Extended Abstract for AISB 2017

Architectures underlying cognition and affect in natural and artificial systems[1]

Aaron Sloman
School of Computer Science, University of Birmingham, UK
http://www.cs.bham.ac.uk/~axs

Abstract

This is a summary of some of the ideas in my invited talk for the Symposium on "Computational modelling of emotion: theory and applications" at AISB 2017. A deep understanding of human (or animal) minds requires a broad and deep understanding of the types of information processing functions and information processing mechanisms produced by biological evolution, and how those functions and mechanisms are combined in architectures of increasing sophistication and complexity over evolutionary trajectories leading to new species, and how various kinds of evolved potential are realised by context-sensitive mechanisms during individual development. Some aspects of individual development add context-specific detail to products of the evolutionary history, partly because evolution cannot produce pre-packaged specifications for complete information processing architectures, except for the very simplest organisms. Instead, for more complex organisms, including humans, different architectural layers develop at different times during an individual's life, partly under the influence of the genome and partly under the influence of what the individual has so far experienced, learnt, and developed. This is particularly obvious in language development in humans, but that is a special case of a general biological pattern (identified in joint work with Jackie Chappell, partly inspired by theories of Annette Karmiloff-Smith, among others). This paper complements a paper presented in the Symposium on Computing and Philosophy at AISB 2017, which develops more general ideas about evolution of information processing functions and mechanisms, partly inspired by Turing's work on morphogenesis: http://www.cs.bham.ac.uk/research/projects/cogaff/sloman-aisb17-CandP.pdf.

1 INTRODUCTION

Biological organisms differ in many ways. Members of the same species can differ according to their stage of development, according to the problems and resources (including information) encountered during their development, according to details of their genome, and details of previous development, growth, and learning opportunities and also in details of their particular environments with different threats, opportunities, resources, obstacles, competitors, helpers, current needs, and so on.

Variations across species are even greater. Over billions of years, biological evolution on this planet has produced a staggering variety of forms of life, differing in physical size, change of size during the life of individuals, life span, sensory apparatus, modes of development and motion, types of environment, modes of interaction with the environment including conspecifics and other life forms, food, prey, predators, forms of information storage, modes of reproduction, and many more. All of these differences (most of which are structural not numerical) can affect mechanisms,

internal states or processes, and externally visible forms of behaviour or expression, including affective states and processes related to motivation, goals, plans, preferences, desires, attitudes, values, hopes, ambitions, decisions, intentions, concerns, moods, and other affective states and processes.

Is this an area that is susceptible of scientific study and accurate modelling, or is there merely a hopelessly unstructured mess/tangle of special cases understood in depth by some novelists, poets, playwrights counsellors and therapists, but unfit to be the subject of scientific investigation?

A similar question might have been asked about chemistry centuries ago when alchemists were faced with a tangled mess of special cases with no means of expanding knowledge except by doing more experiments. But that situation was changed by discoveries about the atomic structure of matter, including the details summarised in the periodic table of the elements, along with advances in chemical understanding based on many experiments and applications of new ideas from quantum mechanics - producing explanations that were not possible in the framework of Newtonian mechanics. Chemical reactions could not be explained by Newton's laws of motion, but new explanatory theories emerged from information about the structure of atoms related to the facts assembled in the periodic table of physical elements, later elaborated by developments in quantum physics able to explain chemical structures and mechanisms including some that are crucial for biological evolution analysed in 1944 by Schrödinger[9].[2]

Since then, although huge gaps remain in our biological knowledge, there have been tremendous advances based on theories in physics and chemistry about possible structures and their interactions, often forming new structures essential to processes of biological reproduction, growth and development.

In contrast, much (so-called) scientific study of minds has relied on correlation-seeking experiments and the use of independently variable components of vectors to describe complexity - which would be hopelessly inadequate even for the study of complex molecules.

