METHODOLOGICAL LINKS BETWEEN
AI AND OTHER DISCIPLINES

Margaret, A. Boden

This paper will be published in _Knowledge; Creation, Diffusion, Utilization_ (_Dec. 1982_)

Cognitive Science Research Paper

Serial no: CSRP 020

The University of Sussex
Cognitive Studies Programme
School of Social Sciences
Falmer
Brighton BN1 9QN

55X

020

METHODOLOGICAL LINKS BETWEEN AI AND OTHER DISCIPLINES

Margaret A. Boden
Professor of Philosophy and Psychology, University of Sussex

(To be published in Knowledge: Creation, Diffusion, Utilization (Dec. 1982).
Also for project of F. Machlup, Knowledge: Its Creation, Distribution, and Economic Significance (ten vols), Vol IV: The Disciplines of Information (Princeton University Press).

Whether our interest is in psychology, philosophy, or linguistics, there is no question but that AI has given us a new standard of rigour, and a new appreciation of the importance of mental process. Linguistics already had rigour but not process, psychology had little of either, and philosophy had less of each. AI provides a range of precisely definable computational concepts, specifying various symbolic representations and transformations, with which to conceptualize the mind. And the technology of programming makes manageable a degree of theoretical complexity that would overwhelm the unassisted human brain. So the inadequacy of theoretical approaches that fail to recognize the complexity of mental structure and process is now evident, and psychology and the philosophy of mind have been influenced accordingly.

One example of a class of empirical psychological work partly inspired by AI-ideas is microdevelopmental research [e.g. Inhelder & Karmiloff-Smith, 1975; Karmiloff-Smith, 1979]. This studies the dialectical interplay between the child's action-sequences and changing cognitive representations (theories, models, heuristics, choice-criteria). The specifics of action are emphasized, on the assumption that the procedural details of performance (not only its overall structure) give clues to the underlying competence. However, the degree of procedural detail -- though high relative to more traditional forms of experimentation in psychology -- is inadequate to express a complete computational model of the psychological processes concerned. It is not a straightforward matter to assess such studies in computational terms, and we need to learn how to refine the theories and methodology of these studies so as to facilitate such assessment.

This case exemplifies the general point that, if we ask whether AI has given us new discoveries as well as a new approach, the reply might be that it has not been as helpful to working psychologists as its supporters initially hoped. There has been an increasing amount of computationally-influenced empirical research in cognitive, developmental, and educational psychology. But (with the arguable exception of vision) we have gained little new insight into the actual details of mental life, as opposed to the sorts of questions that it may be appropriate to ask.

Is this because psychologists have not yet learnt how to apply AI fruitfully to further their research, or because (as some critics claim) it is in principle unsuitable for psychological modelling? This question raises a number of methodological difficulties and conceptual unclarities in applying AI-ideas to other disciplines. Some of these involve commitments on basic theoretical or philosophical issues, and call for cooperative research by people in various specialties.

There is much disagreement -- and not a little scepticism in some quarters -- about the extent to which empirical psychological work should or can be planned and assessed in the light of computational ideas. It is not even agreed whether or not psychologists sympathetic to the computational approach should seek to express their theories in programmable (or programmed) terms, as opposed to merely bearing computational issues in mind in their work. Some AI-workers even believe that doing psychological experiments is not an intellectually justifiable exercise in our present state of ignorance, arguing that we should concentrate on clarifying the range of possible computational mechanisms before trying to discover which ones are actually used by living creatures.

Correlatively, there is disagreement over the psychological relevance of specific examples of work within AI. Some of this disagreement is grounded not in detailed objections, but in broad philosophical differences over the potential psychological relevance of facts about neurophysiology or hardware.

For instance, there are two "streams" of work within AI vision research, each of which has spurred psychological experimentation. The theoretical emphases of these two streams are different, and to some degree opposed. One is focussed on low-level computational mechanisms, while the other is focussed on higher-level, top-down processes in scene-analysis. The former (especially in the work of David Marr and his group [Marr, 1982]) takes account of psychological optics and neurophysiology in some detail. But the latter considers optics only in very general terms, and ignores neurophysiology on the principled ground that physiological (hardware) implementation is theoretically independent of questions about computational mechanisms.

