AARON SLOMAN: A BRIGHT TILE IN AI'S MOSAIC

Margaret A. Boden
University of Sussex

I: The Aims of AI

When AI was still a glimmer in Alan Turing's eye, and when (soon afterwards) it was the new kid on the block at MIT and elsewhere, it wasn't regarded primarily as a source of technical gizmos for public use or commercial exploitation (Boden 2006: 10.i-ii). To the contrary, it was aimed at illuminating the powers of the human mind.

That's very clear from Turing's mid-century paper in *Mind* (1950), which was in effect a manifesto for a future AI. Like Allen Newell and Herbert Simon too, whose ground-breaking General Problem Solver was introduced as a simulation of "human thought" (Newell and Simon 1961). Turing's strikingly prescient plans for a wide-ranging future AI were driven by his deep curiosity about psychology. Indeed, most of AI's technical problems and answers arose in trying to discover *just how* computers could be programmed to model human thought.

Virtually all of the philosophers who read Turing's *Mind* paper ignored that aspect of it (Boden 2006: 16.ii.a-b). They weren't interested in the technical or psychological questions. Instead, they focussed on criticizing the so-called Turing Test--which had been included by the author not as serious philosophical argument but as jokey "propaganda" (Gandy 1996: 125). In other words, they responded not by getting intrigued, or even critically sceptical, about a potential AI, but by considering one ancient philosophical question: whether a machine could think.

While ignoring what Turing saw as the most important part of the paper, however, they were addressing the second main aim of early AI. For the AI pioneers weren't targeting only psychology: they had their guns trained on philosophy, too. In other words, besides wondering how the mind-brain actually works (and how it could be modelled in a computer), they wondered how it is possible for a material system to have psychological properties at all.

Turing himself was interested in the philosophical problem of the nature of mind, even though he rejected the philosophers' usual way of addressing it. As for his AI successors, they hoped to illuminate a variety of longstanding problems--not only in the philosophy of mind, but in philosophical logic and epistemology as well. Marvin Minsky's (1965) paper on self-models and freewill, and the foray into logic and epistemology by John McCarthy and Patrick Hayes (1969), were early examples. So was Allen Newell and Herbert Simon's approach to the nature of reference, choice, and freedom (Newell 1973). (My own 1960s work, likewise, used AI to address problematic issues in philosophy, as also in psychology: Boden 1965, 1969, 1970,

1972. Before then, I had approached these issues in a more traditional manner: Boden 1959.)

Today, AI looks very different. Most current workers avoid psychological modelling, and aren't drawn to philosophy either. Trained as computer scientists at university, they have scant interest in these matters. Originally, everyone came into AI from some other discipline, carrying a wide range of interests with them. But that's no longer so. Indeed, AI has adopted a third aim: most AI research is now directed at the useful gizmos.

(There are some exceptions, of course. Besides individual researchers--including both long-time AI modellers and younger workers--there have been some national research programmes focussed on science rather than gizmos. For example, since 2003 the European Union has provided significant funding for interdisciplinary projects on cognition, involving psychologists, philosophers, and neuroscientists as well as AI programmers. These include the flagship Human Brain Project, and funding devoted to ICT and creativity.)

As for the gizmos, there are lots of those. In fact, this research has been so extraordinarily successful that its results are now largely taken for granted. In other words, AI has become well-nigh invisible to the general public, who benefit from the gizmos every day in countless ways but don't realize that they utilize AI. Even computer professionals, who should know better, sometimes deny the involvement of AI: an AI friend of mine was shocked to be told by a colleague that "Natural language processing (NLP) isn't AI: it's computer science". Partly because of this invisibility, AI is often even said to have "failed" (Boden 2006: 13.vii.b).

There's another reason, however, why AI is commonly said to have failed: its early hopes--and hype--about modelling, matching, and maybe even surpassing human mental powers have not been met. Most of the AI gizmos (which started with work on expert systems: Boden 2006: 10.iv.c, 13.ii.b) address only a very narrowly defined task, although admittedly they frequently achieve far better results than people can. Sometimes, they can surpass the world champion—as IBM's Deep Blue did when playing chess against Gary Kasparov in 1997. But even the more widely-aimed AI systems don't match up to human intelligence.

Despite amazing advances in NLP, for instance (using powerful statistical methods very different from traditional AI), the ability of computers to deal with natural language is far less sensitive than that of reasonably articulate human beings. Common-sense reasoning, too, despite the power of today's data-mining and the notorious mastery of IBM's Watson in playing *Jeopardy* (Baker 2011), has not been reliably emulated. As for the non-cognitive aspects of mind, namely motivation and emotion, these are commonly ignored by AI researchers--or addressed in a very shallow fashion. In short, all these aspects of human mentality are much more challenging than was initially expected.

Even vision, a sense we share with many other animals, has turned out to be more difficult than expected. In 1966, Minsky asked a bright first-year undergraduate (Gerald Sussman) to spend the summer linking a camera to a computer and getting the computer to describe what it saw. (They wanted a vision system for the MIT robot.) This wasn't a joke: both Minsky and Sussman expected the project to succeed (Crevier 1993: 88). Even in the world of shadow-less convex polyhedra, that was a tall order: it couldn't be mastered in a single summer. In more realistic worlds, beset not only by shadows but also by curves, multiple occlusions, and missing and/or spurious parts, visual computation is hugely more complex than had been thought. (It's still the case that computer systems presented with a complex visual scene are very limited in the objects and relationships that they can identify; moreover, they mostly ignore the biological functions of vision—see below.)

II: AI as philosophy

What does all this have to do with Aaron Sloman? Well, he has never been a gizmo-chaser. With a background in mathematics and philosophy, he has always taken both the aims of early AI seriously. That was already evident in his book *The Computer Revolution in Philosophy* (1978), and remained so in his later papers. (His book is now available online, and is constantly updated--so can act as an extra reference for most of the topics mentioned here.)

But for him, the usual intellectual priorities of AI, where psychological and technical questions took precedence over philosophical ones. were equalized--or even reversed. In other words, what I've called AI's "second" aim has been Sloman's first. Despite being a highly accomplished AI programmer, and a key figure in the development of AI in the UK, Sloman still regards himself as first and foremost a philosopher. ("Still", because he was a professional philosopher for some years before encountering AI.)

The familiar philosophical puzzles he has written about include freewill, reference, intentionality, representation, modal reasoning, philosophy of mathematics, causation, and consciousness (Sloman 1974, 1986, 1987, 1996b,c; Sloman and Chrisley 2003). His book has two illuminating chapters on the philosophy of science. Above all, however, he has been concerned with the philosophy of mind—not only the age-old problem of mind and machine (e.g. Sloman 1992, 1993), but also deep questions about the nature of mind that are *not* familiar philosophical chestnuts, as those just listed are.

One of the less familiar puzzles he has addressed is the nature of non-logical or “intuitive” thinking, such as mathematical reasoning based on diagrams. He was drawn to this topic partly because early AI wasn’t able to model it. But his motives weren’t purely technological. Although the paper he published on it appeared in the *Artificial Intelligence* journal, and was later included in a high-profile AI collection on knowledge representation (see below), Sloman chose to describe it in the title as

exploring "Interactions Between Philosophy [*sic*] and AI" (1971: 209). Much of the paper contrasted non-logical ("analogical") reasoning with the type of inference described by the philosopher Gottlob Frege. But the main inspiration for the paper was the philosopher Immanuel Kant. Sloman wanted to defend Kant's (unfashionable) claim that intuitive mathematical reasoning, neither empirical nor analytic, could lead to necessary truths—a defence that he'd mounted at greater length in his DPhil of 1962.

Sloman has never been guilty of irresponsible hype about AI, which he has criticized consistently over the years. For instance, in his 1978 book (section 9.12) he predicted that , by the end of the century, computer vision would not be adequate for the design of domestic robots capable of washing dishes, cleaning up spilt milk, etc. Nevertheless, he has always had high ambitions for the field. Avoiding the narrow alleyways of gizmo-AI, his prime concern has been the mind as a whole.