There is also a wide-spread assumption that all motivation needs to be thought of in terms of the relative attractions (or repulsions) of various kinds of reward (or punishment) with a common (positive or negative) utility measure. This can be compared with the ancient assumption that all physical masses seek the centre of the universe, which is hopelessly inadequate for the explanation of known physical and chemical phenomena.

Even if there are reward mechanisms that explain some motives and preferences, there is much they cannot explain. For example, if someone really enjoys doing mathematical research only because doing it produces some reward (whether chemical or psychological) then in principle he or she should be just as willing to get the reward by doing something much easier than struggling with mathematical problems - e.g. drinking some potion, or stepping into an otherwise harmless machine. But nobody who *really* enjoys doing mathematics would swap the activity for one of its side-effects. Of course, there may be such people for whom doing mathematics is not its own reward, but they still want to do it, e.g. because they enjoy the admiration it produces in others, or because it is a necessary condition for achieving some other goal, such as getting into university, or a useful aid to attracting an intelligent mate.

I believe I first encountered that refutation of popular reward-based theories of motivation in Ryle[8]. There are similar objections to widely used utility-based mathematical theories of decision making, such as theories based on "payoff matrices" (criticised in my 1978 book [12]).

I suggest that evolution frequently made use of *architecture-based* motive-generation mechanisms (ABM) that, unlike *reward-based* motivation (RBM), allow new motives to be triggered by perceived opportunities or situations without the individual having any *ulterior* reward-motive. It suffices that *ancestors* who had such mechanisms acquired useful knowledge that later brought benefits that the individuals could not have predicted, or even thought about. As a result they succeeded in life and produced offspring who were likely to share the same motive-generators. So the ABM mechanisms trigger motives that have been beneficial in one's ancestors, not motives whose achievements produce some special pleasure-juice. These can be thought of as genetically programmed *internal* reflexes comparable to genetically programmed physical protective and feeding reflexes. (For more on the ABM theory see [17].)

How evolution is able to produce such changes is one of the many questions addressed in the Turing-inspired Meta-Morphogenesis project[22].[3],

2 CAN STUDY OF MINDS MIRROR STUDY OF MATTER?

Across all the variation in forms of life, are there any common principles? One seems to be the ability to acquire and use information for purposes of control, such as generating options for consideration, selecting options, working out consequences of various options. There is also information-based control of chemical and physical processes of reproduction, development and growth.

Information is used during interaction with inert physical features of the environment and also during interaction with predators, prey, offspring and other conspecifics - which often requires information about information, e.g. using information about what something else wants or can perceive.

In many cases passive individuals are acted on by the environment, for instance when seeds are dispersed by wind, or when seasonal or daily changes in temperature or availability of light, air or water currents, or supply of nutrients or dangers are out of the control of individuals and they can at most resist, react to avoid or react to make use of (e.g. consume) contents of their environment.

In more complex cases information about threats, opportunities, resources, and obstacles can be acquired and put to use, either immediately or at a later time when a need arises. Coping with threats from other organisms, may involve purely physical avoidance or escape actions. But in some cases it requires other-directed meta-cognition: inferring intentions, knowledge, reasoning processes and choosing means of avoidance or escape accordingly.

So information of many kinds plays many different roles in living things, unlike non living but interacting physical objects and processes, such as weather features, geological features shaped by and shaping one another, including tectonic motion, earthquakes, volcanoes, floods, tornadoes, other weather patterns, seasonal changes caused by motion around the sun and tides caused by rotation of the moon around the earth.

This notion of information is much older than the notion developed by Shannon around 1948. Since Shannon, information is often discussed as if it were primarily the content of messages, with senders and receivers. But sending and receiving messages would be pointless if the message contents had no other use than to be transmitted, received and stored.

The fundamental fact about information that is often ignored in discussions of the nature of information is that it can be used in controlling what happens.

