Invited talk in proceedings PPSN2000
The Sixth International Conference on Parallel Problem Solving from Nature
Paris, France,September 16-20, 2000,
Springer: Lecture Notes in Computer Science, eds Schoenauer et al.

Interacting Trajectories in Design Space and Niche Space: A Philosopher Speculates About Evolution

Aaron Sloman

School of Computer Science, The University of Birmingham
A.Sloman@cs.bham.ac.uk,
http://www.cs.bham.ac.uk/~axs/

Abstract. There are evolutionary trajectories in two different but related spaces, *design space* and *niche space*. Co-evolution occurs in parallel trajectories in both spaces, with complex feedback loops linking them. As the design of one species evolves, that changes the niche for others and vice versa. In general there will never be a unique answer to the question: does this change lead to higher fitness? Rather there will be tradeoffs: the new variant is better in some respects and worse in others. Where large numbers of mutually interdependent species (designs) are co-evolving, understanding the dynamics can be very difficult. If intelligent organisms manipulate some of the mechanisms, e.g. by mate selection or by breeding other animals or their own kind, the situation gets even more complicated. It may be possible to show how some aspects of the evolution of human minds are explained by all these mechanisms.

1 Introduction: design space and niche space

Evolution can be thought of as involving trajectories in two different but related spaces, *design space* and *niche space*. Co-evolution involves multiple parallel trajectories, with complex feedback loops. This paper attempts to explain these ideas and explore some of their implications, building on previous partial expositions ([18, 20, 22, 24, 28–30]).

1.1 What is a niche?

A biological niche is something that can produce pressure for evolutionary change. A niche is sometimes thought of as a geographical location, but since two types of organisms in the same location (e.g. a bee and a wasp flying close together) can have different niches, a niche must be something more abstract. It can be compared with what an engineer would call a ‘requirements specification’.

However a niche is more subtle than a standard set of engineering requirements, since the nature of the niche, i.e. the collection of requirements for an organism to function well, is not some fixed externally determined specification, but depends in part on the organism’s existing features and capabilities.

E.g. its digestive mechanisms partly determine the food it needs, and that determines behavioural capabilities required to obtain such food in a given environment, and that determines the kinds of knowledge the organism requires, or needs to be able to acquire, and so on.

Such considerations support the view of an organism as a collection of co-evolving sub-organisms, an idea also found in Popper’s work [14]. Minsky [11] refers to a ‘society’ of mind. From our view point 'ecosystem of mind' is more apt. In particular, evolution of motivational mechanisms can produce diversity of tastes and preferences as well as skills, abilities, knowledge, etc.

There are thus niches corresponding to components of organisms. The engineering analogue might be a dynamically changing requirements specification for work still to be done on a partially designed and implemented system, where requirements for the remaining designs will depend on design decisions already taken.

An even closer analogue would be the set of requirements for possible improvements in parts of a complete design, on the assumption that other parts, and the environment remain fixed. Different niches are determined by which components and which aspects of the environment are held fixed.

1.2 What is a design (genotype)?

Designs, like niches, are also obviously abstract, since a design may have different instantiations, especially a design for a self-modifying system. Two instances of such a design may be very different in their structure and capabilities. Different instances of *the same* abstract design can also be viewed as instances of *different* more specific designs.

Development of an individual organism involves 'moving' through a space of specific designs. Different trajectories may start in the same region of design space and then diverge, either because of cumulative effects of very small initial differences or because of different environmental influences.

The instances of such designs will have ill-defined boundaries, since, as Dawkins [4] and others have pointed out, the genotype affects not only the individual's physical and behavioural capabilities but also typical products of its behaviour, such as nests, tools, paths, furniture, etc. For our present purposes, it is not necessary to be precise about the boundaries of instances (or organisms).

2 Fitness relationships

Regions in niche space (niche types) are abstract collections of requirements, and regions in design space are sets of designs for types of behaving systems.

The instances of any particular design (genotype) will be capable of functioning more or less well in any particular region of niche space. However this does not mean that there is a numeric, or totally ordered, fitness function representing how well a design fits a niche.