Or perhaps one should rather say *minds* as a whole, since he has considered intelligence in general--in animals, humans, and machines. Besides remarking on the mental architecture of particular species (e.g. humans, chimps, crows, ants...), he has tried to outline the space of all possible minds (Sloman 1978: chap. 6). His work makes it clear that intelligence isn't an all-or-none phenomenon, nor even a continuously varying property. Rather, it's a richly structured space defined by many distinct dimensions, or information-processing procedures, which enable radically different--and to some extent incommensurable--types of intelligence to arise.

To some extent, Sloman's view of mind leant on the philosophy of Gilbert Ryle (1949), whom he had encountered as a DPhil student at Oxford (Sloman 1996c: Acknowledgments; 1978: chap. 4). Besides always sharing Ryle's scorn for "the ghost in the machine", Sloman was—eventually (see below)--inspired by his talk of "dispositions" versus "episodes".

Ryle analysed many psychological terms not as reports of actual events or phenomena, but as denoting long-standing dispositions to behave in a certain way in certain circumstances. He compared jealousy, for example, with brittleness. To say that glass is brittle is to say that *if* it is hit *then* it will probably break; likewise, to say that someone is jealous is to say (for example) that *if* someone sees her husband talking animatedly to another woman *then* she will very likely say something unpleasant to one or both of them. The same is true, he argued, of concepts denoting propositional attitudes--such as *know, believe, desire, prefer, fear*, *expect*, and *hope*. So to believe that *p* is to be disposed to say that *p*, and to behave in ways that would be appropriate (given the person's other beliefs and desires) if *p* were true.

This approach implied that most psychological concepts are logically interlinked with others. In other words, the activation of disposition *a* is likely (by definition) to lead to the activation of dispositions *b, c*, … and to the triggering of episodes *x, y*, and *z.* To be sure, dispositions can be suppressed—much as a piece of glass may never

be dropped, or may be wrapped in a protective cloth to prevent its breaking. But jealousy without *any* tendency to resent, denigrate, or harm the person or persons concerned simply is not jealousy.

Even first-person psychological statements such as *I see blue* or *I have an itch*, said Ryle, are not reports of events in some mysterious non-material world but "avowals" of certain behavioural dispositions. So someone who claimed to see blue, or to have an itch, who did not assert any resemblance with the sky, or make any attempt to scratch, would either be thought insincere or would simply not be understood. As for the feelings sometimes involved in emotions, to say *I feel depressed*, according to Ryle, is not to report an internally accessible conscious state, but rather to perform "a piece of conversational moping: … not discovery [by Cartesian direct access], but voluntary non-concealment" (Ryle 1949: 102).

Ryle was widely accused of behaviourism—and, in my view, rightly so. However, his key term *disposition* was systematically ambiguous, denoting either observable behaviour and/or its underlying causes. Most analytic philosophers in the 1950s, like Ryle himself, interpreted it as a summary *description* of behaviour, not an *explanation* of it. This was largely because they saw explanation as the task of science, not philosophy (see Section v). Later, the term was read by some (not all) philosophers as explanatory, denoting *the mechanism responsible for the relevant behaviour* (Squires 1970).

Initially, Sloman's reading of Ryle's key term was descriptive rather than explanatory. As a result, he rejected Ryle as a behaviourist. Moreover, if he'd been asked to interpret "disposition" as explanation, he would have assumed it to refer to some (unknown) neural mechanism. But his conversion to AI enabled him to see that it could also be a *computational* explanation.

On re-reading *The Concept of Mind* (Ryle 1949), he was especially interested by the fact that--as remarked above--Ryle's dispositions and episodes were interlinked. In other words, one mental state could switch to another mental state, much as one part of a computer program could activate another. His work thereafter can be seen as an attempt to put specific computational flesh onto broadly Rylean dispositional bones.

III: A vision of vision

Sloman's avoidance of AI hype is grounded in his nuanced appreciation of the significant complexity and diversity of human (and much animal) intelligence. That was apparent even in his earliest work on computer vision, the POPEYE project (Sloman 1978: chap. 9).

POPEYE modelled the interpretation not of fully connected drawings of perfect polyhedra (or even polyhedra-with-shadows), but of highly ambiguous, noisy, input-

-with both missing and spurious parts: see Figure 1. And it simulated the complexity of perception to an extent that was highly unusual at the time.

Figure 1 about here

Sloman's thinking about vision (although not the POPEYE program itself), especially in the years following the implementation of POPEYE, stressed the fact that vision is integrated with action and motivation (Sloman 1983). This was a lesson that he had learnt from the psychologist James Gibson (1966). Gibson's theory of perceptual "affordances" held that vision has evolved for a range of different purposes, for which different types of motor action are appropriate.

That is, the primary point of vision is not to build a visual image, nor even—as David Marr would argue later (see Section v)--to represent the location of objects in 3D-space. Rather, it is to prepare for and guide motor behaviour, enabling the organism to achieve evolutionarily significant goals. As well as answering questions about what things are in the environment and just where they are located, such goals include recognizing and following a pathway, avoiding an obstacle, jumping onto a stable support, approaching a potential mate, and deciding which way to move in order to see more of something already glimpsed.

If those are the purposes of vision, a realistic (or even near-realistic) computer model would need to combine many different sorts of background knowledge. Moreover, information processing could occur concurrently in different domains, determining which sub-processes would dominate the scarce computational resources. (This differed from the heterarchy so popular in the early-mid 1970s, wherein there was only one locus of control at any moment, and control was transferred to process *X* by an explicit call from process *Y*: Boden 2006: 778ff.). Each knowledge domain in POPEYE had its own priorities for finding/processing information, and these priorities could change suddenly as a result of new information arriving unexpectedly. Diversity (and flexibility) was increased also by the fact that some of the internal representations constructed by POPEYE were temporary, rather than provisional. (Something provisional may become permanent, but something temporary should not.)

Considered as a practical visual system for a robot, POPEYE wasn't impressive. Quite apart from anything else, it didn't actually contain anything that linked to bodily action (though some aspects of it could have been so linked, if Sloman had had the opportunity to develop it further: see Section v). But Sloman wasn't trying to advance robotics, least of all robotics confined to toy polyhedral worlds. Rather, he was trying—again, a *philosophical* aim--to advance Kant's argument that the mind must provide some prior knowledge for even the "simplest" perceptions to be

possible (Sloman 1978: 230). But whereas for Kant the principles of organization were very general, and innate, for Sloman they also included highly specific learnt examples.

Accordingly, his program modelled the fact that high-level visual schemata can aid recognition enormously. For instance, learnt knowledge of the familiar upper-case sign "EXIT" helps us--and POPEYE—to recognize the four letters in Figure 1. This computational diversity has a chicken-and-egg aspect: if one recognizes a particular set of dots in Figure 1 as co-linear, that can help one to recognize an "E"; but if one has already recognized "EXIT", one will be much more likely to recognize *those very dots* as co-linear (Sloman 1978: 228-232).

The moral of POPEYE, as of Sloman's more recent work on vision (1983, 1989), was that any realistic degree of visual complexity will involve many diverse types of background knowledge, all playing their parts concurrently. As he put it: "Our program uses knowledge about many different kinds of objects and relationships, and runs several different sorts of processes in parallel, so that 'high-level' processes and (relatively) 'low-level' processes can help one another resolve ambiguities and reduce the amount of searching for consistent interpretations. It is also possible to suspend processes which are no longer useful: for example low-level analysis processes, looking for evidence of lines, may be terminated prematurely if some higher-level process has decided that enough has been learnt about the image to generate a useful interpretation. This corresponds to the fact that we may recognize a whole (e.g. a word) without taking in all of its parts" (1978: 229).

In an important sense, however, POPEYE wasn't really—or anyway, it wasn't *only*—about vision. Rather, it was a preliminary exercise in architecture building. For Sloman saw computer vision as a way of keying in to the computational structure of the mind as a whole.

IV: Architectural issues

In his early book, Sloman had discussed the nature of the mind as a whole (1978: ch. 6). He argued, for example, that because emotion is integral to intelligence, truly intelligent robots would have to have emotions too. For instance "they will sometimes have to feel the need for great urgency when things are going wrong and something has to be done about it" (1978: 272; cf. Sloman and Croucher 1981; Sloman 1982).

