EMOTIONAL STATES AND REALISTIC AGENT BEHAVIOUR

Matthias Scheutz
School of Computer Science
The University of Birmingham
Birmingham B15 2TT, UK
E-mail: mxs@cs.bham.ac.uk

Aaron Sloman
School of Computer Science
The University of Birmingham
Birmingham B15 2TT, UK
E-mail: axs@cs.bham.ac.uk

Brian Logan
School of Computer Science Science and Information Technology
University of Nottingham
Nottingham NG8 1BB, UK
E-mail: bsl@cs.nott.ac.uk

KEYWORDS
Affect and emotions, believable agents, reactive versus deliberative architectures, *SimAgent* toolkit

ABSTRACT

In this paper we discuss some of the relations between cognition and emotion as exemplified by a particular type of agent architecture, the *CogAff* agent architecture. We outline a strategy for analysing cognitive and emotional states of agents along with the processes they can support, which effectively views cognitive and emotional states as architecture-dependent. We demonstrate this architecture-based research strategy with an example of a simulated multi-agent environment, where agents with different architectures have to compete for survival and show that simple affective states can be surprisingly effective in agent control under certain conditions. We conclude by proposing that such investigations will not only help us improve computer entertainments, but that explorations of alternative architectures in the context of computer games may also lead to important new insights in the attempt to understand natural intelligence and evolutionary trajectories.

INTRODUCTION

In both artificial intelligence and the design of computer games, the study of emotions is assuming a central role. Building on pioneering early work (Simon 1967; Sloman 1981), it is now widely accepted in the artificial intelligence community that cognition (including intelligence) cannot be understood completely if emotions are left out of the picture. At the same time, the designers of computer games and entertainments have come to realise that emotions or at least mechanisms associated with them are important in the creation of convincing or believable characters.

However, to exploit emotion effectively game designers need to understand the differences between purely cosmetic emotional implementations and deeper interactions between cognitive and affective behaviour. In this paper, we outline a strategy for analysing the properties of different agent architectures, the cognitive and affective states and the processes they can support. We illustrate our argument with a scenario demonstrating the surprising effectiveness of simple affective states in agent control, in certain contexts.

EMOTIONS AND INTELLIGENCE

Minsky (1987) writes in *The Society of Mind*: "The question is not whether intelligent machines can have emotions, but whether machines can be intelligent without any emotions." Like many others (e.g. Damasio 1994; Picard 1997) he claims that higher levels of (human-like) intelligence are not achievable without emotions. Unfortunately the concept "emotion" is understood in so many different ways by different people that this is not a well-defined claim. Moreover, some of the evidence purported to establish a link between emotions and higher forms of intelligence shows only that rapid, skillful decisions, rather than analytical deliberations, are sometimes required for intelligence. As argued in (Sloman 1999a) it does not follow that there is any *requirement* for such episodes to involve emotions, even though emotions are *sometimes* involved in rapid skillful decisions.

The definitional morass can be separated from substantive scientific and technical questions by a strategy which involves exploring a variety of information processing architectures for various sorts of agents. The idea is to use agent architectures to (1) study families of concepts supported by each type of architecture and (2) explore the functional design tradeoffs between different architectures in various contexts. This will help game designers understand the difference between purely cosmetic emotional implementations (e.g. using "emotional" facial expressions or utterances) and deeper interactions between cognitive and affective mechanisms that are characteristic of humans and other animals, where the visible manifestations arise out of processes that are important for the well-being or survival of the individual, or some group to which it belongs.

Some of these are relatively simple, e.g. "alarm" mechanisms in simple *reactive* architectures, which interrupt and override "normal" processing (e.g., being startled by an unexpected noise or movement would be an example of a purely reactive emotion in humans). Other cases are more subtle, e.g. where the use of explicit affective states such as desires or preferences to select behaviours can achieve more flexibility than direct coupling between stimulus and response, for instance allowing both new ways of detecting the presence of needs and new ways of satisfying the same needs in different contexts to be learnt. More sophisticated emotions involving awareness of "what might happen", (e.g. anxiety) or "what could have happened or could have been avoided" (e.g. regret or shame), require more sophisticated *deliberative* architectures with extended representational capabilities.