This can take many forms: in some cases information directly triggers a response, e.g. a defensive reflex such as blinking or rapid withdrawal, or an opportunity taken such as motion towards water, food, shelter or a mate, or use of a body part to acquire or consume something edible. In other cases the information can be stored for future use, e.g. information about where a resource or a danger is located, or information encoded in a genome that is used at a particular stage during during reproductive processes to control aspects of development and growth of tissues and parts of new individuals. Other forms of information in a genome can generate and control behaviours of organisms once they are functional, e.g. controlling breathing, pumping of blood, digestion, begging for food, following parents, and triggering new motives to be acted on later (ABM).

Such *uses* of information could be ignored in Shannon's famous work on information [11] because he was working for a company (Bell Telephone Company) providing information services, for whom the main problems were reliable transmission and storage, not use of information. The use was the concern of their customers.

In contrast, the novelist Jane Austen was very much concerned with ways in which her characters could not only transmit, acquire and store information, but also use it, as discussed in http://www.cs.bham.ac.uk/research/projects/cogaff/misc/austen-info.html. She frequently referred to information, not in Shannon's sense, but in the much older sense in which information is used, not merely transmitted or stored.

3 TWO MAIN VARIETIES OF INFORMATION USE

There are two fundamentally different roles that useful information can have, as Hume noted in distinguishing "is" (information about what is the case) from "ought" (information about what to do) in his argument that "ought" can never be derived from "is". This distinction was elaborated by Elizabeth Anscombe [1] as a difference in "direction of fit".

For an information user there are some information contents (which we can crudely label "desire-like information") whose role in an organism determines what should be done to the world to make the world match the information content, and other information contents (which we can crudely label "belief-like information") whose role is such that the information should be altered when there is a mismatch with how the world is. Both sorts are required for intelligent, or purposeful action, or deliberate inaction.

Moreover, in both cases there is always the possibility of an organism not being in a position to determine whether the information item does or does not match reality - e.g. whether some belief is true, or whether some desire or goal has been satisfied. This can generate a new *second order* desire-like information state, which specifies that an information gap needs to be bridged. That new state can trigger action to fill the information gap - which may either be done relatively simply (e.g. by looking, sniffing, touching, etc.) or by engaging in some sort of information-gathering research,

e.g. to find out whether food is available nearby and if so where it is.

As these examples show, there can be many processes by which combinations of belief-like and desire-like information states can generate actions to determine whether the belief-like states actually fit the world or actions to make the world fit the desire-like states. A rich theory of varieties of cognition and affect can be based on the implications of this distinction as pointed out (by Sloman, Chrisley and Scheutz) in [18] (building on ideas developed by Beaudoin[2]).

The time scales involved and the scale of action required to bridge these information gaps (finding out whether X is true, or making X true) can vary enormously according to the complexity of the information specification and the amount of effort involved in checking whether X fits the facts or making X fit the facts.

Things get even more complex if individuals can have a large and changing collection of desire-like and belief-like information states, unlike a simple thermostat which has a target temperature and a sensor providing information about the gap between the current and target states, along with a mechanism for turning on or turning off a heat generator or heat remover. It is often assumed that all desire-like information states are concerned with achievement or maximisation of some measurable reward or utility, but life is far too complex for that: organisms have many different needs at different stages of development and at different times and places, often needs that coexist and conflict, e.g. a need to approach a source of food when energy stores are low and a need to avoid detection by a dangerous predator or rival. The assumption that these needs can be compared on a common scale are as misguided as the assumption that strength of materials and fuel energy of materials can be compared on a common scale.