If a class of designs can be specified by two parameters (e.g. length and stiffness of a spring), then there is a 2-D space of designs. Adding a measure of fitness of the spring for a particular purpose, produces a 3-D fitness landscape. Typically, design spaces are far more complex than this, and cannot be specified by a fixed number of parameters, e.g. designs for Prolog compilers vary in structure and complexity. Moreover, many designs have no single fitness measure: Prolog compilers vary according to their portability, the speed of compilation, the speed of compiled code, the size of compiled code, the kinds of error handling they support, etc.

So there is a multi-dimensional *fitness relationship* whose complexity will vary according to the design and the niche (Figure 1). An approximation to this is to regard a design and a niche as determining not a fitness *value* but a fitness *vector* whose components may be not only numerical values but descriptions ([8, 16]). E.g. one element of the vector linking a design for a predator with a niche involving certain types of prey might be something like: 'able to catch prey only in bright sunlight'.

In general there will not be an answer to the question: does this change (e.g. a mutation) lead to higher fitness? Rather there will be tradeoffs: the new variant is better in some respects and

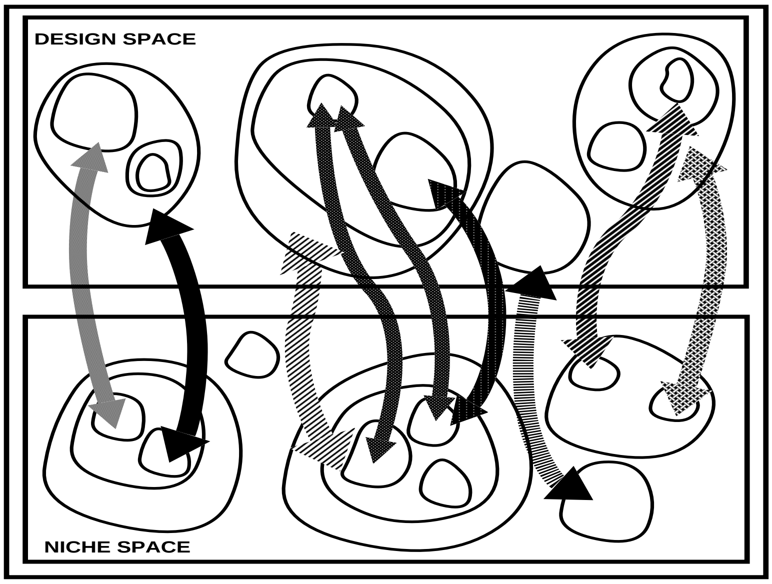

Fig. 1. Design space, niche space and mappings between them
Relations between designs and niches are complex and varied. 'Fitness' vectors describing the relations, indicated by the arrows linking designs and niches, are required rather than fitness values. Different arrows represent different types of fitness relations. Trajectories are not shown here.

worse in others, like products reviewed in consumer magazines. This is not the same as a "neutral" change which makes no relevant difference to the individual's abilities to fit the niche.

Sometimes there is a partial ordering of the fitness descriptions, and sometimes not even that, because there is no way to combine the different dimensions of comparison. Design A might be better than B in one respect, B better than C in another and C better than A in a third.

Tradeoffs between dimensions in the fitness vector may be exploited by the formation of cooperative behaviours and division of labour. Two individuals that excel in different ways (e.g. hunting and farming) may together be more competent than two with equal but intermediate levels of expertise. This is particularly true of individuals in a social group requiring many kinds of expertise. Useful division of labour can also occur across species.

3 Constraints on trajectories

Some regions of design space are not linked by possible trajectories for individual development. An acorn can transform itself into an oak tree, and by controlling its environment we can slightly modify what sort of oak tree (e.g. how big). But no matter how you try to train or coax it by modifying the environment, it will never grow into a giraffe. The acorn (a) lacks information needed to grow into a giraffe, (b) lacks the architecture to absorb and use such information, and (c) lacks the architecture required to modify itself into an architecture that can absorb the information.

Trajectories that are possible for an individual which adapts or changes itself will be called *i-trajectories*. Different sorts of i-trajectories could be distinguished according to the sorts of mechanisms of change.