Over the following years, he focussed increasingly on the control structure of the entire mind, eventually offering computational analyses of motivation and emotion that illuminated even such seemingly computationally recalcitrant phenomena as anxiety and grief. (The most accessible statement of his mature approach is Sloman 2000; for more technical descriptions, see Sloman 1998, 2001, 2003, and Sloman n.d.)

Emotions often involve conscious feelings, but—Sloman argued--these are not all there is to emotion. Emotions are control-structures, participating in the guidance of action and the scheduling of potentially conflicting motives. They enable goals and sub-goals to be chosen appropriately, and—if necessary—to be put on hold, or even dropped, as circumstances change. They interact with perception, and with various types of short-term and long-term memory, alarm systems, and (variable) attention-thresholds.

According to Sloman (and his student Luc Beaudoin), emotions—and whole minds, too--differ from each other in terms of three main architectural levels. These involve what he calls reactive, deliberative, and meta-management mechanisms.

The minds of insects are mostly reactive, depending on learnt or innate reflexes. They are capable only of "proto-emotions": inflexible reactions that have much the same adaptive function as (for instance) fear in higher animals.

A chimpanzee's mind is largely deliberative, capable of representing and comparing past, and possible future, actions or events. So the animal is capable of backward-looking and forward-looking emotions: non-linguistic versions of anxiety and hope, for example.

In general, deliberative mechanisms are more complex than reactive ones. They are also diverse, including various intermediate levels between pure reaction and full-blown planning—which employs multi-step look-ahead with various strategies, and uses meta-management (the third level) to control the planning process (Sloman 2009a). (This diversity is underplayed by the currently fashionable embodiment/enactive movement: e.g. Brooks 1990; Clark 2013. Among other things, such work prioritizes instant control via evanescent "online" signals, at the expense of more long-lasting "offline" processes and data-structures: Sloman 2009b, 2013.)

In human adults, the deliberations can include conscious planning and reasoning, generating more precisely directed emotions accordingly. In general, language makes possible emotions with propositional content, which may be highly specific—and which may vary significantly from one culture to another. The concept of love, for example, differs across cultures: so emotions such as love and grief, and even honour, differ too. In addition, an adult human mind has a rich store of reflexive meta-management mechanisms, which monitor and guide behaviour. Emotions centred on the concept of self—such as vainglory and embarrassment—are now possible, accordingly.

Sometimes, humans seemingly have no choice: a danger that's just been identified *must* be averted, and it must be done *now*. There's no time for conscious deliberations. Reactive mechanisms must take control. But the sense in which a human being (sometimes) has no choice about what to do next is fundamentally

different from the sense in which an insect (always) has no choice. Humans are free, whereas insects aren't. But human freedom doesn't depend on randomness, or on mysterious spiritual influences: to the contrary, it's an aspect of *how our minds work.* Sloman's account of mental architecture shows how our emotions can sometimes compromise our freedom (by leading us to react [*sic*] unthinkingly) even though they also help to make it possible (by controlling appropriate cognitive mechanisms, such as deliberation).

The specific emotions discussed by Sloman include grief and sorrow, closely-related but different emotions that are generated by the death of a loved one. A dog may suffer from sorrow, and appear to mourn its lost master. But grief in a human mourner is a much more complex, and ("Edinburgh Bobby" notwithstanding) more long-lasting, than it is in a dog. Quite apart from being expressible in a host of linguistically distinguishable ways, it leads to continual (though gradually decreasing) interruptions of thinking and behaviour as the mourner remembers, or is reminded of, the lost person. The previously-built motivational structure of caring about and encouraging the goals and interests *of the loved person* has to be gradually dismantled (Fisher 1990). This is not, and cannot be, the work of a minute: mourning inevitably takes time.

Most of what was said in the previous paragraph could have been said by Ryle—or by a competent novelist. But in discussing grief, Sloman and his students (Beaudoin and Ian Wright) used their deep knowledge of AI to suggest a host of specific information-processing mechanisms that could interact to generate the various mental/behavioural phenomena concerned (Wright et al. 1996; Sloman 2000).

Critics will surely complain that grief, over and above its dispositional aspects, involves searing feelings, conscious episodes which—they say--cannot be captured in computational terms. Indeed, many say this about emotions in general. Feelings of grief, or joy, or anxiety … are special cases of what philosophers call *qualia*. Any adequate theory of emotion must therefore make place for qualia. But this is a tall order. For, notoriously, *all* philosophers (Rene Descartes and Ryle included) have difficulty in giving a coherent account of conscious feelings or sensations.

In other words, the AI-friendly thinkers aren't the only ones to encounter trouble here. But, undeniably, trouble there is. Some computationalists have denied the existence of qualia (Dennett 1988; 1991, ch. 12). Sloman did not. Instead, he analysed them as aspects of the virtual machine which is the mind (1999; Sloman and Chrisley 2003).

Specifically, he saw them as intermediate structures and processes generated by an information-processing system with a complex, reflexive, structure. Some qualia (but not all) can be noticed and thought about using self-reports—which might require the system to generate ways of classifying them, using *internal* categories that can't be matched/compared with comparable categories in other virtual

machines (other minds). But the self-reports are something extra. They are directly accessible to the highest level of the system itself, and are sometimes communicated verbally, or expressed behaviourally, to others.

(On this view, some of the very same qualia could exist in simpler organisms that have sophisticated perceptual mechanisms without also having human-like self-monitoring mechanisms for introspection. So a house-fly might have visual qualia of which it simply cannot be aware. Clearly, Sloman's analysis conflicts with any view which requires qualia, by definition, to involve self-knowledge, or to be actually attended to.)

As Ryle would doubtless have been glad to hear, self-reported qualia do not rest on Cartesian "direct access" to some mysterious mental world. The directness, or lack of evidence, of first-person experiential statements is due—so Sloman argues--to the particular kind of (reflexive) computation involved. For example, the meta-management system may have access to some intermediate perceptual data-base (*blue, itch* …) which does not represent anything in the third-person-observable outside world because it is the content of a dream or hallucination. In other cases, it would be part of the process of perceiving something external.

Sloman's pioneering discussions of the integration of cognition, motivation, and emotion were *computational* analyses, in the sense that they conceptualized the mind as an information-processing system and were deeply informed by an extensive knowledge of various types of AI research. And they have been extended (and are still being developed) by him and his colleagues in the same strongly computational spirit. But, for many years, they were not illustrated by functioning computer programs. Even now, Sloman cannot provide a computer model of his architectural theory as a whole.

Since the 1990s, however, he and his students have implemented a model based on his theory of emotional perturbances. This is the series of MINDER programs, developed to illuminate the nature of emotion, and its role in the control of action (Wright and Sloman 1997; see also Beaudoin 1994; Wright 1997). (To *illuminate*, not to capture: these programs model only a very limited subset of Sloman's theory of the mind.)

MINDER simulates the anxiety that arises within a nursemaid, left to look after several babies single-handed. She has only a few tasks: to feed them, to try to prevent them from falling into ditches, and to take them to a first-aid station if they do. And she has only a few motives to follow: feeding a baby; putting a baby behind a protective fence, if one already exists; moving a baby out of a ditch for first-aid; patrolling the ditch; building a fence; moving a baby to a safe distance away from the ditch; and, if no other motive is currently activated, wandering around the nursery. In short, she's hugely simpler than a real nursemaid. Nevertheless, she is prone to emotional perturbations ("proto-emotions") comparable to anxiety—or rather, to several interestingly different types of anxiety.

Sloman's simulated nursemaid has to notice, and respond appropriately to, a number of visual signals from her environment. Some of these trigger (or affect) goals that are more urgent than others: a baby crawling towards the ditch needs her attention sooner than a merely hungry baby does, and one who's about to topple over the edge of the ditch needs attention sooner still. But even those goals which can be put on hold for a while may have to be coped with eventually; and their degree of urgency may rise with time. So a near-starving baby, who has not been fed for hours, can be put back into its cot if another baby is about to fall in the ditch; but the baby who has waited longest to be fed should be nurtured before the ones whose last meal is more recent. As these examples suggest, the nursemaid's various tasks can be interrupted, and either abandoned or put on hold. She—or rather, the MINDER program—has to decide just what the current priorities are. Much as with a real nursemaid, her anxieties increase, and her performance degrades, with an increase in the number of babies—each of which is an unpredictable autonomous system.