Affective states involving evaluation of one's own internal processes, e.g. the quality of problem solving or the worthiness of desires, need a still more sophisticated reflective architecture with a *meta-management* layer (Beaudoin 1994). If the operations of that layer can be disrupted by highly "insistent" motives, or memories or concerns, then typically human types of emotional states may emerge out of the ensuing interactions. For instance, infatuation with a member of the opposite sex, embarrassment, excited anticipation, conflicts between desire and duty, can all interfere with attempts to focus on important tasks, because in these states high level processes are interrupted and diverted. These are characteristic of the emotions portrayed in novels and plays. The underlying processes will need to be understood if synthetic characters displaying such emotions in computer entertainments are ever to become as convincing as human actors.

Understanding the complex interplay of cognition and emotion in all these different sorts of cases requires close analysis of the properties of different architectures and the states and processes they can support.

EXPLORING ARCHITECTURES FOR COGNITION AND AFFECT

The "cognition and affect project" at the University of Birmingham is a long term project to study different kinds of architectures and their properties in order to understand the interplay of emotions (and other affective states and processes) and cognition. It addresses questions such as how many different classes of emotions there are, how different sorts of emotions arise (e.g., which ones require specific mechanisms and which ones are emergent properties of interactions between mechanisms with other functions), how emotions fit into agent architectures, how the required architectures can be implemented, what role emotions play in the processing of information, where they are useful, where they are detrimental, and how they affect social interaction and communication. A better understanding of these issues is necessary for a deep and comprehensive survey of types of agents, the architectures they require, their capabilities, and their potential applications (Logan 1998).

As part of the project, a particular type of agent architecture, the *CogAff* architecture (Beaudoin 1994; Wright 1996; Sloman and Logan 1999; Sloman 1998) has been developed, which divides the agent's cognitive system into three interacting layers (depicted in Figure 1) corresponding to the three types of mechanisms mentioned in the previous section. These are a *reactive*, a *deliberative*, and a *meta-management* layer, all concurrently active, all receiving appropriate sensory input using perceptual mechanisms processing information at different levels of abstraction (as illustrated in Figure 2) and all able to generate action. Each layer serves a particular purpose in the overall architecture, but layers can also influence one another.

The reactive layer implements basic behaviours and reflexes that directly control the agent's effectors and thus the agent's behaviour, using no mechanisms for representing possible but non-existent states. This layer can generate chains of internal state-changes as well as external behaviours. In animals it can include chemical mechanisms, analog circuits, neural nets, and condition-action rules. Different sub-mechanisms may all run in parallel performing dedicated tasks. Sometimes they may be activated sequentially. There is no construction of complex descriptions of hypothesised situations or possible plans, though the system may include pre-stored plans whose execution is triggered by internally or externally sensed events.

The deliberative layer is a first crucial abstraction over the reactive layer in that it is concerned with the processing of "what-if" hypotheticals involved in planning, predicting or explaining past occurrences. Deliberative mechanisms can vary in complexity and sophistication. A full-fledged deliberative layer will comprise at the very least compositional representational capacities, an associative store of re-usable generalisations, as well as a re-usable working memory for constructing proposed plans, conjectures or explanations, which can be evaluated, compared, adopted, or rejected.

The third layer is concerned with self-observation and self-reflection of the agent and provides the possibility for the agent to observe and evaluate aspects of its internal states, and perhaps to control some of them, e.g. by directing attention to specific topics. However, since processes in other layers can sometimes interfere, such control is not total.

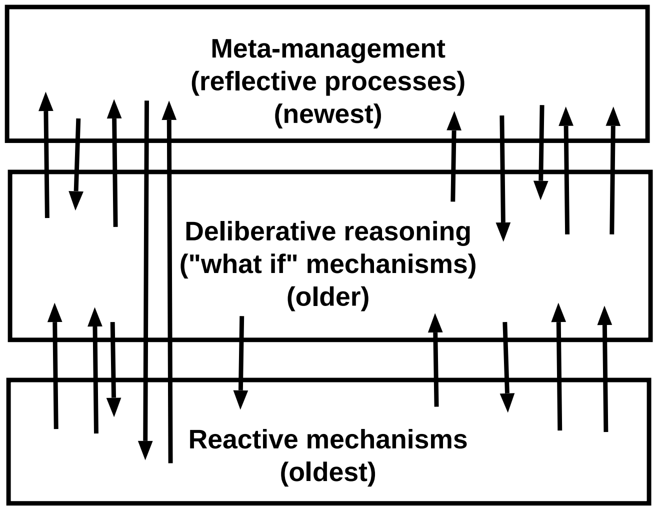

Figure 1: The three layers

We conjecture that these three layers represent major transitions in biological evolution. Although most details of the evolutionary trajectories that can produce such multi-layered systems are unknown it is possible that many of the architectural changes will turn out to be examples of the common process of "duplication and divergence" (Maynard Smith and Szathmàry 1999).

