Meta-morphogenesis and the Creativity of Evolution

Aaron Sloman [1]

Abstract.
Whether the mechanisms proposed by Darwin and others suffice to explain the achievements of biological evolution remains open. One problem is the difficulty of knowing exactly what needs to be explained. Evolution of information-processing capabilities and supporting mechanisms is much harder to detect than evolution of physical form, and physical behaviours in part because much goes on inside the organism, and in part because it often has abstract forms whose physical manifestations do not enable us to identify the abstractions easily. Moreover, we may not yet have the concepts required for looking at or thinking about the right things. AI should collaborate with other disciplines in attempting to identify the many important transitions in information processing capabilities, ontologies, forms of representation, mechanisms and architectures that have occurred in biological evolution, in individual development (epigenesis) and in social/cultural evolution – including processes that can modify later forms of evolution and development: meta-morphogenesis. Conjecture: The cumulative effects of successive phases of meta-morphogenesis produce enormous diversity among living information processors, explaining how evolution came to be the most creative process on the planet.

1 Life, information-processing and evolution

Research in a variety of disciplines has contributed a wealth of observations, theories and explanatory models concerned with the diversity of living organisms on many scales, from sub-microscopic creatures to very large animals, plants and fungi, though many unsolved problems remain about the processes of reproduction, development and growth in individual organisms. Many animal competences are still not replicated in machines. I suggest this is in part because of the difficulty of characterising those competences with sufficient precision and generality. Instead researchers focus on special cases inadequately analysed and their models do not "scale out". By studying many more intermediate stages in evolution and development we may achieve deeper understanding of existing biological information processing, and find clues regarding the layers of mechanisms supporting them.

Conjecture: we cannot understand specific sophisticated animal competences without understanding the creativity of biological evolution that produces not only those designs, but also many others. Studying only a few complex cases of animal cognition, for instance pursuing the (in my view hopelessly ill-defined) goal of "human-level AI" [13], may be like trying to do chemistry by studying only a few complex molecules. Likewise trying to replicate selected aspects of some competence (e.g. 3-D vision) while ignoring others may lead to grossly oversimplified models, such as AI "vision" systems that attach labels (e.g. "mug") to portions of an image but are of no use to a robot trying to pick up a mug or pour liquid out of it. Solutions need to "scale out" not just "scale up".[2]

I'll attempt to explain the conjecture, inviting collaboration on the task of identifying and analysing transitions in information processing functions and mechanisms produced by evolution, in humans and also in other species that inhabit more or less similar niches. This is the "meta-morphogenesis" project.[3] In contrast, recent fashions, fads, and factions (e.g. symbolic, neural, dynamical, embodied, or biologically inspired AI) may all turn out to be limited approaches, each able, at best, to solve only a subset of the problems.

2 Diversity of biological information-processing

Every complex organism depends on many forms of information-processing, for controlling aspects of bodily functioning, including damage detection and repair, along with growth and development of body-parts and their functions, and also for behaviours of whole individuals at various stages of development, and also new learning.

Much research has been done on transitions produced by evolution, but, as far as I know, there has not been systematic investigation of *evolutionary transitions in information-processing functions* and mechanisms and their consequences. In [12] the main transitions in information-processing mentioned are changes in forms of communication, ignoring non-communicative uses of information, e.g. in perception, motivation, decision making, learning, planning, and control of actions [18, 19], which can both evolve across generations and change during development and learning. In some species, there are also changes of the sort labelled "Representational Redescription" in [11]. There are also within-species changes in cooperative or competitive information processing, including variation between communities. Conjecture: changes in information-processing help to *speed up and diversify* processes of evolution, learning and development. For example, evolution of individual learning mechanisms, allowed products of evolution to change more rapidly, influenced by the environment.