This last is a widely shared view in AI (in some quarters approaching the status of a dogma), and one which has caused many physiologically-minded psychologists to doubt the usefulness of AI work. It is a position that is correct in principle but possibly sometimes misleading in practice. In an abstract theoretical sense, all computing devices are equivalent, just as all programming languages are. But to ignore the varying computational powers of distinct (electronic or physiological) hardware may be as stultifying in practice as to try to use a single programming language for all programs. The differences between programming languages often matter: a computation that can be expressed easily if one uses the representational potential of one language may be difficult, or even practically infeasible, if one relies on another. Clearly, further computationally-informed work on neurophysiological mechanisms is needed. It may be that physiology is relevant to the relatively peripheral processing but irrelevant at higher levels, but the precise points at which one may expect physiology to have a casting vote are controversial (some of Marr's earlier work on the cerebellum, for instance, is now attracting interest within AI).

If one could prove that a particular computation simply could not be carried out in realtime by any existing cerebral mechanism, then the use of "alien" computer hardware to effect it would be psychologically irrelevant. However, our ignorance of both computational and neurophysiological constraints preempts such proofs. Nor can we prove that only mechanisms like those in our brains are capable of certain of a programmed model is that it is consonant at some significant level with neurophysiology.

This claim is made, for example, in support of a very recent advance in the computational modelling of vision [Hinton, 1981]. Hinton's work is focussed on low-level, dedicated hardware, mechanisms that are capable of cooperative computation, or parallel processing. Although it is not a simulation of detailed neurophysiology, Hinton believes it to be a prime strength of his model that it is compatible with what is known about nervous function. For instance, it relies on excitatory and inhibitory connections between computational units on various levels that appear to have an analogue in the nervous connectivity of our own visual system.

Critics of AI often complain that one program does not make a theory, any more than one swallow makes a summer. That is, AI is accused of being "empirical" in the sense in which much of medicine is, that it achieves practical results by methods it does not understand and which it therefore cannot responsibly generalize. This is indeed a methodological shortcoming of much AI work -- but not of all. Thus Hinton's research is especially interesting because it provides not only an *example* of a program that achieves a desired result (the "perception" of shape), but also a *general proof* that results of this class can be computed by computational systems of this form that are within specific size-constraints. In brief, he has proved that many fewer computational units are necessary for the parallel computation of shape than one might initially have supposed. This proof lends some more physiological weight to the model, since the human retina apparently has enough cells to do the job.

Because Hinton's model of vision utilizes a type of computation fundamentally different from that of "traditional" AI, it raises the question of just which psychological phenomena AI can be used to illuminate, and which it cannot. Hinton's results suggest that parallel-processing systems can perform shape discriminations -- such as recognition of an overall Gestalt -- commonly believed (even within AI) to require relatively high-level interpretative processes. They suggest also that the way in which an object is represented may be radically different depending on whether it is perceived as an object in its own right or as a part of some larger whole. This might account for the phenomenological differences between perceptual experiences of which we are reminded by those philosophers [e.g. Dreyfus, 1972] who argue that AI is essentially unfitted to model human minds. In general, commonly-expressed philosophical criticisms of AI and cognitive psychology that assume serial processing may be invalidated by these recent developments.

This would be doubly true if the computational techniques of this work on vision can be generalized to other domains. Hinton believes, for example, that his computational model of spatial relations enables motor control to be understood in a new way, one that is significantly analogous to the mechanisms of muscular control in the human body. Phenomenologically influenced philosophers, as well as scientists concerned with the psycho-physiology of movement, commonly complain that AI does not -- or even cannot -- model the body. Many philosophers and psychologists argue that human intelligence is rooted in our embodiment as material beings situated in a material world, and see AI as therefore radically irrelevant. Most current computers do not nave 'Doaies' that can move in and manipulate the external world, and even robots are currently very crude in their motor abilities. But Hinton's preliminary work on motor control suggests an efficient way of computing a jointed limb's movements and pathway through space (a problem that can be solved by traditional computing techniques only in a highly inefficient manner)•