Trajectories that are not possible for an individual machine or organism but are possible across generations will be called *e-trajectories* (evolutionary trajectories). Examples include development of humans and other animals from much simpler organisms and modifications of software structures by genetic algorithms. Conjectured e-trajectories leading to human minds are discussed in

[5] and [28]. If identical individuals inhabit slightly different niches, reproductive success will be favoured by different traits. E.g. in a farming community physical strength may be more important than intelligence, whereas in a nearby industrialised region intelligence is more useful for acquiring resources to raise a family. Thus different e-trajectories can be explored in parallel within a species.

Whether two designs are linked by an e-trajectory or not will depend on the type of evolutionary mechanism available for manipulating genetic structures and the type of ontogenetic mechanism available for producing individuals (phenotypes) from genotypes. In biological organisms the two are connected: the ontogenetic mechanism can also evolve. Lamarckian inheritance (used effectively in some evolutionary computation mechanisms) allows i-trajectories to form parts of e-trajectories.

There are also some changes to individuals that are possible only through external intervention by another agent, e.g. performing repairs or extensions, or debugging software. These are *r-trajectories* (repair-trajectories).

Viewing a species as a type of individual, e-trajectories for *individuals* can be thought of as i-trajectories for a *species*, or a larger encompassing system, such as an ecosystem.

When organisms are able to form societies within which information is transmitted across generations, there are kinds of changes in knowledge, preferences, skills that depend on social and cultural influences. So there are trajectories of a fourth kind *s-trajectories* (social-trajectories).

4 Discontinuous spaces and trajectories

Since designs and niches can vary in degrees and kinds of abstraction, the space of designs and the space of niches each has a very complex topology, with different neighbourhood relations at different levels of abstraction. At a particular level of abstraction a design can be modified to produce different designs at that level. *Neighbouring* designs are those reached by minimal changes in the design.

A specification mentioning the number of legs of an organism could have as a minimal change the addition of a leg (which is not necessarily biologically feasible). This might improve some aspects of fitness for a niche and degrade others. A more specific design, which includes details of the limbs, might allow minimal changes which retain the number of legs, but change aspects of their shape or other features.

A minimal change in a very specific niche for an organism might include an increase in speed of movement of its prey. A minimal change in a more abstract specification of the niche might include introduction of a new kind of predator, or a new type of prey.

We see therefore that there are trajectories in design space and niche space with different levels of abstraction, where the trajectories may include small changes or large changes, continuous changes or discontinuous changes. As one species evolves that will change the niche (the requirements) for others and vice versa, as in arms races.

Neither space is continuous, though there may be regions of continuity where requirements or designs vary only quantitatively: e.g. increasing speed of locomotion. However, Darwinian evolution is inherently discontinuous, since between any two generations there are only finitely many steps. This discreteness also follows from the structure of DNA.

Natural trajectories in the space of very specific designs are usually ‘gradual’ rather than continuous, as they involve many *small* but discrete steps. There can also be quite large steps, e.g. copying some complex structure or capability and then differentiating. In [9] Maynard Smith and Szathmaŕy refer to this as ‘duplication and divergence.’

Human designers, e.g. software engineers, aircraft designers, composers, poets, also often come up with big jumps in relation to prior designs.

4.1 Step size and searching

Small steps allow exhaustive local sampling of possible directions of change, but these may lead to unsatisfactory local optima. So larger steps are desirable. However, as the maximum step size grows, the number of options increases quickly (e.g. as the volume of an N-dimensional hypersphere increases with radius). Informed jumping without exhaustive search requires a grasp of the structure of the problem (the niche) and the region of design space. E.g. understanding why a two-legged chair doesn't stay upright helps one choose to add a third leg rather than explore variations in length or thickness of the two legs, or varying the size of the seat, etc.

Without such problem-specific knowledge an evolutionary search must use general purpose heuristics. Alas, for any such set of heuristics there will be design spaces for which those heuristics fail to find optimal, or even good, designs to fit a particular niche. In finite spaces exhaustive search is always a last resort, but where the space is large enough that is of no practical relevance.

Biological evolution, however, is generally not a search for solutions to some pre-defined problem. It is more like a large sea of viscous fluid seeping into all the nooks and crannies in some complex terrain.

Where a particular portion of the fluid goes, will depend both on the local structure of the terrain and also on the pressure from surrounding portions of fluid. The fluid's movement will change the structure of the terrain, sometimes catastrophically.