Sloman has identified several important limitations of MINDER (Wright and Sloman 1997: sects. 3.7.2, 4.3), some of which could be alleviated or overcome in later versions. But it's important to note that this system could be used to model many different types of autonomous agent besides babies and nursemaids. Indeed, a programming environment based on the early work on MINDER (and first used in Wright 1997) has been placed on the Internet for other AI workers to experiment with. This is the SimAgent toolkit (Sloman and Poli 1995; Sloman 1995), which has been used by a number of researchers outside Sloman's group.

We're now in the 21st Century, with POPEYE and MINDER approaching the status of ancestral forms. Their descendants in Sloman's recent thinking include his current work on autism (soon to be included on his website).

This work-in-progress attempts to throw light on the intriguing drawing-abilities of the autistic child Nadia, and to explain why they regressed when her language skills developed. It also suggests that autism considered as a deficiency in Theory of Mind (Frith 1989/2003; Boden 2006: 7.vi.f-g) is a special case of a more general range of possible developmental abnormalities that can impede later developments. Sloman situates these ideas within the theoretical framework that he (with the ethologist Jackie Chappell) has produced for accounting for "the differences between precocial and altricial species, where the latter have multiple routes leading from genome to behaviours, through competences that develop late and build on competences that developed earlier under the influence of the environment"(p.c.).

Autism isn't the only intriguing topic that Sloman is currently working on. Meta-morphogenesis is another. This is an enquiry into how the possible "design spaces" and "niche spaces" in biological evolution generate, and are generated by, an increasing variety of information-processing mechanisms. How is morphogenesis driven by new forms of representation, ontologies, and architectures? How do

reflexive (meta-management) mechanisms evolve that can self-monitor and self-modify the developing organism? And can special-purpose chemical mechanisms, which involve both continuous and discrete changes, produce results that cannot be produced, or cannot be produced quickly, by Turing-computation?

A third current concern is a development of the Kantian position that he initiated as a graduate student in mathematics, before switching to philosophy. As he puts it: "I've been exploring the conjecture that the precursors of Euclid must have started from something like more general versions of Gibson's abilities to perceive and reason about positive and negative affordances. I've also been trying to isolate the specific forms of human mathematical (especially geometrical) reasoning that current forms of logical, arithmetical, and algebraic theorem proving in AI fail to capture" (p.c.). For example, one can *see* that the angles of a triangle *necessarily* continue to add up to the 'angle' of a straight line as the size and shape of a triangle change, because one has intuitive (non-logical) knowledge of the possibilities involved in the spatial structures concerned. So, again, an intriguing--and potentially game-changing--coupling of philosophy and AI.

Clearly, Sloman's intellectual curiosity, and insightfulness, is as wide-ranging and as sparkling as ever. In short, he's still a bright tile in AI's mosaic. – So:*Watch this space!* More to the point: *Watch his website!*

V: Recognition delayed

There's a puzzle here, however: Sloman's work was under-appreciated for many years. The reason, in a nutshell, is that it didn't (yet) fit in with the *Zeitgeist.*

To be sure, he was always highly respected by the UK's AI community, who could interact with him personally. In verbal discussions, the air would fizz with his searching questions and unexpected insights. The AI-pioneers in the USA knew him personally too, and respected his contributions accordingly. His account of analogical representation (Sloman 1971, 1975), for instance, attracted their interest immediately, and was recognized by them as offering key—albeit "controversial"--ideas in knowledge representation (Brachman and Levesque 1985: 431). But the subsequent generations of USA's AI scientists were less familiar with his research.

In part, that was due to AI's no longer being a tiny research community. In addition, most of his publications were on philosophical topics. (Even these were relatively few in number, since he devoted so much time to helping other people's learning and research: see Section vii.) But the lack of recognition was due also to the fact that POPEYE and MINDER, and the integrative computational philosophy that underlay them, were deeply unfashionable.

Sloman's work on vision was already non-mainstream in the mid-1970s, as we've seen. Most computer vision at that time was conceptualized much more narrowly. But bad--or rather, unlucky--timing soon played a part too.

When David Marr's *Vision* (1982) was published, shortly after his tragically early death, it was instantly influential. Indeed, many people interested in computer vision hadn't waited for the book, having been converted by Marr in the mid-late 1970s. Marr's vision papers were readily available from MIT as AI Memos, one had been published by *Science,* and three by the Royal Society (Marr 1974a,b, 1975a,b,c; Marr and Hildreth 1980; Marr and Nishihara 1978; Marr and Poggio 1976, 1977, 1979). Quite apart from their interest as examples of AI, these publications promised to illuminate the relevant areas of neuroscience--as Marr's earlier work had done for the cortex and cerebellum (Boden 2006: 14.v.b-f). As a result of the huge interest that they aroused, AI scientists asking questions about vision in the late-1970s and 1980s tended to base them in Marr's approach.

In other words, they focussed almost entirely on optically specifiable (and probably innate) bottom-up processes, not on top-down influences from learnt high-level schemata. They ignored questions about the use of vision in motor control. They accepted Marr's key claim that the purpose of vision is to turn the retinal 2D-representation into a representation of the 3D-environment. They emphasized the fact--which Sloman, like most early AI-vision researchers, hadn't stressed (but see 1978: 219)--that we can locate, and describe, visible objects that we have never seen before. And they interpreted the visual scene in terms of surfaces, edges, textures, and 3D-locations—not in terms of identifiable objects, and still less in terms of those objects' potential roles in the organism's behaviour.

As part of the Marrian revolution, Gibson's approach to vision—which had influenced Sloman deeply, as we've seen—was scornfully rejected. Admittedly, this was primarily because Gibson had claimed that low-level vision doesn't involve computation: inevitably, a red rag to all AI bulls (Boden 2006: 7.v.e-f). Marr's theory, of course, identified and modelled many detailed computations going on in low-level vision. Admittedly too, no Marrian would have denied that vision is useful, and often essential, for action. But in their discussions about and modelling of visual perception, the Marrians said nothing about how it is integrated with motor control, or with a changing motivational context. One might almost characterize their approach as the study of *vision without mental architecture*.

In all these ways, then, Marr's work was at odds with Sloman's. As a corollary to becoming deeply unfashionable virtually overnight, Sloman lost his research funding. POPEYE was now so far off the mainstream that it simply stood no chance.

(His account of what happened is given in the historical note added to the online version of his 1978 book. One of the factors he mentions is yet another aspect of the *Zeitgeist*: the widespread move from AI languages such as LISP or POP-11 to more general, more efficient, languages such as Pascal or C/C++. These, he said, make it

very difficult to express "complex operations involving structural descriptions, pattern matching and searching", and to permit "task-specific syntactic extensions ... which allow the features of different problems to be expressed in different formalisms within the same larger program". Today, he says that the same rejection of AI languages may have prevented a wider take-up of the SimAgent toolkit, based as it is on the POP-11 language.)

Sloman's pioneering work on mental architecture, too, was largely ignored by AI scientists for a while. No doubt, that can be explained in part by the unfashionableness of the topic at the time. Only a few computer modellers were thinking about the mind as a whole (e.g. Simon 1962, 1969; Anderson 1983; Minsky 1987; Laird et al. 1987). And, despite Simon's having written about emotions as long ago as the 1960s (Simon 1967), even they were looking at cognition (perception, planning, problem-solving) rather than emotion and/or motivation. Within the psychological and neurological literature too, the emotional aspects of thinking were still mostly ignored. (Minsky was an exception here: his work on "the emotion machine" was circulated for several years before finally being published in 2007.) It wasn't until the 1990s that the concept of "emotional intelligence" became popular, entering not only the newspapers but also AI research (Damasio 1994; Picard, 1997, 1999).

But there was an additional problem. The neglect, especially in the USA's AI community, was due also to the younger generations' dismissal of anything that wasn't actually implemented. As remarked in Section iv, Sloman's studies of architecture, including his account of motivation and emotion, is deeply informed by AI and computational thinking without being presented as computer programs. Now, to be sure, there is MINDER—and the SimAgent environment, too. But for many years no such implementation existed. And even these model only a very small part of Sloman's theory.

