Available online at www.sciencedirect.com
ScienceDirect
Cognitive Systems Research 7 (2006) 372–374
Cognitive Systems RESEARCH
www.elsevier.com/locate/cogsys

Book review

Review of Curious Emotions, by Ralph Ellis; John Benjamins Publishing, 2005 (Advances in Consciousness Research Series)

Action editor: Stefan Wermter

Christopher Willmot

Centre for Hybrid Intelligent Systems, School of Computing and Technology, University of Sunderland, St. Peters Way, Sunderland SR6 0DD, UK

Available online 8 February 2006

At first glance, this does not look like the kind of book which would help cognitive science researchers, as Ralph Ellis is Professor of Philosophy at Clark Atlanta University and has written mainly on consciousness, affect, motivation and thinking. But Ellis' long-term interest in the philosophy of mind makes his observations refreshingly different. He emphasises a viewpoint and approach that few have explored, making some specific suggestions which might prove fruitful research areas.

Most cognitive science research (Arkin, 1998; Arleo, Smeraldi, & Gerstner, 2004; Weber, Wermter, & Zochios, 2004) takes a reactive view, where data in the environment triggers an agent to respond in some way. Ellis turns this on its head. For him it is the agent which is always active within a relatively passive environment. The agent is no *tabula rasa* but already has purposes of its own. Citing evidence from neuroscience and psychology, he presents a "biological organism" constructed as a self-organising system.

In speaking of "biological organisms" there is no assumption that artificial systems could not manifest such structures, thus in principle that such systems could not have conscious minds. The determining factor is not whether the system is composed of certain specific elements such as carbon or silicon, but rather the structural dynamics of the system. To be a biological organism, in this sense, is to be structured in a certain self-maintaining and self-organising pattern. (p. 198)

This review brushes lightly over the philosophical issues, focussing on that "self-organising pattern". In particular, the book presents a radically different role for emotions and their coordinating role in such a system.

Most people, most of the time, do not think much. Rather, when faced with some situation, they *feel* their way to a solution. The hard work of reasoning is only turned to when their first, emotion-based, response fails to produce something useful. By giving emotional "reasoning" a central place in the agent, Ellis opens the door for the development of a human-like intelligence that is unreachable using most current approaches.

Much of Chapter 1 is spent debunking the common fallacy that emotions are responses to stimuli. For Ellis, emotions are the system's way of signalling that something is "off" – the current state has deviated from the desired state in some way. This signal occurs in a self-organising system that is already engaged in some on-going activity. It therefore has a context and a direction. But what the emotion is "about" is only discovered through an exploratory process somewhat similar to active vision. What we perceive is as much a function of what we look for and expect to find, as of what is actually present.

The introduction is well worth reading for Ellis' background assumptions, which differ from the starting point of many cognitive scientists. The conclusion is less useful in this regard, focusing upon philosophical issues. Indeed, for a cognitive scientist the early chapters are the most pertinent.

Chapter 1 provides the underpinning for the rest, describing what emotions are and their role in perception

E-mail address: chris.willmot@sunderland.ac.uk.
URLs: http://his.sunderland.ac.uk, http://osiris.sunderland.ac.uk/~cs0cwi/.

1389-0417/$ - see front matter
doi:10.1016/j.cogsys.2005.12.002

and rationality. This is fleshed out in Chapter 2 which reviews the psychological and physiological evidence Ellis calls upon to support his thesis. Chapter 3 introduces a fundamental concept he calls "extropy", although it is mostly concerned with psychological issues. The material most directly relevant to cognitive science research is in Chapter 4. Relying heavily on the observations of a neuroscientist (Panksepp, 1998) and the role of action imagery in perceiving affordances (Jeannerod, 1994) a bounded yet flexible description of motivation emerges. Many treatments of this subject are either nebulous or unconvincing. But there is enough psychological detail here to suggest how a motivational system could be constructed that is more biologically plausible than a few "drives" represented as numbers.

The reader interested in intelligent systems might be tempted to miss out Chapters 5 and 6 on "Varieties of extended self and personality" and "Learning about emotion through the arts". However, this would be a mistake. These chapters include gems like Ellis' account of the neuroscientific mechanism for free will (p. 147) or how a machine built on enactive principles would be capable of love (p. 181). Chapter 7 is also worth a look as it deals with the emotional brain as an enactive entity and has some suggestions concerning how a self-organising system must be structured to exhibit agency.

There are three reasons why I consider this book to be worth reading. First, it presents a viable alternative to the dominant reactive paradigm. Secondly, Ellis is at pains to stay firmly within the scientific process. And finally, he offers a human centred perspective.