This model is partially inaccurate insofar as it suggests that the terrain, the niche space, is mostly fixed. This is not the case in co-evolution where the "terrain" through which each organism's design moves is constantly changing because of changes in other organisms. Can we find more accurate model?

4.2 Co-evolutionary trajectories

If large numbers of mutually interdependent species (designs) are co-evolving, getting a good understanding of the dynamics of the various trajectories can be very difficult. When organisms become sufficiently intelligent to understand and manipulate some of the mechanisms, e.g. by breeding other animals, by mate selection, or by selectively feeding and protecting their own infants according to their qualities, the situation gets even more complicated.

5 Causation in niche space and design space

Since niches and designs interact dynamically, we can regard them as parts of virtual machines in the biosphere consisting of a host of control mechanisms, feedback loops, and information structures. All of these are ultimately implemented in, and supervenient on physics and chemistry. But they and their causal interactions may be as real as poverty and crime and their interactions. They are not epiphenomena.

The biosphere is a very complex abstract dynamical system, composed of many smaller dynamical systems. Some of them are evanescent (e.g. tornados), some enduring but changing over diverse time scales (e.g. fruit flies, oak trees, ecosystems). Many subsystems impose constraints and requirements to be met or overcome by other subsystems: e.g. one component's design is part of another component's niche.

Through a host of pressures, forces and more abstract causal relations, including transfer of factual information and control information, systems at various levels are constantly adjusting themselves or being adjusted or modified. Some of the changes may be highly creative, including evolution of new forms of evolution, and new mechanisms for copying and later modifying modules to extend a design.

These ideas may seem wild, but they are natural extensions of ideas already accepted by many scientists and engineers, e.g. [7, 2].

Niche-pressure can influence movement of individuals or a species in design space, e.g. via adaptation along an i-trajectory, or movement of a gene pool or a subset of a gene pool, along an e-trajectory. There are many different sorts of causal relations to be analysed, including causal loops. Niches can interact with one another by producing pressure for changes in designs, which in turn can change niches. Such feedback can lead to continual change, to oscillations, or to catastrophes.

Parallel design changes can also occur *within* an individual, e.g. making an organism both physically stronger and better able to recognize complex structures. Problems arise when the changes are not independent: e.g. increasing agility may conflict with increasing strength. Which of two incompatible changes actually occurs may depend on subtle features of the total context.

Since designs have complex structures, a niche for an organism can change simply because of a change *within* its own design, without anything changing in the environment. This can generate positive feedback loops driving designs along e-trajectories without any environmental changes.

Where independent changes in different dimensions are possible, causation may be conditional. In one context one change will be favoured, and a different change in another context. Such changes in different directions can happen in parallel in members of the same species in slightly different niches, possibly leading to useful functional differentiation.

Design space and niche space are “layered”: regions within them are describable at different levels of abstraction and for each such region different “specialisations” exist. Some specialisations of designs are called implementations. The philosopher’s notion of “supervenience” and the engineer’s notion of “implementation” (or realisation) seem to be closely linked, if not identical. There are many confusions about these relations still to be clarified ([15]).

Both are inhomogeneous spaces: local topology varies with location in the space, since the minimal changes possible at various locations in the same space can be very different in type and number. Consider designs of different complexity: there are typically more ways and more complex ways, of altering a complex design than a simple design. So they have neighbourhoods of different structures. By contrast, in most multi-dimensional spaces considered by scientists and engineers (e.g. phase spaces), each point has the same number of dimensions, i.e. the same number and the same types of changes are possible at all points (unless limited by equations of motion).

5.1 Evolution of mind

Both design space and niche space have very complex topologies, including many discontinuities, some small (e.g. adding a bit more memory to a design, adding a new step in a plan) some large (adding a new architectural layer, or a new formalism). Understanding natural intelligence may require understanding some major discontinuities in the evolutionary history of the architectures and mechanisms involved. This in turn may help us with the design of intelligent artefacts.

Important types of discontinuity in design space occur between systems that are able merely to perform certain tasks reactively, perhaps with slight modifications due to learning, and others which can use generalisations they have learnt about the environment to create new plans, i.e. between reactive and deliberative architectures. Deliberative capabilities, supporting ‘what if’ reasoning, require a collection of mechanisms not needed for purely reactive systems.

