Phil Stone's seminar on "Computer Simulation", for which I was to do my first programming. (We wrote our programs in Victor Yngve's early list processing language COMIT, using punched cards for MIT's pre-release prototype of the IBM-360--not officially announced until 1964.)

But all that was still to come. Already primed by Masterman and Pask, my thinking was instantly triggered by this coffee-stained volume. Leafing through it in the bookshop, it seemed to offer a way to tackle just those questions which had bothered me as a schoolgirl.

It was an intoxicating attempt to apply specific computational ideas--hierarchies of Test-Operate-Test-Exit procedures (TOTE-units)--to the whole of psychology. Unlike Masterman, it focussed on process as well as structure. And it ranged from animal learning and instinct, through memory and language, to personality, psychopathology, and hypnosis. Self-confessedly vague and simplistic, and often careless to boot, it was nevertheless a work of vision.

Its computational ideas soon informed my own work. In 1963, I wrote a paper applying them to William McDougall's rich theory of the purposive structures underlying normal and abnormal personality (Boden 1965). And a few years later, I addressed one of my longstanding puzzles by outlining how a robot could have a paralysis conforming not to its actual wires-and-levers anatomy, but to its programmed "concept" of what an arm is (Boden 1970). Its behaviour, I argued, would therefore be describable in intentional terms. That is, what it was "doing," and how it might be "cured," could be stated only by reference to the descriptions and instructions in its program.

In the interim, I'd returned to England (and moved to Sussex in 1965), and was writing my first book: *Purposive Explanation in Psychology* (1972). Begun as my PhD thesis (the first purely theoretical thesis that the Harvard department had ever allowed), this took me eight years to finish. The delay was explained only partly by the amount of intellectual work involved: the publisher's air-mailed advance copy reached me in hospital on the day after the birth of my second baby. (Both were deep purple on arrival.)

In that book, I developed a fundamentally physicalist but non-reductionist account of purpose, and other intentional concepts. That is, I offered an essentially functionalist philosophy of mind--though using my own terminology, not Hilary Putnam's (I came across his work later). I compared my account of mind, and of the mind-body relation, with a wide range of theories in psychology and philosophy. And I focussed most closely on McDougall--not as an unquestioned guru, but as an intellectual sparring partner.

What had drawn me to McDougall was his deep insight into the complex structure of the human mind, and his many explicit arguments against psychological reductionism (2.x.b. and 5.ii.a). Most of these, but not all, I thought to be valid.

It turned out later that there were several personal links, as well. On researching McDougall's life history (1871-1938), I was intrigued to discover that, after completing his degree in medical sciences at Cambridge, he'd taken up the same clinical scholarship at St. Thomas's Hospital which had been offered to me in 1959, and again in 1962. He, like me, had moved from medicine to psychology, with philosophy of mind constantly in the background. And he,

too, had gone from one Cambridge to the other: he was a professor at Harvard for several years.

There was even a link that made me one of his intellectual grandchildren, by means of a sort of apostolic succession. For when I said to Bruner, in the spring of 1964, that instead of doing experiments on information density in disjunctive concepts (a mind-numbing topic which he, intending to be helpful, had suggested to me) I wanted to study McDougall's theory of purpose, he told me--after a gasp of amazement--that McDougall had been his teacher at Duke University.

Whereas I'd been recommended to go to Harvard Graduate School by Charlie Gross, Bruner--so he told me--had been specifically warned against it by McDougall. McDougall's broadly-ranging psychology had been highly influential until it was suddenly eclipsed by behaviourism. He remained bitter about this for the rest of his life, and left the newly-behaviourist Harvard in disgust to set up his own outfit at Duke. When the young Bruner announced his intention of travelling north to Massachusetts, McDougall had gruffly warned him about the intellectual corruption (he never minced his words) that awaited him there.

By the time I arrived at Harvard, a quarter-century later, the behaviourist "corruption" was less strong. Or anyway, that was true in Bruner and Miller's Center for Cognitive Studies, if not in the Psychology Department as such--whose denizens included Burrhus Skinner and Richard Herrnstein. Even so, McDougall's name had vanished from the curriculum. (Hence Bruner's gasp of amazement.) It survived only as one of many items on a three-page mimeographed list of long-dead worthies, circulated as potential essay topics for Gordon Allport's seminar on the history of social psychology.

What alerted me to him initially was the title of his book *Body and Mind* (1911), which Allport had included on the list alongside the author's name. On consulting his work (long-unborrowed from the Harvard library), I found that McDougall's ideas had a welcome subtlety and depth, and a refreshing concern for real life, whether psychiatric syndromes or everyday pursuits. His psychology dealt not only with cognition, but with motivation and emotion too--and, significantly, with how these three types of mental function are closely integrated in individual personalities.