This model may be contrasted with other kinds of layered models, e.g. where information enters the lowest layer, flows up some abstraction hierarchy, causes decision-making at the top, after which commands flow down via successive expansion processes to the lowest layer which sends signals to motors. The *CogAff* model also differs from layered hierarchies where higher layers totally dominate lower layers, e.g. some subsumption models. Notice, moreover, that although the functions of the top two layers are different from those of the reactive layer, they will need to be implemented in reactive mechanisms, much as abstract virtual machines in computers are implemented in low level digital mechanisms performing very different operations.

Besides clearly distinguishing conceptually different capabilities of agents, and among other advantages, this tripartite division of cognitive systems also provides a useful framework for the study of a wide range of emotions. For example, it turns out that it nicely parallels the division of human emotions into three different classes (Sloman 2000a; Sloman 2000b):

- primary emotions (such as "fear" triggered by sensory inputs (e.g. LeDoux 1996) are triggered within reactive mechanisms and influence evolutionarily old responses.
- secondary emotions (such as "shame" or "relief" at what did not happen) are triggered in the deliberative layer and may produce a combination of cognitive and physiological processes.
- tertiary emotions (such as "adoration" or "humiliation", where control of attention is reduced or lost) involve the metamanagement layer, though they may be initiated elsewhere.

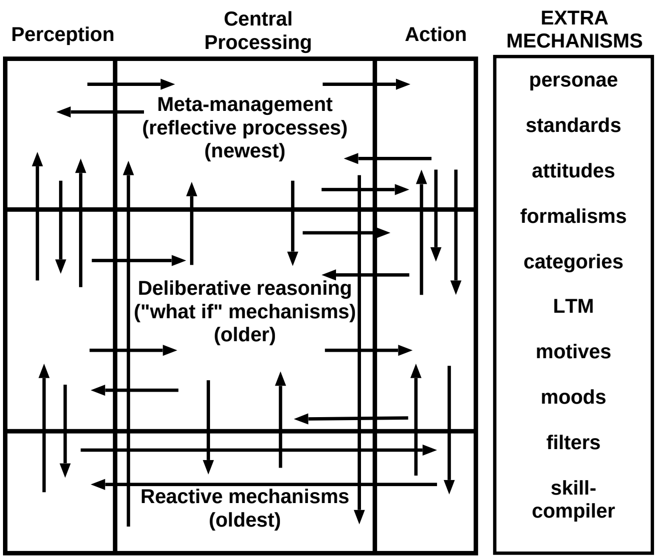

Figure 2: A more detailed version of the *CogAff* architecture

Further analysis of the *CogAff* architecture is likely to reveal finer distinctions that are relevant to understanding human interactions and also the design of "believable" synthetic agents. For instance, it provides scope for many different kinds of learning and development, involving different sorts of changes in different parts of the architecture. Figure 2 indicates impressionistically how different layers in perceptual and motor sub-systems might evolve or develop to serve the needs of different layers in the central system. The figure, however, omits "alarm" mechanisms and many other details described in papers in the Cognition and Affect project directory (see the references in the acknowledgement section).

On-going investigations in our project also include much simpler architectures that might have occurred earlier in evolution and might be found in insects and other animals. These kinds of architectures and their capabilities will be of particular interest to the games industry in the near future. Moreover, as we learn to design and make effective use of more sophisticated architectures they can take advantage of the increased computer power that will become available.

In the following, we will briefly describe one current track of the "cognition and affect project", which studies in particular the role of simple affective states in agent control. This is especially relevant for game programming, where the efficiency of the control mechanism of an agent is of the utmost importance. We will show that the usual trade-off between the behavioural capabilities of an agent and the allocated computational resources, where more effective behaviour requires more complex computations, can sometimes be overcome by using a different overall design. We demonstrate that in a certain class of environments adding simple types of "affective" states (perhaps even describable as "emotional" in some sense) to the agent control, which results only in a minimal increase in processing time, can produce complex and useful changes in the agent's behaviour.

