ELSEVIER

Journal of Cognitive Systems Research 2 (2001) 291–295

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Book review

Action editor: Stefan Wermter

The Mechanization of the Mind: On the Origins of Cognitive Science, Jean-Pierre Dupuy, translated by M.B. DeBevoise, Princeton University Press, 2000, $29.95 / £19.95, 240 pp. ISBN: 0-691-02574-6

The idea that cognitive agents process information by virtue of neural computations performed in their brains is, deservedly or otherwise, a central tenet of contemporary cognitive science. As central perhaps as the idea of natural selection is to evolutionary biology. Yet, while every biologist knows, at least in some mythical version, the story of how Darwin and Wallace introduced the concept of evolution by common descent by means of natural selection, the number of researchers in cognitive science and artificial intelligence for whom a paper published in 1943 by McCulloch and Pitts is but an early neural network model and not one of the foundational events of their disciplines can be worryingly large. Similar things could be said, risking only a slight exaggeration, about the relation between the concept of feedback and purposeful behaviour, about the first uses of terms such as 'information' and 'complexity', and in general about the very idea that the mind can be understood by the scientific study of the *formal* properties of brain mechanisms.

In this book, Dupuy explores the initial stages of the intellectual history of cognitive science. He does so with dexterity, drawing on primary and secondary sources and delivering a vision of the original cybernetics group, its achievements and its shortcomings, which is brilliantly balanced. Indeed, Dupuy proves himself particularly apt to handle the interdisciplinary aspect of this history as he is able to swiftly guide the reader through the subtleties of both continental and analytic philosophy, the nature of modelling and scientific explanation, the physics of dissipative processes, aspects of neurophysiology and embryology, and more, all without once trivializing his subject or falling victim to the use of extremely technical jargon. The book differs from similar projects (e.g., Heims, 1980, 1991) in its focus on the origin and early development of ideas that led to current cognitive science in particular.

Dupuy retraces the intellectual history of cybernetics as beginning with wartime research on feedback systems and code cracking. The incipient movement was reflected in seminal events, notably in publications by McCulloch & Pitts (1943), and Rosenblueth et al. (1943), as well as von Neumann's collaboration in the construction of ENIAC (also in 1943)[1]. But it would be after the war (from 1946 to 1953) that what was to be called the cybernetics group would acquire its character through a series of ten interdisciplinary meetings held in New York and sponsored by the Josiah Macy Jr. Foundation. The name of these meetings would evolve from ''Feedback Mechanisms and Circular Causality in Biological and Social Systems'' to simply ''Cybernetics'' following Heinz von Foerster's proposal to adopt the title of Norbert Wiener's 1948 book (Wiener, 1961). These meetings gathered researchers from mathematics, engineering, neurophysiology, psychology, social sciences and philosophy. Their most salient members included Norbert Wiener,

[1] That same year saw the publication of K. J. W. Craik's *The Nature of Explanation* (Craik, 1943), in which the concept of 'internal models' was introduced as a response to the prevailing behaviouristic approach in psychology. In view of the subsequent development of cybernetic and computational ideas of the mind one must believe it was indeed an auspicious year for the sciences of cognition.

1389-0417/01/$ – see front matter
PII: S1389-0417(01)00037-7

Warren McCulloch, John von Neumann, Claude Shannon, Kurt Lewin, Margaret Mead and Gregory Bateson, the first three being central characters in the book.

Despite unavoidable internal misunderstandings one thing became clear during these meetings: the research goal of the group, at least as championed by Wiener and McCulloch, was the naturalisation of the mind by means of the scientific study of formal models of cognitive mechanisms. As such, cybernetics is a direct ancestor of both mainstream cognitive science and AI. Ideas from cybernetics would influence many other disciplines such as operations research, communication engineering, and control theory, but cognitive science and AI would most notoriously, if not outspokenly, carry forward its flag in the decades to follow. There is also a less well known offspring of cybernetics, made rather inconspicuous by its history of limited funding, which Dupuy calls the second, or second-order, cybernetics. Members of this movement included Heinz von Foerster, W. Ross Ashby, W. Grey Walter and Gordon Pask. Second-order cybernetics and manifestations of cybernetics beyond the American scene (such as the Ratio Club in Britain) get mentioned only briefly in the book and are beyond the bounds of the main argument which is to establish the original cybernetics group as the starting point of modern cognitive science. It is here that historical ironies begin to crop up. While second-order cybernetics departed from the original in essential ways, it never lost contact with the initial movement and saw itself as its natural development. In contrast, cognitive science and AI, whose ruptures with cybernetics, Dupuy argues, were much less significant from the conceptual and methodological perspectives, often acted as if cybernetics had never existed or should be forgotten.