This shouldn't matter: although functioning programs are of course a huge strength, computational *thinking* about intelligence is valuable also. Indeed, the latter must precede the former. Both MINDER and SimAgent , after all, resulted from Sloman's methodology of developing and testing *theoretical* architectural ideas about functional requirements, evolutionary origins, and variants in other species. Nevertheless, in the quest for comprehensive programmed models, he couldn't deliver.

As for the philosophers, their professional *Zeitgeist*, also, prevented them from recognizing Sloman's importance quickly. When AI was still young they knew little or nothing about it. Most got no further than the Turing Test, even though Turing himself had made other philosophically interesting claims in his notorious *Mind* article (Boden 2006: 16.ii.b). Nor were many of them persuaded to learn about it when Sloman published *The Computer Revolution in Philosophy* in 1978. Besides being informed by a detailed knowledge of AI which they lacked, and which they therefore could not understand, the book boldly announced that "within a few years

philosophers … will be professionally incompetent if they are not well-informed about these developments [in computing and AI]" (1978: xiii). That announcement was correct--but perhaps hardly tactful. It was bound to raise many philosophers' hackles.

The philosophers' resistance wasn't based purely in annoyance at being told that they were ignorant. It was underpinned by the fact that most of them believed science to be in principle irrelevant to philosophy. (Hence their reluctance to interpret Ryle's philosophy of mind as offering *explanations:* see Section ii.) Anyone who disagreed was accused of "scientism" and/or "psychologism", which Frege--and his translator John Austin, then the high-priest of Oxford philosophy--had denounced as a near-deadly sin (Frege 1884/1950, Preface). My own first book (Boden 1972), which used both AI and psychology to inform philosophical argumentation about mind and personality, had also suffered from this attitude: one highly complimentary review in a philosophical journal ended by saying "… but you can't really call it philosophy".

Even those philosophers who, in the decade following publication of Sloman's book, became willing to grant that AI could be philosophically interesting often missed the point. For instance, when I was putting together a collection of papers on *The Philosophy of Artificial Intelligence* for Oxford University Press in the late-1980s (Boden 1990), one of the publisher's advisers said that Sloman's 'Motives, Mechanisms, and Emotions' (1987a) should be dropped. He—or (very unlikely) she—announced that it was "unrepresentative" and "irrelevant". The adviser was half-right. It was indeed unrepresentative, for it was years ahead of its time (see Section iv). But "irrelevant"...? The mind boggles. Only a narrowly technological, gizmo-seeking, view could have justified such a judgment. I insisted that the paper be included.

VI: The *Zeitgeist* shifts

However, things change. The change of most relevance here is not in Sloman's theoretical approach, although this has of course developed over the years—and is still doing so. Rather, it is in the surrounding intellectual atmosphere. The *Zeitgeist* of AI, in particular, has altered significantly.

POPEYE—or, more accurately, the general philosophy that underlay POPEYE—has recently had something of a revival. For Sloman's 1989 paper on vision was cited by the neuroscientist who, with a psychologist colleague, had recently caused a sensation by positing two visual pathways in the brain--one for perception, the other for action (Goodale and Humphrey 1998: 201).

According to Melvyn Goodale and David Milner, the dorsal pathway locates an object in space relative to the viewer, who can then grasp it; the ventral pathway may (this point is contested) locate it relative to other objects, and enables the

viewer to recognize it (Goodale and Milner 1992, 2003; Milner and Goodale 1993). (The evidence lies partly in brain-scanning experiments with normal people, and partly in clinical cases: damage to these brain areas leads to visual ataxia and visual agnosia, respectively. For example, one patient can recognize an envelope but is unable to post it through a slot, whereas another can post it efficiently but can't say what it is.)

Like Sloman, these 1990s researchers asked how the different pathways can be integrated in various circumstances. But even they didn't really get the point of his work—which was that there are many visual pathways, not just two, and that these deal with many different types of information (e.g. transient versus long-lasting). Of course, Sloman was talking about neural computation, not neuroanatomy: there may or may not be distinct neuroanatomical pathways related to distinct types of function. (In low-level vision, it appears that there are.) The significant point here, however, is that his emphasis on the functional diversity of visual perception is still unusual. Hence my remark, above, that it has had only "something" of a revival.

Besides bearing comparison, up to a point, with Goodale and Milner's work, Sloman's current approach to vision fits in with recent psychological and neuroscientific research on the internal emulation and anticipation of motor control, and on the perceptual feedback provided by motor action. For instance, on his website he describes his own work as being similar in spirit to the "emulation theory of representation" developed by the philosopher Rick Grush (2004).

Grush's theory lies within the general tradition initiated eighty years ago by Kenneth Craik, wherein behaviour is guided by anticipatory mental/cerebral "models" of various kinds (Craik 1943; Boden 2006: 4.iv). Insofar as POPEYE was focussed on the use of a variety of mental representations, one could say that it, too, was situated in this tradition. But Sloman's focus has shifted. Today, he thinks of vision less as the analysis and interpretation of *representational structures* than as the analysis and interpretation of *informational processes*, with many different knowledge bases, and varying types of representation, acting (cooperating and competing) concurrently.

An even greater change has occurred in the attitude of AI researchers to work on emotion. The AI community has now woken up to the importance of emotion—although their interest is largely motivated by their wish to develop potentially lucrative gizmos. Current research on "companion robots" and the like tries to give AI systems recognition of, responsiveness to, and sometimes even simulation of, human emotions (Dautenhahn 2002; Wilks 2010).

Much of this work, it must be said--and has been said (Sloman 1999, 2001; cf. Picard 1999)--is shallow, for two reasons. First, many researchers still don't appreciate the degree of mental diversity that Sloman sketched years ago. Second, their primary aim (often) is not to understand how emotion functions in the control of other mental processes (including perception, thinking, and motivation as well as

action). Rather, it is to reassure, or even deceive, the human users of computer companions by providing a superficial appearance of emotional understanding and response on the part of the machine.

Whether gizmo-driven or not, however, the new interest in such matters has helped to draw international attention to Sloman's research on emotion, and on mental architecture in general. For instance, in 2002 DARPA invited him to take part in a small workshop on their new cognitive systems initiative, where his work was discussed in one of the introductory papers. Two years later, the AAAI held a cross-disciplinary symposium on "Architectures for Modelling Emotion". The EU also took an interest: it funded the four-year CoSy project, begun in 2004 (Christensen et al. 2010), and the CogX project (2008-2012), led by Sloman's colleague Jeremy Wyatt (<http://cogx.eu>). These projects have produced a number of working models (research demos, not usable gizmos), but—like MINDER—these reflect only a small subset of Sloman's theoretical ideas.

Besides advising on large-scale (collaborative) projects such as these, and receiving many other invitations to speak, Sloman has been appointed as one of the leaders of a project addressing a "Grand Challenge" of British computing (see below). In this project, the architectural functions of emotion (in robots, as well as humans) are more important than the presentation of apparently emotional machine-companions to naïve users.

Even the philosophers—well, some of them—have woken up to the importance of Sloman's work. That's due in large part to a change in the philosophical background. (The *Zeitgeist,* again.) Analytically-minded philosophers are now more ready to take account of scientific concepts and findings than they were in the mid-twentieth century. And some of them have specifically concerned themselves with AI (and sometimes with the concepts of computation and/or information), whether to defend or to reject its potential for illuminating the philosophy of mind. (The defenders include Grush, Jerry Fodor, Daniel Dennett, Paul Churchland, Steven Harnad, Andy Clark, Michael Wheeler, Brian Smith, Ronald Chrisley, Jack Copeland, Luciano Floridi, John Pollock, and myself; the sceptics include John Searle, Hubert Dreyfus, John Haugeland, Timothy van Gelder, Roger Penrose, Selmer Bringsjord, and Ned Block.)

However, most philosophers are still largely ignorant of the AI details, so cannot engage with Sloman's work in a truly productive fashion. Moreover, the Turing Test, not to mention the Chinese Room (Searle 1980), still rears its ugly head far too often. The attempts of Sloman (1996a, 2002), and others, to defuse this ever-ticking bomb have not been taken to heart by his philosopher colleagues.

In addition, philosophers' ignorance is still often bolstered by philosophical principle. Dismissive charges of scientism are mounted by thinkers on the phenomenological side of the Anglo-Saxon/Continental, or realist/constructivist, divide (Boden 2006: 16.vi-viii). Unfortunately, these people—who don't even bother

to read AI-based work--now comprise a larger fraction of the philosophical community than they did when Sloman was a young man.