It is worth mentioning at this point that all our implementations use the specially developed, freely available *SimAgent* Toolkit, which is open with respect to ontologies and various other agent control paradigms (Sloman and Logan 1999). The toolkit provides support for rapid prototyping and has been used in the development of a range of projects from agents in military simulations (Baxter 1996, Baxter and Hepplewhite 1999) to various student projects.

AFFECTIVE STATES AS MECHANISMS FOR AGENT CONTROL

Agents with states are more powerful than agents without states. In general, states in an agent without the third architectural layer are often used to keep track of states of affairs external to the agent and not only the agent's internal states (i.e., information about both the environment and the agent's internal states that is directly available to the agent and often times most relevant to its proper functioning, or, in some contexts, survival).

In other words, states in agents include records of current percepts and contents of the deliberative layer functioning as representational vehicles, which in turn enable reasoning, planning, etc. about the "external world", including possible futures in that world.

Usually, the additional machinery used for such reasoning and planning is quite time and resource intensive and thus requires significant additional computation power to be of any use to the agent. The generative potential of deliberative capabilities often opens up realms that are inaccessible to reactive agents (unless they have vast memories with pre-computed strategies for all possible eventualities), thus justifying this additional computational cost. However, there are cases where the same (if not better) results can be achieved using reactive systems augmented by simple affective states. Such trade-offs are not always obvious, and careful and detailed explorations in design space may be needed in order to find good designs to meet particular requirements.

For instance, we can compare (1) adding a type of deliberative extension to a reactive architecture with (2) adding some simple states recording current needs, along with simple behaviours triggered by those states which modify the agent's reactive behaviours. Option (2) can be loosely described as adding primitive "affective" or "emotional" states. We demonstrate that these states can have a powerful influence on an agent's ability to survive in a certain class of environments (based on Scheutz 2000) containing different kinds of agents, obstacles, predators, food sources, and the like (similar to simulated worlds in certain computer action/adventure games). Different kinds of agents have different goals, but common to all of them is the *implicit* goal of survival, i.e., to get (enough) food and to avoid getting killed (e.g., eaten by another creature). Agents have different cognitive mechanisms that control their actions, but in particular we focus on two kinds of agents, the "affective agents" (A-agents) and "deliberative" agents (D-agents). These differ from purely reactive agents (R-agents) in having extra mechanisms.

A-agents have reactive mechanisms augmented by simple "affective states", whereas D-Agents have simple planning abilities (implemented in terms of condition-action rules) in addition to the same reactive mechanisms (details of the respective architectures are given below). We conducted various experiments in *SimAgent* using different combinations of all three kinds of agents in the environment (everything else being the same). It turns out that the affective agents are more likely to survive for a given time period than the deliberative or purely reactive agents. Yet, the computational resources used by affective agents are significantly less than those required for deliberative agents in this scenario. Of course, we do not claim that agents with this simple affective mechanism will outperform agents with more sophisticated deliberative mechanisms using more computer power.

EXPERIMENTS WITH ARCHITECTURES

To illustrate how A-agents can reliably outperform both D-agents and R-agents in a certain type of world, we use a common reactive architecture for both kinds of agents based on augmented finite state machines, which run in parallel and can influence each other related to the style of Brooks' subsumption architecture (Brooks 1986).

The reactive layer consists of finite state machines that process sensor information and produce behavioural responses using a schema-based approach (in *SimAgent* these finite state machines are realized as a rule system). Essentially, they take sensor information and compute a sensor vector field for each sensor (i..e, the simulated equivalents of a sonar and a smell sensor), which then gets combined in a specific way and transformed into the agent's motor space (e.g., see Arkin 1989).

As mentioned previously, A-agents and D-agents extend R-agents in different ways. A-agents differ from R-agents in that they possess one "inner" state (a primitive type of "hunger") that can influence the way in which sensory vector fields are combined (i.e., this state alters the gain value of a perceptual schema in the transformation function mapping sensory to motor space, see Arkin 1998). Hence, the very same sensory data can get mapped onto different motor commands depending on the affective state. For example, when "hunger" is low, the gain value for hunger is negative and the agents tend to move away from food.

D-agents, on the other hand, possess an additional primitive deliberative layer, which allows them to produce a "detour plan" when their path to food is blocked (by an obstacle, predator, or any other agent). This plan will generate a sequence of motor commands, which override those given by the reactive mechanisms.