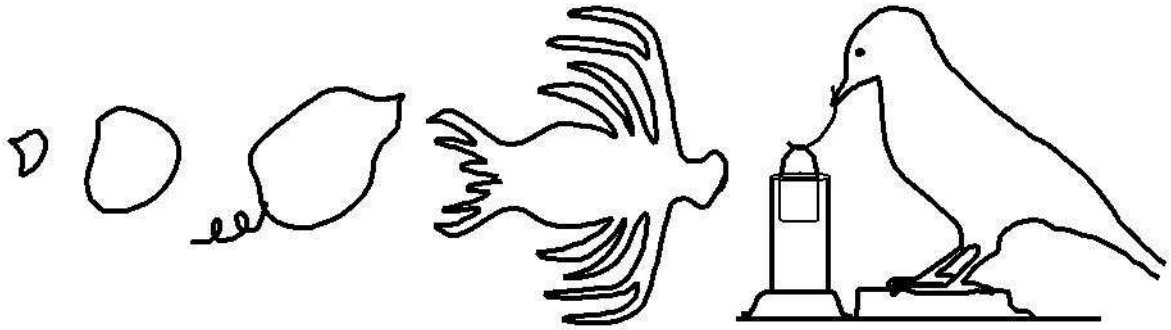

Figure 1: Many discontinuities in physical forms, behavioural capabilities, environments, types of information acquired, types of use of information and mechanisms for information-processing are still waiting to be discovered.

As organisms became more complex with more complex collections of biological needs and capabilities (crudely indicated in Figure 1) the information processing requirements, including both processing of information about what is the case (belief-like information) and information about what should be done (desire-like information) became increasingly complex, involving not only immediate choices between different possible movements, but comparisons involving various time scales and various locations in which actions can be performed.

As a result, evolution produced not only huge variations in physical forms and physical behavioural capabilities, but also huge variations in types of information acquired and used and variations in mechanisms for acquiring storing and using information - leading to further problems of control of those mechanisms - e.g. whether to think about where to get the next meal or how to avoid the

approaching predator, or where to find a mate, or what to do to improve one's information processing abilities or physical abilities of various kinds.

In the case of humans, this led to a vocabulary that referred to varieties of information state (mental state) and processes in which such states change, in addition to a vocabulary referring to varieties of physical state and physical process.

However, in the case of the physical sciences the "ordinary" vocabulary was found to be in need of fundamental expansion to cover states, processes and mechanisms that were previously unknown but provided vastly superior understanding of the physical world than our ancestors had, especially during the last few centuries.

In contrast the sciences of mind are still, to a large extent, like the ancient alchemist science, in a state that is groping towards adequate explanatory concepts and mechanisms. I do not believe that current theories are any more than a pale shadow of the theories required for deep characterisations and explanations of mental phenomena, both in humans, in other animals and in future intelligent machines.

AI has begun to change this, during the last half century or so, but we still have a long way to go, both in understanding and in solving the problems. Current proposals for information processing architectures and mechanisms are still grossly inadequate in comparison with the complexity of the phenomena to be explained. However, there are separate strands of progress in various subfields, such as vision, language, planning, finding formal mathematical proofs and various aspects of motor control. Completeness and integration seem to be a long way off. A very useful survey of recent attempts to explain affective phenomena in humans, or human like machines can be found in [5].

In my presentation I'll offer some conjectures, and evidence, relating to required forms of explanation, including required information processing architectures for explaining minds of various kinds, how they develop, and how they evolve.

4 VARIATIONS IN EPIGENETIC TRAJECTORIES

The description given so far is very abstract and allows significantly different instantiations in different species, addressing different sorts of functionality and different types of design, e.g. of physical forms, behaviours, control mechanisms, reproductive mechanisms, etc. In particular at one extreme the reproductive process may produce individuals whose genome exercises a standard pattern of control during development, leading to "adults" with only minor individual differences.

At another extreme, instead of the process of development from one stage to another being fixed in the genome, it can be created during development through the use of two or more levels of design in the genome, allowing different environments to cause different choices in going from the initial design to the adult form so that at intermediate stages not only are there different developmental trajectories due to different environmental parameters, there are also selections among the intermediate level patterns to be instantiated. For example, for the same species, in one environment development may include much learning concerned with protection from freezing, whereas in another environment individuals may vary more in the ways they seek water during dry seasons, where the differences in adults come partly from the influence of the environment in

selecting genetically available patterns to instantiate during development of individuals. E.g. one group may learn and pass on information about where the main water holes are, and in another group individuals may learn and pass on information about which plants are good sources of water (with nutrients).

All of these things may happen automatically because of patterns and meta-patterns picked up by earlier generations and instantiated in cascades during development.