Some of these evolutionary discontinuities may have occurred in e-trajectories where an old mechanism was copied then modified. E.g. a mechanism which originally associates sensory patterns with appropriate responses could be copied and the new version used to associate *possible* future sensory patterns with predicted sensory patterns or with a set of available responses.

Discontinuities might also be involved in the evolution of the "reflective" abilities described below: not only being able to do X but having and being able to use information on how X was done, or why X was done, or why one method of doing X was used rather than another. (Compare [22, 26, 30].) What sorts of niche pressures in nature might favour such e-trajectories is an interesting biological question.

6 Never-ending discontinuous i-trajectories.

A system which develops, learns or adapts changes its design. I-trajectories, like e-trajectories can be discontinuous (e.g. cell division) and link regions in inhomogeneous spaces. The most familiar examples are biological: e.g. a fertilised egg transforming itself into an embryo and then a neonate. In many animals, including humans, the information processing architecture seems to continue being transformed long after birth, and after the main physiological structures have been established: new forms of control of attention, learning, thinking, deliberating, develop after birth. Ontogeny may partly recapitulate phylogeny: but cultural influences may change this.

Humans follow a very complex trajectory in design space throughout their lives. A good educational system can be viewed as providing a trajectory through niche space which will induce a trajectory in design space in self-modifying brains. A culture provides a set of developmental trajectories.

In general, following a trajectory in design space also involves a trajectory in niche space: the niches for an unborn foetus, for a newborn infant, a schoolchild, a parent, a professor, etc. are all different. Moreover, an individual can instantiate more than one design, satisfying more than one niche: e.g. protector and provider, or parent and professor. To cope with development of multi-functional designs we can include *composite niches* in niche space, just as there are composite designs in design space.

7 Trajectories for Virtual Machines in Software Systems

The distinction between i-trajectories and e-trajectories can be applied to software individuals inhabiting virtual machines. A word processor which follows an i-trajectory in which it adapts itself to different users may or may not be capable of turning itself into a language understanding system through such a series of adaptations, even if evolutionary mechanisms could so transform it. As with organisms there may be e-trajectories linking software designs that are not linked by i-trajectories.

Whether an e-trajectory exists from one software design to another in an artificial evolutionary system depends on (a) whether there is a principled way of mapping the features of the designs onto genetic structures which can be used to recreate design instances via an instantiation (ontogenetic) function, and (b) whether the structures can be manipulated by the permitted operators so as to traverse a trajectory in "gene space" which induces a trajectory in design space via the instantiation function. Whether some sort of evaluation function or niche pressure can cause the traversal to occur is a separate question [13]. E-trajectories can exist which our algorithms never find.

8 Evolution of Human-like Architectures

We have argued in [19, 27] and elsewhere (*contra* Dennett's "intentional stance") that many familiar mental concepts presuppose an information processing architecture. We conjecture that it involves several different sorts of coexisting, concurrently active, layers, including an evolutionarily old "reactive" layer involving dedicated highly parallel mechanisms each responding in a fixed way to its inputs. These may come from sensors or other internal components, and the outputs may go to motors or internal components, enabling loops. Some reactive systems have a fixed architecture except insofar as weights on links change through processes like reinforcement learning. Insects appear to have purely reactive architectures implementing a large collection of evolved behaviours. As suggested in [23, 28], sophisticated reactive architectures may need a global "alarm" mechanism to detect urgent and important requirements to override relatively slow "normal" processes. This can interrupt and redirect other subsystems (e.g. freezing, fleeing, attacking, attending).

A hybrid architecture, as shown in Figure 2, could combine a reactive layer with a "deliberative" layer which includes the 'what if' reasoning ability needed to create new temporary structures representing alternative possibilities for complex future actions, which it can then compare and evaluate, using further temporary structures describing similarities and differences. This plan-construction requires a long term memory associating actions in contexts with consequences. After creating and selecting a new structure the deliberative system may execute it as a plan, and then discard it. 'Skill compilers' can use practice to save new plans within the reactive layer for future use in fast, fluent mode. This happens when humans learn to drive a car or speak a new language.

