Computational Emergence and Computational Emotion

D.N. Davis

Neural, Emergent and Agent Technologies Group, Department of Computer Science,

The University of Hull, Kingston-upon-Hull, HU6 7RX, England

D.N.Davis@dcs.hull.ac.uk

ABSTRACT

This paper presents an analysis of the issues pertaining to computational emergence and emotion in (cognitive) agent systems whether their ecology is carbon or silicon-based. We consider how a developing computational theory of cognition can be used to monitor and manage interactions with and within complex systems and so harness unwanted and emergent states and behaviours before the computational system becomes dysfunctional. A comparison is made between these emergent computational states and emotional states. The opening scenario provides a global systems context for the interaction between man and the cyber-domain. The final focus of this paper weaves these threads together in a consideration of the nature of goals, emotion and philosophical issues related to (computational) modelling.

1. Introduction

The best way to learn from your mistakes is to survive them : W. Vollman, The Rifles :256 [1]

If age is an evolving matrix within the sphere of existence, then consider the ramifications of a near exponential rise in the technological sub-matrix that we can refer to as the cyber-domain. What is the nature of this domain? Who is using the technology of this domain, and to what ends? Furthermore, are there certain categories of repeated failures from which we can learn how to design and develop more robust complex systems?

Simons, sometime ago [2], suggested that we would see a convergence of technologies. Today, this view seems, at least superficially, correct with the increasing domination of the IT market by certain international and multinational players. With the increasing uptake in specific computational software platforms by larger businesses, the smaller players refocus their products to align with these platforms. Again with the increasingly standardisation of communication across heterogeneous platforms, a similar convergence seems apparent. While it is unclear whether this is problematic in itself, what is of more immediate concern is that the design philosophies underpinning these computational platforms that provide the focus of this convergence are, in themselves, seriously flawed.

There are further issues. If we consider trade and finance, we can note the recurrence of technology accelerated global breakdowns such as the events surrounding Black Thursday and the UK's withdrawal from the European Economic Community’s common fiscal policy. Small perturbations in the Tokyo markets diversify through the Chinese and Malaysian markets to Europe and the Americas in a 24 hour-a-day global market. Is this really chaotic, and out of control? Is it an emergent property of the type discussed by Simons and Kelly [3].

In medicine, we seen increasing homogeneity within the technological frameworks that underpin the health systems within any one nation, and perhaps across societies of nations (e.g. the North American Trade Alliance, the European Economic Community). In culture, the popularisation of the World Wide Web and Internet for entertainment and education relies on a common metaphor. In short the global technological cybernetic mesh is becoming homogenised.

Sage at SMC98 [4] made a number of points about large systems, and in particular that the structure and behaviour of these complex systems is not dictated uniquely by the edicts of a leader, but emerges in a natural manner through the interactions of the agents with these systems. If this holds true then the uses of these converging cybernetic systems will, because of the different socio-cultural perspectives and individual needs of the users, lead to a diversification in the way these systems actually run. This, of course, assumes that the design philosophy underpinning these systems does not dictate otherwise. If this is not the case, then the agents interacting with these systems will become either disillusioned, and abandon these systems, disenfranchised, and forgo their socio-cultural perspectives and individual needs, or look to other systems, and so create a new cybernetic niche space. There are, of course, a plethora of other and alternative outcomes. The ensuing questions from an analysis of this are manifold:

- Are human beings becoming disenfranchised in the decision-making loops and paths with this increase in (not necessarily intelligent) autonomy? A particularly catastrophic example of this is the Chernobyl incident wherein a complex control system was built in such a way that the overriding of the automatic system produced an unwanted emergent property.

- Is the social-cultural diversification of the Homo sapien centred planet being replaced with the blandishment of Machine sapien?

- In what ways can this socio-cultural matrix of homo-silicon interaction be changed, and to what ends should these changes be directed?

There is also a growing concern of the technological adept about the increasing technocracy and their sometimes over simplistic generalisations and lack of deep understanding of what can happen to and in complex dynamic systems. Dynamism here is not only mere concrete changes - for example computers on the move and in control of aircraft, or the increasingly frequent hardware upgrades, but also the acceleration in data volume and transfer in this silicon based ecology.

2. Emergence

Computational emergence can be categorised [4,5] across four main types. Diachronic emergence, which describes computational states that emerge over time, typically measured in terms of evolutionary periods. The benefits from using evolutionary algorithms to design problems; the resulting optimised designs are a diachronic result emerging from the use of such systems. Gestalt emergence which describes the processes whereby recognisable patterns of processing emerge, much like the interpretations observers place on the patterns displayed by Conway's Game of Life. Representational emergence with which representational schemes related to behaviour in at least one layer in a multi-layer computational system emerges at other layers. Representational emergence describes, for example, the results of the chunking mechanism in (symbolic) computational cognitive architectures such as SOAR [6]. Functional emergence describes situations whereby system capabilities emerge through design of supporting mechanisms.