But it seems that evolution has found ways of providing even richer developmental variation, by allowing the information gathered by young individuals not merely to select and use pre-stored design patterns, but to create new patterns by assembling fragments of information during early development and using newer, more abstract processes to construct new abstract patterns, partly shaped by the environment, but with the power to be used across variations in that environment.

This was called "Representational Re-description" by Karmiloff-Smith in [7]. The best known example of this is the way in which children develop (rather than learn) new languages through cooperation with conspecifics, illustrated most dramatically by Nicaraguan deaf children who produced a new sign language because their sign language teachers had had deprived childhoods because they had not learnt sign languages early enough.[4] See also [10].

Only such a mechanism with cascading alternations between data-collection and abstraction formation (by instantiating higher level previously evolved abstractions, not by forming statistical generalisations) could account for both the diversity of human languages and the power of each one, all supported by a common genome interacting with widely varying developmental environments.

I agree with Karmiloff-Smith that this process is not restricted to language development, but occurs throughout childhood (and beyond) in connection with many aspects of development of information processing. An early version of this idea, crudely depicted in Figure 2, was presented in [4], though there are many details still to be developed.

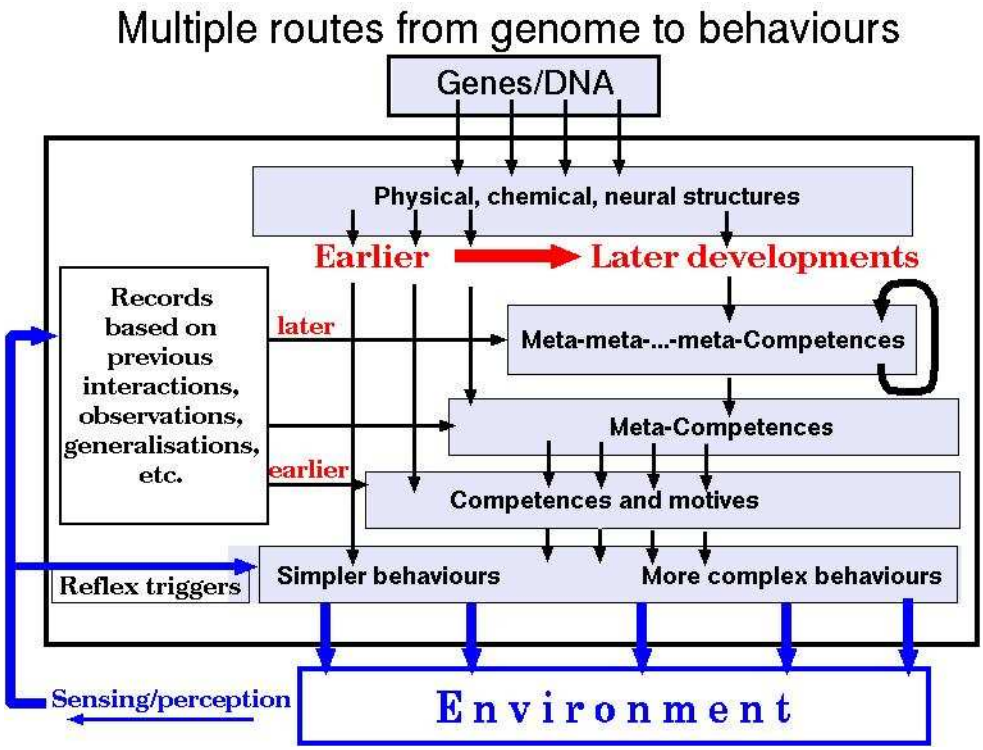

Figure 2: The varieties of developmental trajectory proposed by Chappell & Sloman. Later processes can be triggered by delayed genome products interacting with environmental information acquired at earlier stages. (Chris Miall helped with the original diagram.)

This is very different from a form of learning or development that uses a *uniform method* for repeatedly finding patterns at different levels of abstraction, e.g. using statistical generalisations.