Global alarm mechanisms may be needed for coping with dangers and opportunities requiring rapid reactions. In mammals this seems to use a variety of old and new mechanisms, including the limbic system. Emotional processes are often involved [3, 6].

A deliberative mechanism will (normally) be discrete, serial, and therefore relatively slow, whereas a reactive mechanism can be highly parallel and therefore very fast, and may include some continuous (analog) mechanisms, possibly using thresholds. Resource limits in deliberative mechanisms may generate a need for an attention filter of some kind, limiting the ability of reactive and alarm mechanisms to interrupt high level processing.

By analysing tradeoffs we may be able to understand how niche-pressures can lead to development of combined, concurrent, deliberative and reactive architectures in certain organisms. Other types explored alternative trade-offs, e.g. with large numbers of simple and expendable individuals.

Everything that can be done by a hybrid architecture could in principle be done by a suitably complex reactive architecture e.g. a huge, pre-compiled lookup table matching every possible history of sensory inputs with a particular combination of outputs. However, pre-requisites for such an implementation may be prohibitive: much longer evolution, with more varied evolutionary environments, to pre-program all the reactive behaviours, and far more storage to contain them, etc. For certain agents the universe may be neither old and varied enough for such development nor big enough to store all the combinations required to match a deliberative equivalent with generative power. Perhaps evolution "discovered" this and therefore favoured deliberative extensions for some organisms.

A deliberative mechanism changes the niches for perceptual and motor mechanisms, requiring them to develop new layers of abstraction, as indicated in Figure 2. Likewise, development of new, higher level, abstractions in perceptual and motor systems may change the niches for more central mechanisms, e.g. providing new opportunities for learning and simplified planning.

Meta-management. Reflection on and retrospective evaluation of actions can often lead to future improvements. This is also true of *internal* actions. Thus besides abilities to perceive the environment and how external actions change it, there is a use also for internal self-monitoring, self-evaluation, self-modification (self-control) applied to *internal* states and processes. This could explain the evolution of a third architectural layer, as indicated in Figure 2.

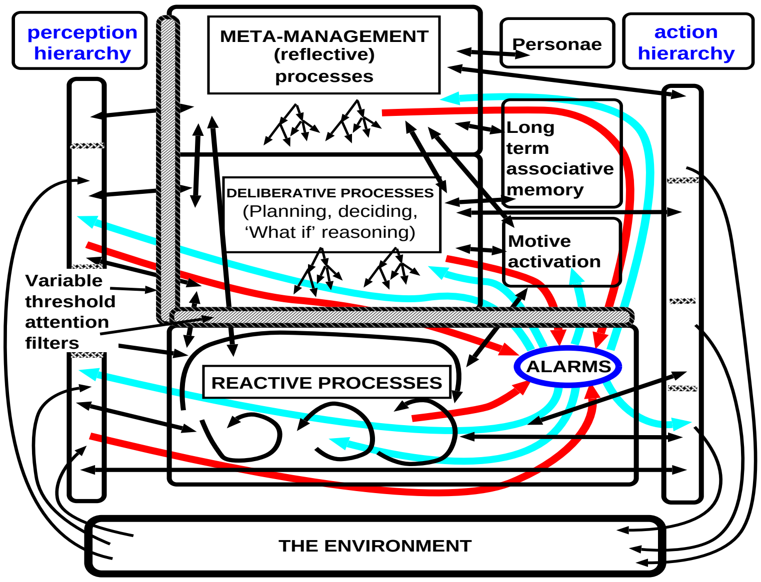

Fig. 2. Design for a human-like system
The reactive, deliberative and meta-management (reflective) layers evolved at different times, requiring discontinuous changes in the design, and providing significantly new capabilities. Three classes of emotions correspond to events involving the three layers. Simpler organisms, e.g. insects and reptiles, have fewer layers. (Figure revised April 2004)

Sensory "qualia" arise in self-monitoring mechanisms with access to intermediate sensory information structures not normally attended to. Different kinds of sensory qualia depend on different perceptual abstraction layers. Such "self-knowledge" is distinct from normal perception providing knowledge about the environment. "Meta-management" capabilities produce other sorts of qualia related to thinking processes, deliberation, desires, etc.