The ideas of the later Ludwig Wittgenstein (1953), who denied any place for computational (i.e. sub-personal) theories in psychology, have been used to attack cognitive science in general (e.g. Bennett and Hacker 2003). The Wittgensteinian philosopher Richard Rorty explicitly hoped for "the disappearance of psychology as a discipline distinct from neurology," including the demise of *computational* psychology (1979:121). To make matters worse, several prominent writers originally trained in the analytic tradition, such as John McDowell (1994), have adopted a phenomenological view according to which no naturalistic explanation of psychology is in principle possible. Even the founder of Turing-machine functionalism, who once urged philosophical comparisons between minds and computers, has reneged and turned to broadly constructivist accounts (Putnam 1967, 1982, 1988, 1997, 1999). In short, many philosophers today are just as loath to take Sloman's work seriously as they were in the 1970s.

Happily, more appreciation has come from other areas of cognitive science, as we've seen. If too many philosophers still steer clear of Sloman's work, because of their ignorance of AI and/or their suspicion of scientism, today's AI researchers do not.

In some AI-watchers' minds, to be sure, the appreciation pendulum has swung much too far in Sloman's favour--as he's the first to admit. Having become well known in AI circles for his work on emotions, he received an unexpected, and ridiculous, request. In his words: "I've even had someone from a US government-funded research centre in California phone me a couple of months ago [i.e. mid-2002] about the possibility of modelling emotional processes in terrorists. I told him it was beyond the state of the art. He told me I was the first person to say that: everyone else he contacted claimed to know how to do it (presumably hoping to attract research contracts)" (Sloman p.c.). (Probably, those other people weren't merely being opportunist, making promises they couldn't keep in order to board the Pentagon/Whitehall/EU band-waggons now funding research on "emotional" robots and "social" human-computer interactions. In addition, they didn't realize the depth and complexity of the mental-computational architecture that's required to generate emotional phenomena.)

Some people might accuse me, too, of valuing his work too highly. For in the final chapter of my recent book on the history of cognitive science, I listed a couple of dozen instances of research in this interdisciplinary field that I regard as especially promising--and said that if I were forced to choose only one, it would be Sloman's approach to integrated mental architecture (Boden 2006: 1449). Indeed, I'd already done that, when (on the 50th anniversary of the 1953 discovery of the double helix) the British Association for the Advancement of Science invited several people to write 200 words for their magazine *Science and Public Affairs* on "what discovery/advance/development in their field they think we'll be celebrating in 50

years' time". Others might well have prioritized a different item—or perhaps something not included on my list at all.

My choice, admittedly, was influenced by my own long-time interest, since high-school days, in personality and psychopathology (Boden 2006: Preface.ii). But it wasn't idiosyncratic. Two years later, the UK's computing community (of which AI researchers are only a subset) voted for "The Architecture of Brain and Mind" as one of the seven "Grand Challenges" for the future. (See http://www.uk.crc.org.uk/Grand_Challenges/index.cfm and http://www.cs.sir.ac.uk/gc5.) What's more, Sloman was appointed as a member of the five-man committee carrying this project forward.

In brief, many AI scientists, if not the committed gizmo-seekers, would endorse my valuation. They might do so while 'twinning' Sloman with Minsky, whose broadly similar research has considered architectural issues in more detail than is usual within AI (Minsky 1985, 2007). They might even judge Minsky's work above Sloman's, especially if they have little interest in philosophical questions. But I'd be very surprised if anyone seriously concerned with the nature of minds as a whole were not to appreciate Sloman's contribution.

VII: Coda

I've focussed only on Sloman's own intellectual work. But I must also mention his importance as an instigator of AI research and education, in the UK and elsewhere.

Having spent a year at the University of Edinburgh's Machine Intelligence Unit, with Donald Michie and Bernard Meltzer (and many younger researchers in AI), he was a main driver in setting up the Cognitive Studies Programme at the University of Sussex in the early-1970s. The other founders of this interdisciplinary venture included the charismatic Max Clowes (an imaginative early researcher in computer vision: Clowes 1967, 1969, 1971), Alistair Chalmers (a highly computer-literate social psychologist), and myself (already using AI to understand the mind: Boden 1965, 1970, 1972, 1973). As the world's first academic programme to integrate AI with philosophy, psychology, and linguistics, COGS became internationally recognized, and widely influential.

As part of this educational project, Sloman (with colleagues such as John Gibson and Steven Hardy) developed the highly user-friendly POPLOG programming system, soon to be used by other universities and by various commercial institutions. Later, he took a key role in the UK's government-backed Alvey Programme, which encouraged AI knowledge transfer between academia and industry (Boden 2006: 11.iv-v). Partly due to his influence, the Alvey remit was broadened from logic programming and expert systems to include vision and neural computation too.

Over the years, a number of other important advisory/administrative roles in the UK's and Europe's computing community followed. For instance, in 2003 he was one of six AI researchers consulted about the new EU Cognitive Systems Initiative. And since 2009 he has been heavily involved in the UK's Computing at School initiative—trying, among other things, to make this focussed more on using AI/computing to understand the mind and less on using or generating gizmos, a.k.a. apps (http://www.computingatschool.org.uk>).

In short, even setting aside his own research, Sloman has been--and continues to be--prominent in virtue of the insightful advice he has given on AI and computing in the UK and beyond.

Finally, Sloman as a person. He could not have achieved the degree of intellectual leadership he has exercised at many different levels without being someone who drew affection, as well as respect, from others. That affection was largely earned by his own unfailing respectfulness for those who came in contact with him. Add to this, his exceptional generosity in helping his colleagues and students--a generosity that devoured precious time that could have been spent more selfishly.

I myself have benefitted from this on various occasions. My book *Artificial Intelligence and Natural Man* (1977) was much improved by his advice, and contained this seemingly bland but actually heartfelt Acknowledgment: "I am deeply grateful to Aaron Sloman for his careful reading of the draft manuscript, and for many conversations on related topics". In addition, he has helped me countless times, with admirable patience, to cope with the technology—as he has done for many others, too. In the 50 years that I've known him (since 1962), Aaron has been my most intellectually stimulating colleague, and a very dear friend.

References

Anderson, J. R. (1983), *The Architecture of Cognition* (Cambridge, Mass.: Harvard University Press).

Baker, S. (2011), *Final Jeopardy: Man vs. Machine and the Quest to Know Everything* (Boston: Houghton Mifflin Harcourt).

Beaudoin, L. P. (1994), *Goal Processing in Autonomous Agents*, Ph.D. thesis, School of Computer Science, University of Birmingham, available at http://www.cs.bham.ac.uk/research/cogaff/.

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

Boden, M. A. (1959), 'In Reply to Hart and Hampshire', *Mind*, NS 68: 256-60.

Boden, M. A. (1965), 'McDougall Revisited', *Journal of Personality,* 33: 1-19. Reprinted in M. A. Boden, *Minds and Mechanisms: Philosophical Psychology and Computational Models* (Ithaca: Cornell University Press, 1981), pp. 192-208.

Boden, M. A. (1969), 'Machine Perception', *Philosophical Quarterly*, 19: 32-45.

Boden, M. A. (1970), 'Intentionality and Physical Systems', *Philosophy of Science*, 37: 200-14.

Boden, M. A. (1972), *Purposive Explanation in Psychology* (Cambridge, Mass.: Harvard University Press).

Boden, M. A. (1973), 'How Artificial is Artificial Intelligence?', *British Journal for the Philosophy of Science*, 24: 61-72.

Boden, M. A. (1977), *Artificial Intelligence and Natural Man* (New York: Basic Books). (2nd edn., expanded, 1987. London: MIT Press; New York: Basic Books.)

Boden, M. A. (ed.) (1990), *The Philosophy of Artificial Intelligence* (Oxford: Oxford University Press).

Boden, M. A. (2006), *Mind as Machine: A History of Cognitive Science* (Oxford: Clarendon/Oxford University Press).

Brachman, R. J., and Levesque, H. J. (eds.) (1985), *Readings in Knowledge Representation* (Los Altos, CA: Morgan Kauffman).