Besides those strong points, his writings abounded with philosophical as well as empirical questions. These attracted me, although they were anathema to most (Anglo-American) textbook writers of the 1950s and 1960s, who believed--wrongly--that psychology had finally 'escaped' from philosophy. In truth, they'd simply accepted the current philosophical fashion (operationalism, or logical positivism), without stopping to question it carefully.

That's not to say that I agreed with all his philosophical arguments, or that I accepted his robust defence of "animism". Far from it. McDougall had been combatively anti-mechanistic, even claiming that purposive behaviour requires a special form of energy (*horme*), intrinsically directed to instinctive goals. That, for me, was a step too far: purposive explanation is one thing, purposive energy quite another. But it was important to understand why, when considering intentional phenomena, he'd felt it necessary to say this.

As part of my critique (Boden 1972), I suggested--what probably made him turn in his grave--that his many insights about personality and psychopathology could be simulated in computational terms. If this was indeed possible, those theoretical insights could be saved, and even clarified, without positing any mysterious energy. And if it could be done for McDougall's avowedly anti-mechanistic psychology, it could in principle be done for any other. (Sigmund Freud's purposive theory would have been less suitable as an exemplar, for he believed that psychology does have a mechanistic base.)

In sketching specific ways of doing this, I had to extrapolate from the computer models that already existed. In 1963, when the book was begun, there were only a handful of candidates.

By the time it was finished, early in 1971, there were many more. These ranged from work in computer problem solving, through programs for vision and language, to models of analogy, learning, and various aspects of personality. For instance, preliminary reports on Terry Winograd's research, whose official publication in 1972 suddenly raised the visibility of computer modelling in the wider intellectual community (see 9.xi.b and 10.iv.a), had been made available informally in 1970.

These new examples, whether successful in their aims or not, were clearly relevant to my book's central claim: that purposive behaviour is intelligible in computational terms, and could in principle be simulated in computers. But although I was able to refer briefly to a few of them, they were in general too numerous--and many were too late--to be added to my already lengthy manuscript.

So I decided that, as soon as my first book was finished, I would write an extended footnote to it. This would detail what could--and, just as importantly, what could not--be done by computer modelling in the early 1970s, and what would be needed for the many remaining obstacles to be overcome.

The resulting footnote, *Artificial Intelligence and Natural Man* (Boden 1977), ran to 537 pages. (By this time, the term "artificial intelligence" had largely replaced "computer simulation".) It devoted a chapter each to the philosophical,

psychological and social implications of AI. Indeed, in one sense they were what the book was really about. Most chapters, however, described the AI as such.

For the sake of readers knowing nothing about AI and highly suspicious of computers to boot, it contained not one line of code. It was also highly critical, in the sense that it identified countless mismatches between existing AI programs and real minds. Nevertheless, it was assigned--alongside Patrick Winston's very different *Artificial Intelligence* (1977)--as a compulsory text for AI courses at MIT and Yale. I was told it was the first time they'd assigned two books, rather than just one. It was also used as the basis of various psychology courses in the UK and USA, including the Open University's first Cognitive Psychology course. (Later, in 1993, I was delighted to be elected an early Fellow of the American Association for Artificial Intelligence, in part for having written it.)

That book *was* fully comprehensive. Had Squoggins been working at the time, he would very likely have been included. For it mentioned virtually every AI program of any interest, including many available only as privately circulated reports or working papers. And it gave closely detailed explanations and critiques of many of them. It ranged over diverse aspects of mind: from language and vision to neurosis and creativity--on which last I promised myself a whole volume, later (Boden 1990a, 2004). And it identified theoretical challenges many of which *still* remain to be met. (So the second edition, in 1987, was unchanged except for an added 'up-date' chapter.) In short, it provided a near-exhaustive description of the state of the art of AI at the time.

Such a project is no longer possible: even 1080 A4-sized pages aren't enough for a fully comprehensive account (Russell and Norvig 2003). If it's not possible for AI, still less is it feasible for cognitive science as a whole. Too much water has flowed under the bridge. There are too many unsung Squogginses out there, and too many branching implications that could be explored. That's even more true if one takes a *historical* approach, for then the potential subjects multiply yet more relentlessly.

I hope these far-from-comprehensive pages will tell an illuminating story, nonetheless. As Naude realized (fueling his attack on historians in general), the facts one chooses to relate, and how one decides to link them, will depend on one's background and perspective. In this book, I've aimed to show how cognitive science has developed so as to help solve problems about mind, brain, and personality that have intrigued me ever since I was a girl.

M.A.B.--Brighton, January 2006.