Robots with these capabilities might begin to wonder how their mental processes are related to their physical implementation, just as human philosophers do. Some of them, not fully understanding the notion of virtual machine functionality and the varieties of forms of supervenience, might even produce spurious but convincing arguments that they have conscious processes which cannot be explained by or fully implemented in physical processes. They may wonder whether humans are actually *zombies* with all the behavioural capabilities of conscious robots, but lacking their consciousness. I believe this solves the so-called "hard" problem of consciousness, see [1]. (Earlier papers exploring these ideas are in the bibliography).)

Such agents (with a combination of reactive, deliberative and self-management sub-architectures) may combine to form social systems. Questions about trajectories in design space and niche space arise for social systems also. Human social systems, follow s-trajectories and develop information and rules which are transmitted to individuals, including rules that control meta-management (e.g. through guilt).

9 Conclusions and Further Work

Milner [10] noted that in computer science theory often follows the lead of engineering intuitions. The process of mathematical formalisation can lag far behind. Likewise attempts to study and formalise the space of possible designs and the various trajectories in design space will lag behind intuitive understanding gained from empirical research in biology, psychology, and computational explorations.

Many have attempted to formalise features of evolution, individual learning, development etc. Kauffman [7] describes mathematical laws which constrain biological mechanisms and processes in surprising ways. The ideas discussed here deal with phenomena which still seem to be too ill defined for mathematical formulation and computational modelling. However that may change.

We need to find more precise ways of describing architectures, designs, niches and their causal interactions, to improve on the high level concepts used only intuitively at present. This will involve both abstracting from domain specific details, so as to replace empirical concepts with mathematical concepts, and also enriching our understanding of the details of the processes, so that we can characterise and model the dynamics.

If the intuitive notions of niche, genotype etc. in biology can be made sufficiently precise to enable us to understand precisely the relationships between niches and designs for organisms, this may provide a better understanding of the dynamics and trajectories in biological evolution, including the evolution of evolvability.

This could lead to advances in comparative psychology. Understanding the precise variety of types of functional architectures in design space and the virtual machine processes they support, will enable us to describe and compare in far greater depth the capabilities of various animals. We'll also have a conceptual framework for saying precisely which subsets of human mental capabilities they have and which they lack. Likewise the discussion of mental capabilities of various sorts of machines could be put on a firmer scientific basis, with less scope for prejudice to determine which descriptions to use. E.g. instead of arguing about which animals, which machines, and which brain damaged humans have consciousness, we can determine precisely which sorts of consciousness they actually have.

We could also derive new ways of thinking about human variability and the causes and effects of mental illness, brain damage, senile dementia, etc. This could have profound practical implications.

Finally it should be noted that designs do not presuppose a designer and requirements (niches) do not presuppose a requirer.

Notes and Comments

This work is partly funded by the Leverhulme Trust. My ideas have benefited from discussions with many colleagues in the School of Computer Science at the University of Birmingham and Brian Logan in the School of Computer Science & IT University of Nottingham. Most of the papers from our group are available at the Cognition and Affect project web site:

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

Our software tools for exploring agent designs are included with the Free Poplog software repository:

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

References

1. Chalmers, D.J.: *The Conscious Mind: In Search of a Fundamental Theory.* Oxford University Press, New York, Oxford, 1996.