The book is dedicated to fleshing out these ironies; a project which is interesting in itself but at the same time also relevant for contemporary research because many of the arguments offered against the early cybernetics cannot today be said to have been fully addressed by mainstream cognitive science. These include the issues of autonomous activity and large-scale processes in the nervous system, the blind trust on, and possible misperception of, the significance of Church-Turing thesis, and the constitution of the identity of a cognizer out of subjectless processes.

Dupuy convincingly debunks various myths about the early cybernetics movement. For instance, cybernetics was *not* perceived by its members to be a new science that would extend the methods of physics to a novel class of systems in order to accommodate mental phenomena. This perception, however, is not uncommon today in those who associate early cybernetics with subsequent developments such as von Bertalanffy's General Systems Theory (von Bertalanffy, 1968). Quite on the contrary, the cybernetic project was often seen as the 'colonisation' of the mental and the social by contemporary physics and mathematics, thus provoking defensive reactions among many of its members and straight rejection by many researchers in neurophysiology and Gestalt psychology who would regard cybernetics at best as a simplistic endeavour.

Cybernetics' central tenet, Dupuy argues, was that the mind was a mechanism – there was no reason why mental activity could not be explained as a manifestation of the laws of physics. Yet, this open metaphysical stance was accompanied by a less overt postulate: if the mind was a machine, it must be a *logical* machine. Springs and levers were ruled out as mental, unless they could be shown to functionally instantiate some logical circuitry. The backdrop for this belief was the logical revolution of the 1930s with the theorems by Gödel and Turing, and especially the Church-Turing thesis. Cyberneticians would make perhaps too much of this thesis, sometimes even presenting it as a proven fact so that they would happily convince their audiences that any behaviour that could be described by a finite number of rules could be instantiated by a logical computing machine.

What gave particular impetus to the cybernetic project, given this context, was the idealisation by McCulloch and Pitts of neural circuits in the brain which demonstrates the possible equivalence in principle between neural networks and Turing machines. Under this idealisation, individual neurons act as threshold devices that, when arranged into circuits, can be used to compute logical functions. As a theory of brain function, it presented an atomistic, digital and logical view. The functional elements in the brain were individual cells, and their function was that of on/off logic gates. All other physiological details were just that: details that could slightly affect computations, but not change the essence of neural function. Shannon and Wiener's information theories (Shannon, 1948; Wiener, 1961) would nicely complement this view, opening vast programmatic possibilities which cybernetic research would exploit.

This neuron doctrine found little support from outside the cybernetics group and less than full support from within. Together with the primary sources proper to the cybernetic meetings, Dupuy also explores events in which cybernetic ideas had to step outside the inner circle into more mainstream neurophysiology and psychology. The logical and atomistic view of the brain/mind did not catch on so easily in these events. One notable example was the Hixon Symposium in 1948 (Jeffress, 1951). Its participants included McCulloch and von Neumann on the cybernetic front, neurophysiologists Karl Lashley and Ralph Gerard (also a participant at some of the Macy conferences), Gestalt psychologist Wolfgang Köhler and embryologist Paul Weiss.

Neurophysiologists expressed their worry that the McCulloch and Pitts model was perhaps too simplistic. Their qualms involved what they saw as the runaway mathematisation of neurophysiological data with little or no 'return' to real nervous systems as the natural place to test hypotheses. Gerard was particularly defensive against what he considered as an empirically unjustified atomistic view of neural function. He advocated a more continuist perspective, where modulation by chemical and hormonal fields could also be seen as central processes of brain function (Wiener's idea of a digital circuit modulated by analogical variables was intended to find a middle ground between McCulloch and Gerard).