Both perception and action are fundamentally enactive for Ellis: carried out by agents already acting for some purpose. Perception occurs because the agent is seeking something, not because data in the environment generates some response. On this view learning involves the whole agent and contrasts sharply with machine learning which has more in common with habituation or conditioning. A fundamental aspect of human learning is that there is always a choice whether, and quite how, it should be applied. This book shows how these choices would be possible for a self-organising dynamical system. Such decision making, which avoids determinism on the one hand or randomness on the other, is similar to the autonomous agents of Kauffman (2000).

For Ellis, all psychological and biological processes must ultimately be explainable in terms of physics and chemistry. However, he argues vehemently against the reductionist tendency which holds that all phenomena are explainable at that lowest level. A self-organising system built from physical and chemical processes can genuinely cause behaviour, i.e. it cannot be reduced to a causal chain at a lower level. It works like this.

Actions are performed by a variety of inter-dependent subsystems (e.g., homeostatic control of energy or novelty). These are selected and invoked by a self-organising dynamical system (the agent). Each subsystem affects the background conditions needed for other subsystems, but may be realized by a variety of alternative mechanisms (e.g., sensing someone by hearing *or* seeing).

So the self-organising system incorporates a selection of subsystems as it continually tries to achieve equilibrium. Each subsystem uses whatever mechanisms are currently available within its repertoire. The agent's behaviour cannot be reduced to the mechanisms as no single one provides a necessary or sufficient causal support for another (although some mechanism always does). The causal chain occurs at the subsystem level as changes in one alters the background conditions for others. However, this configuration of subsystems only exists because the agent set it up in the first place. Therefore the self-organising system can exhibit genuine, irreducible, agency.

Unfortunately, Ellis gives no details of what the subsystems might be or how they might work, appealing vaguely to a combination of interlinked homeostatic feedback loops and the fact that all causal links are only valid given certain assumptions (background conditions determined by the system). This seems a recipe for chaos, which he limits by requiring that the system as a whole must (somehow) have a tendency to stability.

Ellis offers a human-centred perspective. The categories of thought people have, shape what they are able to think about. Physicists see things from a mathematical perspective and so prefer AI implimentations that have theoretical derivation. Someone primarily valuing biology or neuroscience will see AI solutions in those terms. If we are to develop machines with human-like qualities, it would be useful to develop categories of thought concerning how people behave – what human-ness is like. Without this viewpoint, AI is impoverished. Psychologists have been studying people scientifically for a century and we would benefit from listening to what they have to tell us.

There is a telling example in the introduction. A philosopher, when asked about his feelings for his wife, described the constant conjunction between things she did and certain emotions he had, concluding that there was a statistical probability that one was caused by the other. Ellis reminds us that such an abnormal inability to read ones own feelings has a clinical name, alexithymia. Following this stimulus–response paradigm cannot result in anything better than alexithymic machines because of the categories of thought used in their design.

On the other hand, using insights from psychology alongside methods derived from biology and mathematics will help us develop more human-like intelligence. Interacting homeostatic systems can account for some semi-cyclic behaviours, although Ellis reminds us that realistic systems would have a tendency to find the lowest energy configuration. But human behaviour revels in complexity. For example, why do we enjoy watching sport? Or how can we account for the pleasure experienced watching a tragic play? Ellis suggests that "extropy", a preference for higher

energy levels, is enough to counteract the natural trend towards stagnation. It is this seeking of complexity for its own sake that is behind the book's title "Curious" Emotions. For him there are three basic motivations: homeostasis (which includes consummatory drive reduction), a preference for higher energy configurations (extropy) and a need to preserve the systems boundaries.

This book presents a model for an agent based upon a "*self-organising system that can appropriate, organise and replace its own micro-constituents*". Emotions coordinate the subsystems; aiming to maintain homeostasis, protect boundaries and seek complexity. Ellis offers a theory of how emotions mediate high-level behaviours such as curiosity and play. While suggesting a physiological mechanism for consciousness, he tackles the various philosophical problems raised by this issue. Although high level and somewhat sketchy, the book proposes a design underpinned by scientific evidence including some testable hypotheses that follow from the enactive view. I would recommend this book as it shows an alternative way of seeing how intelligent systems could be built that are biologically inspired and exhibit human-like behaviour.

References

Arkin, R.C. (1998). *Behaviour-based robotics*. Bradford: MIT Press.

Arleo, A., Smeraldi, F., & Gerstner, W. (2004). Cognitive navigation based on nonuniform garbor space sampling, unsupervised growing networks and reinforcement learning. *IEEE Transactions on Neural Networks, 15*(3), 639–653.

Jeannerod, M. (1994). The representing brain: neural correlates of motor intention and imagery. *Behavioral and Brain Sciences, 17*(2), 187–245.

Kauffman, S. (2000). *Investigations*. New York: Oxford University Press.

Panksepp, J. (1998). *Affective neuroscience*. New York: Oxford University Press.

Weber, C., Wermter, S., & Zochios, A. (2004). Robot docking with neural vision and reinforcement. *Knowledge Based Systems, 12*(2–4), 165–172.