Even where psychologists deliberately match experimental results against theories expressed in programmed form (e.g., Newell & Simon's [1972D work on problem-solving), the psychological relevance of the computational model is debatable. It is not always clear just which aspects of a program one might plausibly expect to be open to empirical test. Some aspects are not intended to have any psychological reality, but are included merely to produce a program that will run. However, one cannot be sure that none of these last have any psychological significance, since it is a prime claim of AI that it can highlight procedural lacunae in our theories and offer us new concepts with which to jump the gap. Nor is the methodology of protocol-matching unproblematic: what is one to conclude from the fact that no behavioural protocol is observed to match a specific process posited by the programmed theory, or that some matching protocol j£ observed? These problems (which have analogous forms to trouble all experimental psychologists) have been discussed by both proponents and opponents of AI, but there is no consensus about the extent to which they cast doubt on a computational approach to empirical psychology.

Of the many people who would concede that certain aspects (at least) of vision, language-use, and problem-solving might yield to an AI approach, some may feel that social psychology, for instance, has nothing to gain from computational insights [e.g. Gauld & Shotter, 1977D. This should not be too hastily assumed, however. Work within AI on the structure of action and the attribution of intentions is relevant to theoretical discussions in social psychology. In general, AI supports the view that there may be generative rules underlying social interaction, or that social perception is a structured interpretative activity. But although these ideas are essentially consonant with a computational viewpoint, specifying them in a particular case is a notoriously difficult matter.

A general account of what sorts of psychological phenomena are or are not grist to the Al-mill would of course be \/ery useful. But firm intellectual ground could be provided for such an account only by a systematic theory of representation. Philosophical discussions of the nature of intentionality are clearly relevant [e.g. Fodor, 1981; Dennett, 1979D. Some philosophers [e.g. Searle, 1980D argue that AI cannot model genuine (biological) intentionality, although the discussions in recent issues of the peer-commentary journal Behavioral and Brain Sciences show this claim to be highly controversial. But even Searle admits that it can provide a scientifically useful metaphor for intentionality. This is why AI is potentially relevant to studies that are normally thought of as being "humanistically^{11} oriented, such as social and clinical psychology [Boden, 1972D. Given that "representational^{11} processes in computer models can function as heuristically fruitful analogues of representational processes in our minds, the problem remains of providing an account of the range and efficacy of such processes.

inference in significantly different ways. Hinton^{f}s work previously mentioned is one of the many examples that address such issues. Another is Amarel's C1968D comparison of solutions of the "Missionaries and Cannibals^{11} problem grounded in six representations of increasing power; and a third is Sloman^{f}s C1978D discussion of "analog^{11} representations, which *are* interpreted by exploiting the similarity between their own structure and that of the thing represented. However, there is — as yet — little systematic understanding of the power and limitations of different representations. Work in computational logic is pertinent, if it can show whether or not a certain type of representation or computational process is in principle capable of modelling a specific type of knowledge or simulating a given class of psychological process.

General results in the philosophy of science apply to Al-based psychology no less than to non-computational theories. Some such results provide for a rebuttal of common criticisms of the computational viewpoint. For instance, even were it to turn out that AI is not appropriate to the modelling of many psychological phenomena, one should not forget the Popperian point that we would still have learnt something by the enterprise. Science involves conjecture and refutation, and it is an advance to know that a specific conjecture has been empirically rejected. Nor should one forget that some tricky methodological problems apply not only to Al-based psychology but to other theories too. Thus critics of AI often remark — truly — that we cannot conclude from the fact that a computer program achieves a result in a certain way that the mind achieves it in the same way. This is a special case of the general truth that if our theory fits the facts, it may not be the only one to do so. Because of this, conclusive verification of <u>any</u> scientific theory is in principle impossible.

Work in AI concerns the nature and functioning of knowledge, and one may hope for an increasing degree of cooperation between AI and philosophical epistemologists. Traditional approaches to reasoning (whether deductive, inductive, or probabilistic) are over-idealized. They ignore epistemologically important features of intelligent inference, features that apply to all finite minds and cannot be dismissed as mere ^{M}psychologism^{M} irrelevant to normative epistemology. AI offers richer and more rigorous descriptions of the various data and procedures that comprise knowledge, and of the computational constraints that necessitate this rich variety.