There are moreover (at least) three categories of problems associated with those designers who rely on and describe emergence: sub-cognitive bias, behavioural bias, and individualistic bias. Of the different types of sub-cognitive bias, the current paper suggests that a form of sub-cognitive bias (emotional bias) is likely to occur in the scenarios described above because of both behavioural bias and individual bias. Emotional bias can be defined as "*the likelihood of responding with one kind of emotion more than another*" [7]. In a homogeneous computational framework, to achieve specific goals certain very specific patterns of behaviour are required of all users. However, because of the emotional response of individuals to any specific pattern of behaviour, the emerging interaction may not necessarily be fruitful unless the computational framework has the capability to respond to the user's emotional state.

Irrespective of the tasks modelled and whether they are internal to a complex system or the human interaction with such systems, perturbant computational behaviour analogous to human emotive states can occur. This type of problematic emergent behaviour does not easily fit with the four categories already introduced

3. Perspectives on Emotions

Cognitive science is a young, developing and changing area of academic interest. Like most such areas, it is prone to trends of favoured research areas. Recently dynamic systems metaphors have surfaced as a useful analogy for what may occurring in those processes we label as cognition. We have already demonstrated it [8] as a supportive design and methodological platform for providing a coherent approach to areas such as motivation, planning and attention. Here we conjecture that a more appropriate emotion engaging perspective on goals can only be provided through a consideration of the niche spaces and emergent properties of emotional behaviour.

The control state approach to mind was popularised by Simon [9], and has seen resurgence with the work of Sloman and others [10]. While it only provides an incomplete description of cognition (and related behaviours), it does provide a useful starting point for considering the cognitive processes required of an agent that has to demonstrate believable emotional behaviour. What this approach attempts to do is provide an ontological decomposition of the categories of cognitive activity a human mind may perform in attempting to achieve, set or develop long-term motivational goals. While it may try to relate these to other types of psychological states, it needs to be stated that the approach is based on a representational stance. From this perspective, emotions are affective internal states analogous to Attitudes. *Attitudes* are predispositions to respond or act (either internally or externally) to specific (perceptual or cognitive) cues and can involve intricate collections of beliefs, motivators etc..

We contest that reflexive, reactive, deliberative and reflective behaviours [11] are required in agents that ultimately aim to mirror the depth and breadth of human mental activity. Not all of these activities can be modelled using approaches based on the representational stance.

Intelligent behaviour is the result of interaction with the world in the achievement of long-term niche space goals, and can be generated without the explicit reasoning (and representation) that symbolic AI proposes. Furthermore, intelligence is an emergent property of complex systems, and therefore in the eye of the beholder, not some innate, isolated property; i.e.. the description of intelligence is, to some greater or lesser extent, dependent upon the type of biases described above. This is true, whether that intelligence is emotion or rationale based.

Emotion can be described as "*a state usually caused by an event of importance to the subject. It typically includes (a) a conscious mental state with a recognisable quality of feeling and directed towards some object, (b) a bodily perturbation of some kind, (c) recognisable expressions of the face, tone of voice, and gesture (d) a readiness for certain kinds of action*" [7]. Hence, emotions, in socio-biological agents, are affective mental, conative and-or cognitive, states and processes. Conative states differ from cognitive states in being non-symbolic dynamic processes not readily represented using the control state space approach to mind. These dynamic sub-cortical processes can be hormonally and chemically based; they do not however map easily into the representations used in symbolic reasoning. This category of processes do, however, need to interface with the types of categorisations that can be modelled using cognitive metaphors.

In fact, some form of continuum can be readily devised (or construed) for many of these processes as a n-space matrix with one axis labelled as conative to cognitive. We could state that any particular emotional interaction could be sited on some level between a hot conative state to a cold cognitive state. Further valences can be used to provide a range of emotive states, for example from love to hate, whether categorical (hate-loathe-like-love) or a more continuous label (for example normalised as a real number between 0 and 1). A further axis can relate to personal states of danger, from example from insecure to secure, sadness - happiness, trust - doubt - distrust. There are thousands more words and phrases that could be used to amplify this space, and many ways of trying to group these. At any one time, the emotional state of any socio-biological agent could then be represented, however simply, as some vector in this n-space. The development of this as a computational metaphor within our dynamic cognitive architecture may be problematic, but useful as it may provide a more readily accessible environment for people to make more effective uses of computational cyberspace.

4. Goals and Cognition

Goleman [12] and others suggest that the rational-based intelligence quotient provides a too narrow description of truly intelligent systems, and this orthodoxy needs amplifying with an emotional quotient. Most artificial intelligent systems are based on rational cognitive models. Here we consider how this perspective on the nature of intelligence fits with the concept of goals.