To be more precise, a D-agent uses information about the location of food, and the locations of obstacles and dangers to compute a trajectory, which avoids the obstacles to get to the food. It then starts moving to points on the trajectory. An "alarm" system can interrupt the execution if the agent comes too close to an obstacle, and trigger replanning, in which case the agent will attempt to make a more extensive detour. The same can happen again with the new plan. Once the execution of a plan is finished, the agent uses the same reactive mechanisms as other kinds of agents to move towards the food, which should now not be obstructed, unless the world has changed!

As one would expect, the differences in the architecture give rise to different behaviour of the agents: R-agents are always interested in food and go for whichever food source is nearest to them (often manoeuvring themselves into fatal situations). They can be described as "greedy".

Similarly, D-agents are also always interested in food, yet they attempt to navigate around obstacles and predators using their (limited) planning capacity though constantly driven by their "greed". Although their deliberative abilities make good use of all locally available information, this can have the consequence that the agent ends up too far from food and starves in situations where it would have been better to do nothing for a short period of time. By then the obstructing obstacles and predators might no longer be blocking the direct route to food. D-agents (like R-agents) constantly move close to danger in their attempts to get to food, and can therefore die for food which they do not yet really need.

A-agents, on the other hand, are only interested in food when their energy levels are low (i.e., they are not constantly "greedy", and seek food only when "hungry"). Then they behave like R-agents in that they chase down every food source available to them. Otherwise they tend to avoid food and thus competition for it, which reduces the likelihood of getting killed because of colliding with other competing agents or predators.

We conducted three experiments, in which R-agents, D-agents, and A-agents had to compete for survival in a given environment. Each experiment consists of four agents of one or two types, given 120 runs of 1000 simulated time-steps. In experiment 1 we studied R-agents and A-agents, in experiment 2, R-agents and D-agents, and in experiment 3, A-agents and D-agents.

	3 Rs 1 A	1 R 3As	4 As	4 Rs
A-agents	808	772	815	---
R-agents	607	559	---	681

Experiment 1

	3 Rs 1 D	1 R 3Ds	4 Ds	4 Rs
D-agents	773	715	705	---
R-agents	681	655	---	681

Experiment 2

	3 As 1 D	1 A 3Ds	4 As	4 Ds
A-agents	824	824	815	---
D-agents	726	746	---	705

Experiment 3

The above tables (Experiment 1, Experiment 2 and Experiment 3) show for each experiment and each agent the average number of time-steps that an agent of the respective kind (R, D, or A) survives in the given environment.

The experiments show that in the situations studied with the deliberative mechanisms provided, there is no need to resort to high level deliberative control in agent simulations, as reactive mechanisms plus affective states can do the same job more efficiently.

This is not to argue that deliberative architectures are never needed. Enhancing the D-agents with the A-agents' ability to use an "affective" state would, in different contexts, make the D-agents superior. In general, simple affectively augmented reactive mechanisms will not suffice where the ability to analyse and reason about past, present, or future events is required, for instance when attempting to understand why a construction overbalanced, or when selecting the shortest route across unfamiliar terrain with obstacles such as walls and ditches. Moreover, if the world were more sparsely populated with food and obstacles, waiting for hunger to grow before moving towards food would not be a good strategy. In fact, simulations show that A-agents do worse than R-agents or D-agents in such environments.

Our experiments are intended to demonstrate the importance of analysing trade-offs in particular contexts, since our intuitions are not always reliable. Furthermore they show how simple affective mechanisms can be surprisingly effective in some contexts. These experiments may also be relevant to understanding some evolutionary trajectories from reactive to deliberative organisms.

CONCLUSION

We agree with many others that for synthetic characters to be convincing and interesting they will need to have emotional and other affective capabilities. We do not, however, believe that the ambiguities of this claim are widely recognised nor that the relevant design options and tradeoffs are generally analysed in sufficient detail. To illustrate this we have presented an example showing how such analysis might be done and supported by simulation experiments. The mechanisms employed in this simulation, which add simple types of affective states to produce complex and useful behavioural changes, should be re-usable in a variety of games contexts.

This is all part of the larger "Cognition and Affect" project in which we are investigating a variety of design options and evolutionary trade-offs, and which demonstrates that the use of architecture-based concepts can help to clarify the confused terminology regarding emotions and other mental concepts (Sloman 2000c).