Current Al-research into the logic of ^{H}non-monotonic reasoning^{11} and "truth-maintenance,^{11} for example, asks how a belief-system can be organized so as to cope with the fact that a proposition may be intelligently "proved^{11} to be true, yet turn out later to be false. Traditional logicians may wince at this description, but finite minds have to construct their knowledge under this epistemic constraint. Closely-related work on "frames" considers the ways in which single exemplars or stereotypes can be used in a flexible fashion for intelligent (though fallible) reasoning. Current discussion of "naive physics" examines the everyday (pretheoretical) understanding of concepts such as <u>cause</u>, <u>shape</u>, <u>thing</u>, <u>pathway</u>, <u>inside</u>, <u>fluid</u> ..., and should help to clarify traditional problems concerning concepts like these CHayes, 19793.

As these examples suggest, AI calls for a closer relation between epistemoLogy and empirical science than is usually thought proper by philosophers. Work on non-monotonic reasoning can correctly be described as a "logical" enquiry, and in principle could have arisen in a non-computational context. In practice, however, it is AI which has enabled us to recognize the complexity of the problems involved in formalizing everyday inference, and which has extended traditional formal approaches by offering new (computational) concepts suited to express epistemic matters. Developmental psychology (both Piagetian and non-Piagetian) has much to say on what might be called "naive physics" — as also do studies of the perceptuo-motor basis of language (such as the "psycholexicology" of Miller and Johnson-Laird C1976D). Biological and physiological considerations are relevant in view of the sensorimotor ground of our knowledge, and there is a growing recognition of the extent to which the newborn baby is already equipped with computational structures and procedures fitted to the interpretation of its life-world. Some recent work in the philosophy of mind CChurchland, 1979D similarly argues that epistemology cannot ignore our material and biological embodiment — but it suffers from a failure to consider the computational point of view.

Thus we need an interdisciplinary epistemology, in which computational insights are integrated with philosophical understanding and with psychological and biological knowledge. Indeed, the need for a genuine interdisciplinarity is a prime lesson of the computational approach. Workers in AI have much to learn from the insights of psychologists, linguists, physiologists, biologists, and philosophers, who in turn can benefit from their computationally-informed colleagues Ccf. Boden, 1981D. Mere intellectual communication across the boundaries of these several disciplines is not enough. We also need mutually cooperative research by people who (albeit specializing in one area) have a familiarity with other fields, and a commitment to their intellectual integration. This vision of "Cognitive Science" will require modification of current educational practices, so that students are no longer socially separated — and even intellectually opposed — by traditional academic labels.

Reference to education reminds us of the pragmatic, as opposed to the methodological, implications of AI. I have in mind here not primarily the many commercial and administrative applications of AI, though these will radically affect our social relations and institutions. Rather, I mean the way in which the spread of computer analogies of the mind may influence the way people think about themselves and society. As I have argued elsewhere CBoden, 1977, ch. 15D, AI is not only not dehumanizing, but is — potentially — positively rehumanizing. There are at least two senses in which this is so.

First, the view of intelligence springing from AI is active and constructive, rather than passive and defeatist like that which all too commonly informs current educational (and mental-testing) practices. For example, the Al-grounded educational approach developed by Papert (who is currently working with President Mitterand's "Computers for the People" project in Paris) deliberately fosters constructive self-criticism, so that children concentrate on the specifics of how to get better at doing something, rather than giving up in despair at their lack of "talent" [Papert, 1980]. Again, Al-based "CAI" (Computer Assisted Instruction) focusses on the pupil's active construction and exploration of the relevant domain of knowledge [Brown & Sleeman, 1982]. In this it differs significantly from the "mechanistic" approach of traditional "teaching-machines".

Second, because AI deals with representational systems, it has a conceptual base that can admit discussion of human subjectivity. This is why, as I remarked earlier, social and clinical psychology can make use of the computational approach. In general, this approach is consonant with humanistic or hermeneutic (interpretative) theories of psychology, rather than with those psychological theories, such as behaviorism, grounded in the objective natural sciences. Correlatively, hermeneutic or intentionalist philosophies of mind are closer in spirit to AI than most of their proponents believe.