The term goal can mean a number of things, depending upon (social, linguistic) context and the perspective offered by any particular chosen academic discipline. Goals are typically related to problem solving., and other related forms of cognitive behaviour. From a predominantly psychological perspective, we can consider two definitions.

Duncker [13] describes a problem as arising when a problem solver ‘has a goal but does not know how this goal is to be reached’. From the Gestalt perspective (of which Duncker was part) the pattern of internal behaviour required to achieve the goal is unclear. However, of interest, is that the categories of *cognitive behaviour* (sic) with which the School of Gestalt Psychology were attempting to describe all problem solving ability map onto two of the four categories of emergent behaviour introduced above; specifically representational and gestalt emergence.

Adams [14] describes problem solving in terms of frameworks that sometimes require re-representation This is analogous to Newell and Simon’s [15] description of problem solving as a search through problem space, which can be described in terms of states and paths between states, where the solution of a problem is defined as the goal state. Typically the search for a path from an initial (or given) state to a goal state requires guidance, or the use of a set of criteria or cost functions. Newell and Simon describe this in terms of means-end analysis. Newell develops this further elsewhere [6]

From these and other analyses, we can provide a five-point perspective on problem solving as:

- the transformation of an initial state into the goal state where there is no obvious path or method of achieving a solution; otherwise the *problem* (sic) is trivial;
- a cognitive behaviour and therefore an internal, not directly observable, behaviour. There are, however, psychological and knowledge engineering techniques, imperfect as such, that enable such processes to be externalised;
- a process involving the manipulation of knowledge (typically referred to as symbolic, but need not necessarily be so [16]);
- a directed process; i.e.. there is an *intention* to produce a solution;
- a personal experience, dependent upon existing knowledge and its development.

Aspects of this model of problem solving can be seen in many artificial intelligence models of problem-solving [17] and in the application of techniques such as Protocol Analysis to knowledge acquisition. Problems can be categorised [18] in a number of ways as

- Routine problems, with well defined paths from given to goal state;
- Well described problems, where a clear representation is possible;
- Inducing structure problems where patterns are generalised from examples;
- Transformation problems in state space, for example the Towers of Hanoi;
- Arrangement problems, requiring the recombination of mainly pre-existing problems and partial solutions [19];
- Ill defined problems, where the problem is described (if at all) in quite vague terms.

What may be of interest, in terms of some future research is the investigation and modelling of human and machine responses to these categories of problems, particualraly with regard to success and failure. It remains unclear whether a general pattern of response will emerge, and whether any such pattern would be emotional and/or rationale. There would seem to be a continuum (perhaps in something other than a linear space) from the trivial (non-problem) problem to the most ill defined of problems, where the cognitive behaviour necessary to further the given state (if it can be described as such) is contingent upon creativity. Creativity is of course usually, but not always, related to emotional states. Where representation is possible, problems can in effect be said to be understood [20], if not necessarily solved, or even currently solvable, and may be represented in terms of some arrangement of sub-goals.

Related to these perspectives of the representation of problems are different modalities of cognitive behaviour. From the Gestalt and similar subsequent perspectives, problem solving involves a reorganisation of the problem, to provide structural insight. This is effectively a creative (emergent) process that depends largely on reproductive thinking (the application of existing thought patterns) and productive thinking (the generation of a novel thought pattern), with successive stages of reformulation involving increasing more specific (though not necessarily symbolic) representations. This model provides an explanation of conceptual habituation, where the accumulation of past experience can produce rigidity of thought patterns (in a similar way to over-training in an artificial neural net classifier), resulting in routinisation, and an inability to produce novel thought patterns required of creative problem solving. This routinisation is analogous to the possibilities hinted at earlier in describing the increasing homogeneity of global cybersystems.

5. Computational Framework

This developing, but admittedly quite sketchy, theory of affective goals and computational emergence will map onto a distributed multi-layer computational model which makes use of tightly and loosely coupled agents working within a holistic system. This computational metaphor is analogous to those used in describing blackboard systems, group-ware and the dynamic systems approach to cognition. The computational modelling of such an agent system would require several types of functionality and knowledge handling, We contest that this is beyond the type of control systems usually associated with human interaction, and more appropriate to the sophistication associated with distributed agent (nee blackboard) systems. This kind of approach also coincides with the dynamic systems approach to cognition [21]. From this perspective, cognition is viewed as the changing focus of a number of interacting (sometimes extant, sometimes dormant) processes. Any particular reasoning or problem-solving capability can be tuned to a number of related, but different tasks, and that the achievement of complex tasks requires the use of many qualitatively different capabilities.