Brooks, R. A. (1990), 'Elephants Don't Play Chess', *Robotics and Autonomous Systems,* 6: 3-15*.*

Christensen, H. I., Kruijff , G-J.M., and J.L. Wyatt (eds.) (2010), *Cognitive Systems.* Cognitive Systems Monographs vol 8 (Berlin: Springer).

Clark, A. J. (2013). 'Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science', *Behavioral and Brain Sciences*, in press.

Clowes, M. B. (1967), 'Perception, Picture Processing, and Computers', in N. L. Collins and D. M. Michie (eds.), *Machine Intelligence 1* (Edinburgh: Edinburgh University Press), 181-97.

Clowes, M. B. (1969), 'Pictorial Relationships—A Syntactic Approach', in B. Meltzer and D. M. Michie (eds.), *Machine Intelligence 4* (Edinburgh: Edinburgh University Press), 361-83.

Clowes, M. B. (1971), 'On Seeing Things', *Artificial Intelligence*, 2: 79-116.

Craik, K. J. W. (1943), *The Nature of Explanation* (Cambridge: Cambridge University Press).

Crevier, D. (1993), *AI: The Tumultuous History of the Search for Artificial Intelligence* (New York: Basic Books).

Damasio, A. R. (1994), *Descartes' Error: Emotion, Reason, and the Human Brain* (New York: Putnam).

Dautenhahn, K. (ed.), (2002), *Socially Intelligent Agents: Creating Relationships with Computers and Robots* (Boston: Kluwer Academic).

Dennett, D. C. (1988), 'Quining Qualia', in A. Marcel and E. Bisiach (eds.), *Consciousness in Contemporary Science* (Oxford: Oxford University Press), 42-77.

Dennett, D. C. (1991), *Consciousness Explained* (London: Allen Lane).

Fisher, M. (1990), *Personal Love* (London: Duckworth).

Frege, G. (1884/1950). *The Foundations of Arithmetic*. Trans. J. L. Austin. (Oxford: Oxford University Press), 1950.
Frith, U. (1989/2003), *Autism: Explaining the Enigma* (Oxford: Blackwell; 2nd edn., rev., 2003).

Gandy, R. (1996), 'Human Versus Mechanical Intelligence', in P. J. R. Millican and A. J. Clark (eds.), *Machines and Thought: The Legacy of Alan Turing. Vol. I* (Oxford: Oxford University Press), 125-136.

Gibson, J. J. (1966), *The Senses Considered as Perceptual Systems* (Westport, Conn.: Greenwood Press).

Goodale, M. A., and G. K. Humphrey (1998), 'The Objects of Action and Perception', *Cognition* (67): 181-207.

Goodale, M. A., and Milner, A. D. (1992), 'Separate Visual Pathways for Perception and Action', *Trends in Neuroscience*, 13: 20-23.

Goodale, M. A., and Milner, A. D. (2003), *Sight Unseen: An Exploration of Conscious and Unconscious Vision* (Oxford: Oxford University Press).

Grush, R. (2004), 'The Emulation Theory of Representation', *Behavioral and Brain Sciences* (27): 377-442.

Laird, J. E., Newell, A., and Rosenbloom, P. (1987), 'Soar: An Architecture for General Intelligence', *Artificial Intelligence*, 33: 1-64.

McCarthy, J., and Hayes, P. J. (1969), 'Some Philosophical Problems from the Standpoint of Artificial Intelligence', in B. Meltzer and D. M. Michie (eds.), *Machine Intelligence* 4 (Edinburgh: Edinburgh University Press), pp. 463-502.

McDowell, J. (1994), *Mind and World* (Cambridge, Mass.: Harvard University Press).

Marr, D. C. (1974a), 'The Computation of Lightness by the Primate Retina', *Vision*, 14: 1377-1388.

Marr, D. C. (1974b), *A Note on the Computation of Binocular Disparity in a Symbolic, Low-Level Visual Processor*. Cambridge, Mass.: MIT AI-Lab Memo no. 327. Reprinted in L. Vaina (ed.), *From the Retina to the Neocortex: Selected Papers of David Marr* (Boston: Birkhauser,1991), 231-238.

Marr, D. C. (1975a), *Analyzing Natural Images: A Computational Theory of Texture Vision.* AI Memo 334. Cambridge, Mass.: MIT AI Lab., June 1975.

Marr, D. C. (1975b), *Early Processing of Visual Information*. AI Memo 340. Cambridge, Mass.: MIT AI Lab., December 1975. Officially published in *Philosophical Transactions of the Royal Society: B,* 275 (1976), 483-524.

Marr, D. C. (1975c), 'Approaches to Biological Information Processing', *Science*, 190: 875-876.

Marr, D. C. (1979), 'Representing and Computing Visual Information', in P. H. Winston and R. H. Brown (eds.), *Artificial Intelligence: An MIT Perspective,vol. 2* (Cambridge, Mass.: MIT Press), 17-82.

Marr, D. C. (1982), *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (San Francisco: Freeman).

Marr, D. C., and Hildreth, E. (1980), 'Theory of Edge-Detection', *Proceedings of the Royal Society: B,* 207: 187-217.

Marr, D. C., and Nishihara, H. K. (1978), 'Visual Information Processing: Artificial Intelligence and the Sensorium of Sight', *Technology Review*, 81: 2-23.

Marr, D. C., and Poggio, T. (1976), 'Cooperative Computation of Stereo Disparity', *Science*, 194: 283-287.

Marr, D. C., and Poggio, T. (1977), 'From Understanding Computation to Understanding Neural Circuitry', *Neuroscience Research Program Bulletin,* 15: 470-488.

Marr, D. C., and Poggio, T. (1979), 'A Computational Theory of Human Stereo Vision', *Proceedings of the Royal Society: B*, 204: 301-328.

Milner, A. D., and Goodale, M. A. (1993), 'Visual Pathways to Perception and Action', in T. P. Hicks, S. Molotchnikoff and T. Ono (eds.), *Progress in Brain Research*, vol. 95 (Amsterdam: Elsevier), 317-337.

Milner, A. D., and Goodale, M. A. (1995), *The Visual Brain in Action* (Oxford: Oxford University Press).

Minsky, M. L. (1965), 'Matter, Mind, and Models', *Proceedings of the International Federation of Information Processing Congress*, 1, 45-49 (Washington D.C.: Spartan).

Minsky, M. L. (1985), *The Society of Mind* (New York: Simon & Schuster).

Minsky, M. L. (2007) *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of Human Mind* (

Newell, A. (1973), 'Artificial Intelligence and the Concept of Mind', in R. C. Schank and K. M. Colby (eds.), *Computer Models of Thought and Language* (San Francisco: Freeman), pp. 1-60.

Newell, A. (1980), 'Physical Symbol Systems', *Cognitive Science, 4: 135-83.*

Newell, A., and Simon, H. A. (1961), 'GPS - A Program that Simulates Human Thought', in H. Billing (ed.), *Lernende Automaten* (Munich: Oldenbourg), 109-124. Reprinted in E. A. Feigenbaum and J. A. Feldman (eds.), *Computers and Thought* (New York: McGraw-Hill, 1963, pp. 279-293.

Picard, R, W. (1997), *Affective Computing* (Cambridge, Mass.: MIT Press).

Picard, R. W. (1999), 'Response to Sloman's Review of Affective Computing', *AI Magazine,* 20/1 (March), 134-7.

Putnam, H. (1967), 'The Nature of Mental States'. First published as 'Psychological Predicates' in W. H. Capitan and D. Merrill (eds.), *Art, Mind, and Religion* (Pittsburgh: University of Pittsburgh Press), 37-48. Reprinted in H. Putnam, *Mind, Language, and Reality: Philosophical Papers, vol. 2* (Cambridge: Cambridge University Press, 1975), 429-440.

Putnam (1982), 'Why There Isn't a Ready-Made World', *Synthese*, 51: 141-167.

Putnam, H. (1988), *Representation and Reality* (Cambridge, Mass.: MIT Press).

Putnam, H. (1997), 'Functionalism: Cognitive Science or Science Fiction?', in D. M. Johnson and C. E. Erneling (eds.), *The Future of the Cognitive Revolution* (Oxford: Oxford University Press), 32-44.