Instead, on this model, the genome encodes increasingly abstract and powerful creative mechanisms developed at different stages of evolution, that are "awakened" (a notion also used by Kant[6]) in individuals only when their time is ready, so that they can build on what has already been learned or created in a manner that is tailored to the current environment.

5 CHANGING DEVELOPMENTAL TRAJECTORIES

As living things become more complex, increasingly varied types of information are required for increasingly varied uses. The processes of reproduction normally produce new individuals that have seriously under-developed physical structures and behavioural competences.
Self-development requires physical materials, but it also requires information about what to do with the materials, including disassembling and reassembling chemical structures at a microscopic level and using the products to assemble larger body parts, while constantly providing new materials, removing waste products and consuming energy. Some energy is stored and some is used in assembly and other processes.

The earliest organisms can acquire and use information about (i.e. sense) only internal states and processes and the immediate external environment, e.g. pressure, temperature, and presence of chemicals in the surrounding soup, with all uses of information taking the form of immediate local reactions, e.g. allowing a molecule through a membrane.

Some of the changes in types of *information*, types of *use of information* and types of *biological mechanism for processing information* have repeatedly altered the processes of evolutionary morphogenesis that produce such changes: a positive feedback process. A familiar example is the influence of mate selection on evolution in intelligent organisms, since mate selection is itself dependent on previous evolution of new cognitive mechanisms. This is a process with multiple feedback loops between new designs and new requirements (niches), as suggested in [15]. Compare also the author's presentation at the Computing and Philosophy symposium at this conference.

As Figure 1 suggests, evolution constantly produces new organisms that may or may not be larger than predecessors, but are more complex both in the types of physical action they can produce and also the types of information and types of information-processing required for selection and control of such actions.

These ideas, and those in [7] suggest that one of the effects of biological evolution was fairly recent production of extremely, but not totally, abstract construction kits that come into play at different stages in development, that produce much more rapid changes in variety and complexity of information processing across generations than ever before. This idea is fairly familiar as regards the role of a common genetic inheritance in enabling hugely varied languages to be developed by humans in different cultures. This pattern can be generalised to other aspects of development, as suggested in Figure 2. (There are loose connections with Chomsky's ideas on evolution and development of language. I don't think he ever realised that human language evolution and development must be a special case of something deeper and more general.)

There is still much work to be done regarding the space of possible information processing architectures capable of supporting diverse kinds of variety among humans and other animals. I suggest that within a century or two our ideas about how human minds work, and the requirements for modelling them in intelligent machines, will have changed at least as much as our ideas about physics and chemistry have changed since the time of Galileo. Some suggestions, regarding mechanisms and architectures can be found in [13] (compare [24]), [20], [19], [3], [14], [25], [23], [16], [21].

References

[1]
G.E.M. Anscombe, *Intention*, Blackwell, 1957.

[2]
L.P. Beaudoin, *Goal processing in autonomous agents*, Ph.D. dissertation, School of Computer Science, The University of Birmingham, Birmingham, UK, 1994.

[3]
L.P. Beaudoin and A. Sloman, 'A study of motive processing and attention', in *Prospects for Artificial Intelligence*, eds., A. Sloman, D. Hogg, G. Humphreys, D. Partridge, and A. Ramsay, 229-238, IOS Press, Amsterdam, (1993).

[4]
Jackie Chappell and Aaron Sloman, 'Natural and artificial meta-configured altricial information-processing systems', *International Journal of Unconventional Computing*, **3**(3),

211-239, (2007).

[5]
Eva Hudlicka, ‘Affective BICA: Challenges and open questions’, *Biologically Inspired Cognitive Architectures (2014)*, **7**, 98-125, (2013).

[6]
Immanuel Kant, *Critique of Pure Reason*, Macmillan, London, 1781. Translated (1929) by Norman Kemp Smith.

[7]
A Karmiloff-Smith, *Beyond Modularity: A Developmental Perspective on Cognitive Science*, MIT Press, Cambridge, MA, 1992.

[8]
G. Ryle, *The Concept of Mind*, Hutchinson, London, 1949.

