REVIEW

Stephen R. Graubard (ed.), *The Artificial Intelligence Debate. False Starts, Real Foundations*. Cambridge Mass., Londen: The MIT Press, 1988.

This volume, previously published as a 'Special Issue' of the journal Daedalus, contains fourteen papers on Artificial Intelligence. Its authors come from such diverse disciplines as computer science, mathematics, biology, sociology and philosophy. In order of appearance of their article in the book, they are: Seymour Papert, Hubert L. Dreyfus & Stuart E. Dreyfus, Robert Sokolowski, Pamela McCorduck, Jack D. Cowan & David H. Sharp, Jacop T. Schwartz, George N. Reeke & Gerald M. Edelman, W. Daniel Hillis, David L. Waltz, Anya Hurlbert & Tomaso Poggio, Sherry Turkle, Hilary Putnam, Daniel C. Dennett and finally John McCarthy.

In its infant days, most AI researchers thought that AI could be easily achieved. This optimism was based on the belief that the problems any intelligence - biological or artificial - would have to cope with, constrained the design of the problem solver so narrowly that only a few design-possibilities would remain open. Moreover, it was thought that the design could be developed by relying on only a few general mechanisms. Thus the faith in some kind of 'General Problem Solver', governed by e.g. search, logic or production rules. It was also taken for granted that intelligence was a matter of software, and not of hardware. This led to the further assumption that any general-purpose computer - in practice a serial von Neumann machine - could be programmed to embody intelligence. The fact that human intelligence is realized in brains, was considered as accidental.

Nowadays, this view of AI is on its retreat. It is widely criticized, from in- and outside the AI community. Most of the authors in this book share this sceptical attitude. It turned out that designing programs by exploiting constraints arising from the nature of the problems they should solve, did not lead to general, but instead to specific programs. Such programs worked only within the limited domain they were designed for, and even within this domain, they lacked plasticity. This means they weren't capable, for example, of handling any situation that differed only slightly from those that were anticipated by the programmer. In almost all of the papers in this volume, a form of this complaint is to be found. Thus, it is said of traditional AI programs, that they are 'ad hoc', that they lack 'context-sensitivity' or 'common sense' (see e.g. the articles by Dreyfus & Dreyfus, Reeke & Edelman, Waltz and Putnam). Another line of

criticism concerns traditional AI's neglect of biological evidence and its reliance on the von Neumann machine. Thus it is argued that programs that would solve the same problems as humans do in the same amount of time, can never be implemented in a von Neumann machine. Because such type of computer is serial, it is argued, it is bound to be far too slow. Most of the authors prefer some kind of parallelism, which is faster because it allows many computations to be carried out at once.

One potential reaction to such criticism is to try to change the tradition from within. Such an approach is taken by John McCarthy, who - being acutely aware of the problems facing earlier logic-based systems - advocates the implementation of new kinds of logic, e.g. non-monotonic logics (a non-monotonic logic is one in which the addition of new axioms can block conclusions that could be made before the addition).

Another reaction to the aforementioned criticism, is to break with the tradition and to seek inspiration elsewhere. This, of course, is what is done by those engaged in currently in vogue (Artificial) Neural Networks (ANN's) research, or connectionism. It should come as no surprise that all papers in this volume (except those by McCarthy and Sokolowski) discuss, in more or less detail, connectionism.

Connectionists are influenced by biology, but even more by (statistical) physics. Among the features of their models they esteem important and interesting are:
- massive parallelism, which enables networks to operate within real (biological) time constraints
- implicit and adaptable storage of knowledge, which implies a certain plasticity and context-sensitivity.

Connectionist models are reviewed in the article by Cowan & Sharp. In this, and other papers, the authors are more or less optimist about the potentials of ANN's to solve the problems that faced traditional AI programs (e.g. the papers by Dreyfus & Dreyfus, Waltz, Hillis and Turkle).

In other contributions, however, connectionism itself is the object of criticism. Ironically, ANN's are found guilty of some of the same sins connectionists found traditional AI guilty of. It is objected, for example, that connectionists' 'neurons' are simplistic compared to real neurons. It is also argued that the mechanisms that operate in ANN's (e.g. relaxation or backward error propagation) are too slow to work in biological time. Connectionists are also criticized for looking for a few general, not well-understood, mechanisms to solve all problems of intelligence, while nature is bound to have offered a wide variety of solutions to the different cognitive challenges any organism

faces. Developments of such criticisms can be found in the articles by Papert, Schwartz, Reeke & Edelman and Hurlbert & Poggio.

Some authors also show a more general scepticism towards AI, be it of the traditional or connectionist kind. J.H. Schwartz, for example, estimates the total memory capacity of the brain to be about 4×10^{16} bytes, while technology will be able to achieve only one percent of this capacity in foreseeable time. In other papers - not by coincidence by philosophers - the well-known 'in principle' obstacles to formalization, and hence computerization, appear. They are (among others): common sense, desire and induction (see the papers by Dreyfus & Dreyfus, Sokolowski and Putnam).

If the discussion of connectionism is what unifies this book, this does not mean it is the sole issue. Various articles raise important questions and offer interesting perspectives. Schwartz, for example, considers the question of the potential contribution of computer scientists to theoretical neurobiology. Dennett asks whether philosophy can learn anything from AI. Reeke & Edelman offer a synopsis of their theory of Neuronal Group Selection, while Hurlbert & Poggio offer a sketch of Marr-style vision research. In Sherry Turkle's paper, the potential impact of connectionist and society of mind-models on a revival of psychoanalysis is at issue. Many other themes are developed in this volume.

This could lead one to conclude that the book is rich in content but also that it is a bit of a mixed bag. A way to anticipate and prevent the latter conclusion could have been to include an introduction. Alas, this hasn't been done. To make matters worse, the book also lacks an index.

Erik Myin
Research Assistant NFSR
Vrije Universiteit Brussel
Belgium