6. Discussion

This paper has considered how existing cybernetic systems may in effect be disenfranchising their human creators. It is suggested that by considering goal-based processing that takes account of emotional involvement, a more harmonious path may result. In effect, we are suggesting that by performing a means-end analysis on existing cybersystems, we have come to question the value judgements being made upon the synergy that is socio-human-computer interactions. One concern is whether any change in track will make a difference, given that such complex systems are, in the final analysis, necessarily emergent.

If cybernetic systems are being used to solve problems, most current metaphors and descriptions of problem solving involve the idea of directed thinking which subsumes reasoning. De Bono [22], for example, rails against the rigidity of conceptual delineation. Others have raised similar arguments against representation and the issue of intentionality from Bretano through Quine to Popper [23]. Consider Popper’s perspective on worlds 1, 2 and 3. World-1 is the physical world - the computer hardware and its devices. World-2 is the world of conscious experiences - the interaction of the users and the computers. World-3 is logical content of books, libraries, computer memories and such like i.e. theories, concepts etc.. While our theories of emergence may fit within this gestalt, true emergence does not. Similarly if we consider problem solving from a connectionist perspective (the computational gestalts) and the behaviour-based fields, we see a reliance on emergent properties. As this and other papers highlight, we need further research and analysis on the nature of emergence to understand how to design and develop more effective systems, irrespective of the underlying computational stance of their creators.

From a very different and ancient perspective [24], goals are merely a psychological illusion seen through the perception of introspection. As such the pursuit of goal-based interaction is equivalent to chasing a mirage. However, they still have a philosophical reality, and some cognitive efficacy.

While this paper provides an incomplete analysis of the cognitive environment appropriate from an engaging and meaningful interaction within the socio-cultural-homo-machine-ecology, it is suggested that such like analyses ought to be considered when developing complex systems, whether they are to be used for entertainment, education, business or medicine.

7. References

[1] W. Vollman, The Rifles, André Deutsch Ltd, 1994.

[2] G. Simons, Silicon Psychosis: Derangement in the Global Network, Harvester Wheatsheaf, 1989.

[3] K. Kelly, Out of Control: The New Biology of Machines, Fourth Estate Ltd, 1994.

[4] A.P. Sage, Agents and Complexity Theory

in the Engineering and Management of Large Systems, IEEE International Symposium on Systems, Man and Cybernetics, San Diego, USA, October 1999.

[5] N. Gilbert and R. Conte, Artificial Societies: The computer simulation of social life, UCL Press, 1995.

[6] A. Newell. Unified Theories of Cognition. Harvard University Press, 1990.

[7] K. Oakley & J.M. Jenkins, Understandng Emotions, Blackwell Publishers, 1996.

[8] D.N. Davis, Synthetic Agents: Synthetic Minds?, Frontiers in Cognitive Agents, IEEE International Conference on Systems, Man, and Cybernetics, IEEE Press, 1998.

[9] H.A. Simon, Models of Thought, Yale University Press, 1979.

[10] A.Sloman. Motives mechanisms and emotions. Emotion and Cognition, 1(3):217--234, 1987. Reprinted in M.A. Boden (ed), The Philosophy of Artificial Intelligence, `Oxford Readings in Philosophy' Series, Oxford University Press, 231--247, 1990.

[11] D.N. Davis, Reactive and motivational agents. In: Intelligent Agents III - Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages, J.P. Muller and M.J. Wooldridge and N.R. Jennings (Eds.), Springer-Verlag, Heidelberg, 1996.

[12] D.P. Goleman, Emotional Intelligence, Bloomsbury, 1995.

[13] K. Duncker, On problem solving. Pyschological Monographs, 58(5), 1945.

[14] J.L. Adams, Conceptual Blockbusting, W.H. Freeman, 1972.

[15] A. Newell and H.A. Simon, Human Problem Solving. Prentice-Hall, 1972.

[16] S.M. Kosslyn, Information representation in visual images, Cognitive Pyscholology, 7, 341-370, 1975.

[17] J.G. Greeno, Nature of problem solving abilities. In: W.K. Estes (Ed). Handbook of Learning and Cognitive Processes. (Vol. 5), Erlbaum, 1978.

[18] M.W. Eysenck, Blackwell Dictionary of Cognitive Psychology, Blackwell Publishers, Oxford, 1990.

[19] B. Hayes-Roth, Intelligent control. Artificial Intelligence, 59:213—220, 1993.

[20] G. Polya. How to Solve It. Princeton University Press, 1945.

[21] R.F. Port and T. Van Gelder, Mind As Motion : Explorations in the Dynamics of Cognition, MIT Press, 1995.

[22] E. de Bono, Water Logic: The Alternative to I am Right You are Wrong. Viking Press, 1993.

[23] K. Popper, Conjectures and Refutations, Routledge and Kegan Paul, 1974.

[24] L. Tzu, Tao Te Ching, (Approx, 400B.C.) (Trans: D.C. Lau), Penguin, 1963.