[9]
Erwin Schrödinger, *What is life?*, CUP, Cambridge, 1944.

[10]
Ann Senghas, ‘Language Emergence: Clues from a New Bedouin Sign Language’, *Current Biology*, **15**(12), R463-R465, (2005).

[11]
Claude Shannon, ‘A mathematical theory of communication’, *Bell System Technical Journal*, **27**, 379-423 and 623-656, (July and October 1948).

[12]
A. Sloman, *The Computer Revolution in Philosophy*, Harvester Press (and Humanities Press), Hassocks, Sussex, 1978. http://www.cs.bham.ac.uk/research/cogaff/62-80.html#crp, Revised 2015.

[13]
A. Sloman, ‘Towards a grammar of emotions’, *New Universities Quarterly*, **36**(3), 230-238, (1982).

[14]
A. Sloman, ‘The mind as a control system’, in *Philosophy and the Cognitive Sciences*, eds., C. Hookway and D. Peterson, 69-110, Cambridge University Press, Cambridge, UK, (1993).

[15]
A. Sloman, ‘Interacting trajectories in design space and niche space: A philosopher speculates about evolution’, in *Parallel Problem Solving from Nature - PPSN VI*, ed., *et al.* M.Schoenauer, Lecture Notes in Computer Science, No 1917, pp. 3-16, Berlin, (2000). Springer-Verlag.

[16]
A. Sloman, ‘The Cognition and Affect Project: Architectures, Architecture-Schemas, And The New Science of Mind’, Technical report, School of Computer Science, University of Birmingham, Birmingham, UK, (2003). (Revised August 2008).

[17]
A. Sloman, 'Architecture-Based Motivation vs Reward-Based Motivation', *Newsletter on Philosophy and Computers*, **09**(1), 10-13, (2009).

[18]
A. Sloman, R.L. Chrisley, and M. Scheutz, 'The architectural basis of affective states and processes', in *Who Needs Emotions?: The Brain Meets the Robot*, eds., M. Arbib and J-M. Fellous, 203-244, Oxford University Press, New York, (2005). http://www.cs.bham.ac.uk/research/cogaff/03.html#200305.

[19]
A. Sloman and M. Croucher, 'Why robots will have emotions', in *Proc 7th Int. Joint Conference on AI*, pp. 197-202, Vancouver, (1981). IJCAI.

[20]
A. Sloman and M. Croucher. You don't need a soft skin to have a warm heart: Towards a computational analysis of motives and emotions, 1981. (Now available at the Birmingham CogAff site).

[21]
Aaron Sloman, 'Virtual Machine Functionalism (The only form of functionalism worth taking seriously in Philosophy of Mind and theories of Consciousness)', Research note, School of Computer Science, The University of Birmingham, (2013).

[22]
Aaron Sloman, 'Virtual machinery and evolution of mind (part 3) meta-morphogenesis: Evolution of information-processing machinery', in *Alan Turing - His Work and Impact*, eds., S. B. Cooper and J. van Leeuwen, 849-856, Elsevier, Amsterdam, (2013).

[23]
Aaron Sloman and colleagues, 'Origins and Overview of The Cognition and Affect (CogAff) Project'. This web site is updated from time to time, 2010.

[24]
Sophia Vasalou, *Wonder: A Grammar*, SUNY Press, Apr 2015.

[25]
I.P. Wright, A. Sloman, and L.P. Beaudoin, 'Towards a design-based analysis of emotional episodes', *Philosophy Psychiatry and Psychology*, **3**(2), 101-126, (1996).

Footnotes:

[1] **NB** Expanded version of summary in the Symposium proceedings

[2] Some annotated extracts are available here
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/schrodinger-life.html

[3] http://www.cs.bham.ac.uk/research/projects/cogaff/misc/meta-morphogenesis.html

[4] https://www.youtube.com/watch?v=pjtioIFuNf8

File translated from TEX by TTH, version 4.08.
On 15 Jan 2018, 21:25.