Finally we propose that such investigations will not only help us improve computer entertainments, but that explorations of alternative architectures in the context of computer games may lead to important new insights in the attempt to understand natural intelligence and evolutionary trajectories.

NOTES AND ACKNOWLEDGEMENTS

This research is funded by a grant from the Leverhulme Trust.

Our software tools, including the freely available *SimAgent* toolkit (previously referred to as "Sim_Agent"), are available from the Free Poplog Site:

`http://www.cs.bham.ac.uk/research/poplog/freepoplog.html`.

An overview of the *SimAgent* toolkit is available here:

`http://www.cs.bham.ac.uk/~axs/cog_affect/sim_agent.html`

Technical reports elaborating on the points made in this paper can be found in:

`http://www.cs.bham.ac.uk/research/cogaff/`

REFERENCES

Arkin, R.C. 1989. "Motor schema-based mobile robot navigation." *International Journal of Robotic Research* 8: 92–112.

Arkin, R.C. 1998. *Behavior-Based Robotics*. MIT Press, Cambridge, MA..

Baxter, J.W. 1996. "Executing Plans in a Land Battlefield Simulation". *Proceedings of the AAAI Fall symposium on Plan execution: Problems and issues*, November 1996, pp. 15–18.

Baxter, J.W. and Hepplewhite, R. 1999. "Agents in Tank Battle Simulations". *Communications of the Association of Computing Machinery*, vol 42, no 3, pp. 74–5.

Beaudoin, L.P. 1994. *Goal processing in autonomous agents*. PhD thesis, School of Computer Science, The University of Birmingham. (available at `http://www.cs.bham.ac.uk/research/cogaff/`).

Brooks, R. 1986 "A robust layered control system for a mobile robot." *IEEE Journal of Robotics and Automation*, RA-2:14–23.

Damasio, A.R. 1994. *Descartes' Error, Emotion Reason and the Human Brain*. Grosset/Putnam Books, New York.

LeDoux, J.E. 1996. The Emotional Brain. *Simon & Schuster*, New York.

Logan, B.S. 1998. "Classifying agent systems." In *Proceedings AAAI-98 Conference Workshop on Software Tools for Developing Agents*, B.S Logan and J. Baxter, eds., Menlo Park, California.

Minsky, M.L. 1987. *The Society of Mind*. William Heinemann Ltd., London.

Picard, R. 1997. *Affective Computing*. MIT Press, Cambridge, Mass, London, England.

Scheutz, M. 2000. "Surviving in a hostile multiagent environment: How simple affective states can aid in the competition for resources." In *Proceedings on the 20th Canadian Conference on Artificial Intelligence*. Springer Verlag.

Simon, H.A. 1967. "Motivational and emotional controls of cognition". Reprinted in *Models of Thought*, Yale University Press, 29–38, 1979.

Sloman, A. and Croucher, M. 1981. "Why robots will have emotions." In *Proceedings of the 7th International Joint Conference on AI*, Vancouver, 197–202.

Sloman, A. and Logan, B. 1999. "Building cognitively rich agents using the *SimAgent* toolkit." Communications of the Association of Computing Machinery, 42(3):71–77.

Sloman, A. 1999. "Review of Affective Computing by R. Picard 1997" *The AI Magazine*, 20(1):127–133.

Sloman, A. 2000a. "*How many separately evolved emotional beasties live within us?*" In *Proceedings workshop on Emotions in Humans and Artifacts*, Vienna, August (to appear).

Sloman, A. 2000b. "Architectural requirements for human-like agents both natural and artificial. (what sorts of machines can love?)." In *Human Cognition And Social Agent Technology, Advances in Consciousness Research*, Kerstin Dautenhahn, ed., John Benjamins, Amsterdam, 163–195.

Sloman, A. 2000c. "Architecture-based conceptions of mind." In *Proceedings 11th International Congress of Logic, Methodology and Philosophy of Science*, Synthese Library Series, Dordrecht, Kluwer (to appear).

Smith, M. J. and Szathmàry, E. 1999. *The Origins of Life: From the Birth of Life to the Origin of Language*. Oxford University Press, Oxford.

Wright, I.P., Sloman, A. and Beaudoin, L.P. 1996. "Towards a design-based analysis of emotional episodes." *Philosophy Psychiatry and Psychology*, 3(2):101–126.