The autonomous (and massive) activity of neuronal circuitries was what cybernetics was ignoring, according to Lashley. Real neurons cannot be said to be inactive or at rest most of the time 'waiting' for the activity wave that the sensory input is supposed to induce. On the contrary, evidence shows that real brains are constantly exhibiting various dynamic patterns of activity, which are *modulated* by the sensory stimulus, but not determined by it. Weiss would support this view. The fundamental property of nervous systems, according to him, was their autonomy and internal coherence; stimuli could trigger or release different modes of organised patterns in brain activity, but the different possibilities were determined only by the brain's structure. A picture notably opposed to the associationist view where output structure correlates with input structure and which seemed almost axiomatic for McCulloch (or indeed for most computational perspectives on brain mechanisms). Similar worries have been expressed anew in recent years (Varela et al., 1991).

It was not mathematisation that worried Gestaltists like Köhler; both he and his disciples, notably Kurt Lewin, also an early Macy participant, thought psychology could follow the model of field physics. Logical atomism they found less palatable. McCulloch and Pitts applied their model to the problem of perceiving universal forms, such as squares. Köhler's field theory (he had already presented his preliminary experiments on the search for a neurophysiological counterpart of the figural aftereffects at the fourth Macy conference) was diametrically opposed. He looked for structural similarities between the objects of perceptions and brain currents (as elicited by the coherent activity of electrical fields where individual neural activity was not a major player). Though his theory would prove to be a dead end, many of the criticisms that he and other Gestaltists levelled against cybernetics, such as its inability to account for the nature of colour perception or figural aftereffects, proved to be justified.

The irony exemplified by the Hixon encounter is that the perspectives that were presented as opposing the cybernetic position were those of autonomy, hierarchical self-organisation, and circular causation, perspectives that today we tend to associate *with* the cybernetic credo, perhaps because of confusion with the second phase of cybernetics, or perhaps because of unwitting historical distortions promoted by its offspring disciplines.

To my mind, the most revealing confrontation of ideas came from within cybernetics itself, in what Dupuy calls the "Ashby case". W. Ross Ashby presented a couple of papers at the ninth Macy conference in 1952. One of them was on his famous Homeostat (see Ashby, 1960). This machine would be able to adapt to perturbations by the effect of random internal reconfiguration. The mechanism of plastic change would operate for as long as the system's essential variables were out of bounds. Ashby in this way tried to explain in mechanistic terms adaptation in real organisms who must conserve certain physiological variables within permitted limits in order to survive. He saw adaptation as the regaining of homeostatic stability whenever it was challenged during interactions with the environment. *Random* internal reconfiguration was not intended as a faithful model of analogous mechanisms of internal plasticity in real organisms, but it strongly proved the point that, even in this extreme case in which mechanisms are dumb and directionless, adaptation still occurred and the system behaved in a teleological manner when viewed as a whole in relation to its environment. The system could even be trained to respond in desired ways by a regime of punishments.

This result shocked other cyberneticians who could not be convinced that there was no trick behind the Homeostat's design. So ingrained was apparently the idea that brain mechanisms involved some sort of complex computation, that it was not easy for them to accept that anything but the *appearance* of purposeful behaviour could be produced by such a 'mindless' machine. In this rejection, early cybernetics revealed its true metaphysical colours. It was not just that the brain/mind was mechanistic, but in addition those mechanisms should be somehow 'mindful' in themselves, and this meant choosing mechanistic explanations that were in a certain sense isomorphic with the cognitive phenomenon to be explained, a practice inherited by AI and cognitive science, and which remained unchallenged until very recently by dynamical perspectives on cognition (e.g., Port and van Gelder, 1995).