Putnam, H. (1999), *The Threefold Cord: Mind, Body, and World* (New York: Columbia University Press).

Rorty, R. (1979), *Philosophy and the Mirror of Nature* (Princeton: Princeton University Press).

Ryle, G. (1949), *The Concept of Mind* (London: Hutchinson's University Library).

Simon, H. A. (1962), 'The Architecture of Complexity', *Proceedings of the American Philosophical Society*, 106 (1962), 467-482.

Simon, H. A. (1967), 'Motivational and Emotional Controls of Cognition', *Psychological Review*, 74: 29-39.

Simon, H. A. (1969), *The Sciences of the Artificial. The Karl Taylor Compton Lectures* (Cambridge, Mass.: MIT Press). (2nd & 3rd edns., 1981 and 1996.)

Sloman, A. (n.d.), The *CogAff* group's website: www.cs.bham.ac.uk/research/cogaff

Sloman, A. (1971), 'Interactions between Philosophy and Artificial Intelligence: The Role of Intuition and Non-Logical Reasoning in Intelligence', *Artificial Intelligence*, 2: 209-225.

Sloman, A. (1974), 'Physicalism and the Bogey of Determinism', in S. C. Brown (ed.), *Philosophy of Psychology* (London: Macmillan), 283-304.

Sloman, A. (1975), 'Afterthoughts on Analogical Representation', in R. C. Schank and B. L. Nash-Webber (eds.), *Theoretical Issues in Natural Language Processing: An Interdisciplinary Workshop in Computational Linguistics, Psychology, Linguistics, and Artificial Intelligence,* Cambridge, Mass., 10-13 June (Arlington, Va.: Association for Computational Linguistics), 164-168. Reprinted in Brachman and Levesque 1985, pp. 431-39.

Sloman, A. (1978), *The Computer Revolution in Philosophy: Philosophy, Science, and Models of Mind* (Brighton: Harvester Press). Out of print, but available - and continually updated - online at http://www.cs.bham.ac.uk/research/cogaff/crp/.

Sloman, A. (1982), 'Towards a Grammar of Emotions', *New Universities Quarterly*, 36: 230-238.

Sloman, A. (1983), 'Image Interpretation: The Way Ahead?', in O. J. Braddick and A. C. Sleigh (eds.), *Physical and Biological Processing of Images* (New York: Springer-Verlag), 380-40.

Sloman, A. (1986), 'What Sorts of Machine Can Understand the Symbols They Use?', *Proceedings of the Aristotelian Society*, Supp., 60: 61-80.

Sloman, A. (1987a), 'Motives, Mechanisms, and Emotions', *Cognition and Emotion*, 1: 217-233. Reprinted in Boden 1990: 231-247.

Sloman, A. (1987b), 'Reference Without Causal Links', in J. B. H. du Boulay. D. Hogg and L. Steels (eds.), *Advances in Artificial Intelligence – II* (Dordrecht: North Holland), 369-381.

Sloman, A. (1989), 'On Designing a Visual System: Towards a Gibsonian Computational Model of Vision', *Journal of Experimental and Theoretical AI*, 1: 289-337.

Sloman, A. (1992), 'The Emperor's Real Mind, Review of Roger Penrose's *The Emperor's New Mind: Concerning Computers Minds and the Laws of Physics'*, *Artificial Intelligence*, 56: 355-396.

Sloman, A. (1993), 'The Mind as a Control System', in C. Hookway and D. Peterson (eds.), *Philosophy and the Cognitive Sciences* (Cambridge: Cambridge University Press), 69-110.

Sloman, A. (1995), 'Sim_Agent help-file', available at ftp://ftp.cs.bham.ac.uk/pub/dist/poplog/sim/help/sim_agent. See also 'Sim_agent web-page, available at http://www.cs.bhma.ac.uk/axs/cog_affect/sim_agent.html.

Sloman, A. (1996a), 'Beyond Turing Equivalence', in P. J. R. Millican and A. J. Clark (eds.), *Machines and Thought: The Legacy of Alan Turing, vol 1* (Oxford: Oxford University Press), 179-220.

Sloman, A. (1996b), 'Towards a General Theory of Representations', in D. M. Peterson (ed.), *Forms of Representation: An Interdisciplinary Theme for Cognitive Science* (Exeter: Intellect Books), 118-140.

Sloman, A. (1996c), 'Actual Possibilities', in L. C. Aiello and S. C. Shapiro (eds.), *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifth International Conference (KR '96)* (San Francisco: Morgan Kaufmann), 627-638.

Sloman, A. (1998), 'Ekman, Damasio, Descartes, Alarms and Meta-management', *Proceedings of the International Conference on Systems, Man, and Cybernetics SMC98)* (San Diego: IEEE Press), 2652-7.

Sloman, A. (1999), 'Review of [R. Picard's] Affective Computing', *AI Magazine*, 20:1 (March), 127-133.

Sloman, A. (2000), 'Architectural Requirements for Human-like Agents Both Natural and Artificial. (What Sorts of Machines Can Love?)', in K. Dautenhahn (ed.), *Human Cognition and Social Agent Technology: Advances in Consciousness Research* (Amsterdam: John Benjamins), 163-195.

Sloman, A. (2001), 'Beyond Shallow Models of Emotion', *Cognitive Processing: International Quarterly of Cognitive Science*, 2: 177-198.

Sloman, A. (2002), 'The Irrelevance of Turing Machines to Artificial Intelligence', in M. Scheutz (ed.), *Computationalism: New Directions (Cambridge, Mass.: MIT Press),* 87-127.

Sloman, A. (2003), 'How Many Separately Evolved Emotional Beasties Live Within Us?', in R. Trappl, P. Petta and S. Payr (eds.), *Emotions in Humans and Artifacts* (Cambridge, Mass.: MIT Press), 29-96.

Sloman, A. (2009a), 'Requirements for a Fully Deliberative Architecture (or Component of an Architecture)'. Available on the CogAff website: http://www.cs.bham.ac.uk/research/projects/cog-aff.

Sloman, A. (2009b), 'Some Requirements for Human-like Robots: Why the recent over-emphasis on embodiment has held up progress', in B. Sendhoff, E. Koerner, O. Sporns, H. Ritter, and K. Doya (eds.), *Creating Brain-like Intelligence: From Basic Principles to Complex Intelligent Systems.* Lecture Notes in Computer Science vol. 5436 (Berlin: Springer-Verlag), pp. 248-277.

Sloman, A. (2013) 'What Else Can Brains Do?: Commentary on A. Clark's 'Whatever Next?', *Behavioral and Brain Sciences*, in press,

Sloman, A., and Chrisley, R. L. (2003), 'Virtual Machines and Consciousness', in O. Holland (ed.), *Machine Consciousness* (Exeter: Imprint Academic): pp. 133-172.

Sloman, A., and Croucher, M. (1981), 'Why Robots Will Have Emotions', *Proceedings of the Seventh International Joint Conference on Artificial Intelligence* (Vancouver), 197-202.

Sloman, A., and Poli, R. (1995), 'Sim_Agent: A Toolkit for Exloring Agent Designs', in M. Wooldridge, J.-P. Muller, and M. Tambe (eds.), *Intelligent Agents, ii* (Berlin: Springer-Verlag), 392-407.

Squires, R. (1970), 'Are Dispositions Lost Causes?', *Analysis,* 31: 15-18.

Turing, A. M. (1950), 'Computing Machinery and Intelligence',_*Mind*, 59 (1950): 433-60. Reprinted in (Boden 1990: 40-66).

Wilks, Y. A., (ed.) (2010), *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues* (Amsterdam: John Benjamins).

Wittgenstein, L. (1953), *Philosophical Investigations*, trans. G. E. M. Anscombe (Oxford: Blackwell).

Wright, I. P. (1997), *Emotional Agents*, PhD. thesis, School of Computer Science, University of Birmingham, available at <http:///www.cs.bham.ac.uk/research/cogaff/>.

Wright, I. P., and A. Sloman (1997), *MINDER1: An Implementation of a Protoemotional Agent Architecture.* Technical Report CSRP-97-1, University of Birmingham, School of Computer Science, available at ftp://ftp.cs.bham.ac.uk/pub/tech-reports/1997/CSRP-97-01.ps.gz.