2. Cohen, J., Stewart, I.: *The collapse of chaos*. Penguin Books, New York, 1994.
3. Damasio, A.R.: *Descartes' Error, Emotion Reason and the Human Brain*. Grosset/Putnam Books, New York, 1994.
4. Dawkins, R.: *The Extended Phenotype: The long reach of the gene*. Oxford University Press, Oxford, New York, 1982.
5. Dennett, D.C.: *Kinds of minds: towards an understanding of consciousness*. Weidenfeld and Nicholson, London, 1996.
6. Goleman, D.: *Emotional Intelligence: Why It Can Matter More than IQ*. Bloomsbury Publishing, London, 1996.
7. Kauffman, S.: *At home in the universe: The search for laws of complexity*. Penguin Books, London, 1995.
8. Logan, B.S, Sloman, A.: Agent route planning in complex terrains. Technical Report CSRP-97-30, University of Birmingham, School of Computer Science, 1997.
9. Maynard Smith, J., Szathmaŕy, E.: *The Origins of Life: From the Birth of Life to the Origin of Language*. Oxford University Press, Oxford, 1999.
10. Milner, R.: Semantic ideas in computing. In Ian Wand and Robin Milner, editors, *Computing Tomorrow: Future research directions in computer science*, 246–283. Cambridge University Press, 1996.
11. Minsky, M.L.: *The Society of Mind*. William Heinemann Ltd., London, 1987.
12. Peterson, D.M. (editor): *Forms of representation: an interdisciplinary theme for cognitive science*. Intellect Books, Exeter, U.K., 1996.
13. Poli, R: Logan, B.S.: On the relations between search and evolutionary algorithms. Technical Report CSRP-96-07, School of Computer Science, The University of Birmingham, March 1996.
14. Popper, K.R.: *Unended Quest*. Fontana/Collins, Glasgow, 1976.
15. Scheutz, M.: When physical systems realize functions... *Minds and Machines*, 9:161–196, 1999. 2.
16. Sloman, A.: How to derive "better" from "is". *American Phil. Quarterly*, 6:43–52, Jan 1969.
17. Sloman, A.: The mind as a control system. In C. Hookway and D. Peterson, (eds), *Philosophy and the Cognitive Sciences*, 69–110. Cambridge University Press, 1993.
18. Sloman, A.: Explorations in design space. In A.G. Cohn, editor, *Proc. 11th European Conf. on AI, Amsterdam, August 1994*, 578–582, Chichester, 1994. John Wiley.
19. Sloman, A.: Semantics in an intelligent control system. *Philosophical Transactions of the Royal Society: Physical Sciences and Engineering*, 349(1689):43–58, 1994.
20. Sloman, A.: Exploring design space and niche space. In *Proc. 5th Scandinavian Conf. on AI, Trondheim*, Amsterdam, 1995. IOS Press.
21. Sloman, A.: Towards a general theory of representations, in Peterson (1996). In Peterson, D.M: (ed), *Forms of representation: an interdisciplinary theme for cognitive science*, 118–140. 1996.
22. Sloman, A.: What sort of control system is able to have a personality. In Trappl, R., Petta, P.: (eds), *Creating Personalities for Synthetic Actors: Towards Autonomous Personality Agents*, 166–208. Springer (Lecture Notes in AI), Berlin, 1997.
23. Sloman, A.: Damasio, Descartes, alarms and meta-management. In *Proc. Int. Conf. on Systems, Man, & Cybernetics (SMC98), San Diego*, 2652–7. IEEE, 1998.
24. Sloman, A.: The "semantics" of evolution: Trajectories and trade-offs in design space and niche space. In Coelho, C: (ed), *Progress in Artificial Intelligence, 6th Iberoamerican Conf. on AI*, 27–38. Springer, Lecture Notes in Artificial Intelligence, Lisbon, October 1998.
25. Sloman, A.: What sort of architecture is required for a human-like agent? In Wooldridge, A. Rao, A.: (eds), *Foundations of Rational Agency*, 35–52. Kluwer Academic, Dordrecht, 1999.
26. A. Sloman and B. S. Logan. Architectures and tools for human-like agents. In Ritter,F., Young, R.M., (eds), *Proc. the 2nd European Conf. on Cognitive Modelling*, 58–65, Nottingham, UK, 1998. Nottingham University Press.
27. Sloman, A.: Architecture-based conceptions of mind. In *Proc. 11th Int. Congress of Logic, Methodology and Philosophy of Science*, Cracow, Poland, August 1999.
28. Sloman, A.: Architectural requirements for human-like agents both natural and artificial. (what sorts of machines can love?). In Kerstin Dautenhahn, editor, *Human Cognition And Social Agent Technology*, Advances in Consciousness Research, 163–195. John Benjamins, Amsterdam, 2000.
29. Sloman, A., Logan, B.S.: Building cognitively rich agents using the Sim_agent toolkit. *Communications of the Association of Computing Machinery*, 42(3):71–77, March 1999.
30. Sloman, A and Logan, B. Evolvable architectures for human-like minds. In *Proc. 13th Toyota Conf., on Affective Minds Shizuoka, Japan, Nov-Dec 1999*. Elsevier, 2000.