Ashby put a mirror on early cybernetics by showing that the possibility of a true naturalisation of intentionality could have a much more radical interpretation. Intelligence need not be embodied *in* mechanisms. In fact, to think so is to fall into a category error, like thinking that 'speed' or 'all terrain performance' can be found by lifting a car's bonnet and examining its interior. In contrast, intelligence, like speed, is better seen as the outcome of the actions of a whole entity in relation to its medium. Internal mechanisms must obviously be adequate to support intelligent and goal-driven behaviour, but they need not be clever or intentional in themselves. In ''March 1952, Ross Ashby left the paleo-cyberneticians dumbfounded and defenseless. In the history of cybernetics and, by extension, of cognitive science, their encounter marked the end of one era and the beginning of another'' (p. 155). Dupuy sees this event as the origin of second-order cybernetics, which was indeed a new phase in cybernetic thought, but I think it would go against Dupuy's own arguments to suggest that Ashby's lesson on the non-necessity of isomorphism was fully taken in by cognitive science or AI.

Throughout the book, Dupuy's discussions on the difficulties of interdisciplinarity provide a secondary thematic thread. The book abounds in examples of terminological and conceptual misunderstandings, both within and outside the cybernetics group such as, for instance, the resilient and always damaging confusion between information in Shannon's or Wiener's sense and information in the everyday sense of meaningful content.

Dupuy also highlights a series of missed opportunities for the cybernetic project. For instance, he argues that continental phenomenology would have provided a better philosophical stance to inform many of the assumptions that went into cybernetic research. He also discusses its failed rendezvous with what was to become theoretical biology, in particular molecular biology and what we now know as biotechnology. These fields make use of many cybernetic concepts for instance in the understanding of genetic regulatory networks. This potential was already perceived in the 1960s by biologist C. H. Waddington who organised a series of workshops on theoretical biology in which one can find pieces of work that would not have been out of place in a cybernetics meeting.

Overall, Dupuy succeeds in challenging quite a few preconceptions about early cybernetics. Even those readers who regard this era with admiration and acknowledge their intellectual debt to this pioneering group may find a few surprises. Does this make this book a revisionist project? It would probably be fair to say so if cognitive science had their own historical version of what cybernetics was, rather than just a myth. It is, I believe, more accurate to consider this book in an archaeological vein, for the reader is often exposed to first hand accounts and direct primary sources which are not commonly found in textbooks or university syllabuses. Although the book achieves its purposes, its main shortcoming is that, despite extensive discussion of the broader context of philosophy of mind the reader is left with only a piece of the story. The opening question of why was cybernetics such a poorly loved parent is not given a fully satisfactory answer, as the book does not quite enter into the early stages of symbolic AI and cognitivism and their surface break with the early cybernetic tradition. It is here that we should look for a better perspective on many of the ironies in our misperceptions of early cybernetics. But, by putting a magnifying glass on a crucial period in the intellectual history of the last century, Dupuy has opened up a field for further discussion and hopefully better understanding of where cognitive science stands and where it is going.

Ezequiel A. Di Paolo
School of Cognitive and Computing Sciences
University of Sussex
Brighton, BN1 9QH
UK

References

Ashby, W. R. (1960). *Design for a brain: The origin of adaptive behaviour* (Second edition). London: Chapman and Hall.

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

Heims, S. J. (1980). *John von Neumann and Norbert Wiener: The mathematics and technologies of life and death*. Cambridge, MA: MIT Press.

Heims, S. J. (1991). *Constructing a social science for postwar America: The cybernetics group*. Cambridge, MA: MIT Press.

Jeffress, L. A. (Ed.). (1951). *Cerebral mechanisms in behavior: The Hixon symposium*. London: Wiley.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*, 115–133.

Port, R. & van Gelder, T. (Eds.). (1995). *Mind as motion: Explorations in the dynamics of cognition*. Cambridge, MA: MIT Press.

Rosenblueth, A., Wiener, N., & Bigelow, J. (1943). Behavior, purpose, and teleology. *Philosophy of Science*, *10*, 18–24.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal*, 27, 379–423, 623–656.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT Press.

von Bertalanffy, L. (1968). *General systems theory: Foundations, development, applications*. New York: Braziller.

Wiener, N. (1961). *Cybernetics, or control and communication in the animal and the machine*. (Second edition). Cambridge, MA: MIT Press.