This remains true even if one accepts the claim of some philosophers (e.g. Searle) mentioned above, that the processes in computer programs are not really representations, and do not really possess intentionality, but that these terms as used by the computer scientist are parasitic on their use in the human psychological context. The point is that the representational metaphor (for such it is, on this view) is one that is suited to express psychological phenomena (which alone are truly representational or intentional) precisely because it is drawn from those parts of our everyday conceptual scheme that concern these matters. For concepts to be fruitful in the theory and methodology of an empirical psychology, it is not required that they be interpreted as literal descriptions of the phenomena, just as we need not see the atom as literally a solar system in order to benefit from the notion of "planetary" electrons. So, whether computer programs specify representational processes or merely "representational" ones, they are conceptually close to hermeneutic forms of psychology rather than to those forms that ignore subjectivity.

Educational projects within society at large are needed to alert people to these facts. For most people associate computers with relatively stupid "brute force" programs (such as those used to calculate gas-bills), and think of them as machines and therefore as "mechanistic". They fail to realize that computational machines are radically different from non-computational machines, and that they are not "mechanistic" in the sense which implies a denial of subjectivity. The mistaken, though widespread, assumption that AI models of man are mechanistic in this sense may make people experience a threat to -- or even an undermining of -- their personal autonomy and moral responsibility. Behaviorism in psychology and the philosophy of mind has been often, and justly, criticized for its underestimation or denial of these psychological characteristics. But the computational approach, if properly understood, is not open to such criticisms. To realize this is to disarm the computational bogeyman.

oooooOOOooooo

REFERENCES

AMAREL, S. (1968) "On Representations of Problems of Reasoning about Actions," in D. Michie (ed.), Machine Intelligence 3 (Edinburgh Univ. Press), pp. 131-172.

Mass.: Harvard University Press).

BODEN, M. A. (1977) *Artificial Intelligence and Natural Man* (New York: Basic Books).

BODEN, M. A. (1981). "The Case for a Cognitive Biology," in M. A. Boden, *Minds and Mechanisms: Philosophical Psychology and Computational Models* (Ithaca, N.Y.: Cornell Univ. Press), pp. 89-112.

BROWN, J. S., & D. SLEEMAN, eds. (1982) *Intelligent Tutoring Systems* (New York: Academic Press).

CHURCHLAND, P. M. (1979). *Scientific Realism and the Plasticity of Mind* (Cambridge Univ. Press).

DENNETT, D. C. (1979) "Intentional Systems," in D. C. Dennett, *Brainstorms* (Cambridge, Mass.: Bradford Books), pp. 3-22.

DREYFUS, H. L. (1972) *What Computers Can't Do: A Critique of Artificial Reason* (New York, Harper & Row).

FODOR, J. A. (1981) *Representations* (Cambridge, Mass.: Bradford Books).

GAULD, A., & J. SHOTTER (1977) *Human Action and Its Psychological Investigation* (London: Routledge, Kegan Paul).

HINTON, G. E. (1981) "Shape Representation in Parallel Systems," *Seventh Int. Joint Conf. Artificial Intelligence*, (Vancouver), 1088-1096.

HAYES, P. J. (1979). "The Naive Physics Manifesto," in D. Michie (ed.), *Expert Systems in the Micro-Electronic Age* (Edinburgh Univ. Press), pp. 463-502.

INHELDER, B., & A. KARMILOFF-SMITH (1975) "If You Want to Get Ahead, Get a Theory," *Cognition*, 3, 195-212.

KARMILOFF-SMITH, A. (1979) "Micro- and Macro-Developmental Changes in Language Acquisition and Other Representational Systems," *Cognitive Science*, 3, 81-118.

MARR, D. (1982) *Vision* (Cambridge, Mass.: MIT Press)

MILLER, G. A., & P. N. JOHNSON-LAIRD (1976) *Language and Perception* (Cambridge, Mass.: Belknap Press).

NEWELL, A., & H. A. SIMON (1972) *Human Problem Solving* (Englewood Cliffs, N.J.: Prentice-Hall).

PAPERT, S. (1980) *Mindstorms: Children, Computers, Powerful Ideas* (New York: Basic Books).

SEARLE. J. R. (1980) "Minds, Brains, and Programs (with peer commentaries)", *The Behavioral and Brain Sciences*, 3, 417-457.

SLOMAN, A. (1978) "Intuition and Analogical Reasoning," in A.Sloman, *The Computer Revolution in Philosophy* (Atlantic Heights, N.J.: Humanities Press), pp. 144-176