Forms of representation and ontologies. We have known for decades that how information is represented can significantly affect uses of the information, including tradeoffs between rigour and efficiency, ease of implementation and expressive power, applicability of general inference mechanisms and complexity of searching. I suspect that similar constraints and tradeoffs, and probably many more were "discovered" long ago by biological evolution. As far as I know nobody has surveyed the tradeoffs and transitions that are relevant to uses of information in organisms. There are comparisons between the generality of logic and the

---

[1] School of Computer Science, University of Birmingham, UK web: http://www.cs.bham.ac.uk/~axs A longer, open access, version of this paper is freely available at http://tinyurl.com/BhamCog/12.html#1203

[2] Compare McCarthy's requirement for "elaboration tolerance"

[3] http://tinyurl.com/BhamCog/misc/m-m.html

usefulness of domain specific "analogical" representations [17, Chap 7]; and between representing structures, properties and relationships with high precision and "chunking" information into fuzzy categories, useful, for example, in learning associations, making predictions and forming explanations, each covering a range of possibilities with small variations [36]. Evolution seems to have discovered the importance of such discretisation, including meeting requirements related to learning about generalisations that hold across time and space, for instance generalisations about the properties of different kinds of matter, and generalisations about consequences of various types of action in various conditions.

Somatic and exosomatic ontologies A survey of varieties of *information contents* available to organisms would include types restricted to internal and external sensor states and effector signals, i.e. *somatic* information, and also the *exosomatic* ontologies used in organisms that evolved later, referring to objects, relationships, processes, locations, routes, and other things outside themselves. Still more sophisticated organisms can speculate about and learn about the hidden contents of the different kinds of matter found in the environment, including humans developing theories about the physics and chemistry of matter, using newly created exosomatic, theory-based (ungroundable) ontologies.[21]

Ontologies with relations Exosomatic ontologies typically locate objects, parts of objects, structures, events and processes in both space and time, so that they have spatial and temporal relationships. Information about relationships can be essential for some forms of action, e.g. direction and distance to something dangerous or something desirable, or whether helpless offspring are hidden in a tunnel or not. Spatial relations can involve different numbers of entities - X is above Y, X is between Y and Z, X is bigger than the gap between Y and Z, etc. Some objects, and some processes, have many parts with multiple relationships between them, and processes include various ways in which relationships can change, continuously or discretely. (Compare [14].) Do we know which species can acquire and use relational information, and when or how it first evolved, or how many forms it can take, including logical (Fregean) and analogical (e.g. diagrammatic, pictorial, model-based) representations? Early biological relational representations were probably molecular. Multi-strand relations involve objects with parts related to other objects with parts e.g. parts of a hand and parts of a mug. Which animals can reason about multi-strand processes?

Information about causal relationships is essential for making plans and predictions. It is not clear what sorts of causal understanding different organisms can have. Jackie Chappell and I have argued for at least two different sorts of causal knowledge (a) correlational/statistical causation (Humean) and (b) structural, mathematically explainable causation.[4] When did they evolve?

How should scalar variation be represented? A common assumption by researchers in several disciplines is that organisms and intelligent robots necessarily represent spatial structures and relationships using global metrics for length, area, volume, angle, curvature, depth speed, and other scalar features. These modes of representation first occurred in human thought only relatively recently (following Descartes' arithmetisation of geometry), so they may not be available to young children and other animals: perhaps evolution produced much older, and in some ways more powerful, ways of representing and using spatial relationships, without numerical coordinate systems? I suspect that Descartes' forebears, many animals, and pre-verbal children in our culture make use of networks of partial orderings (of distance, direction, angle, curvature, speed, size, and other properties) enhanced with semi-metrical relations refining orderings (e.g. X is at least three times as long as Y but not more than four times as long). Specifying exactly how such representations might work remains a research problem. Obviously, many animals including nest-building birds, primates, hunting mammals, and elephants understand spatial structures and affordances in ways that are far beyond the current state of computer vision/robotics. Neuroscientists and vision researchers in psychology seem to lack a theoretical framework to describe or explain such competences. One problem in such research is a tendency to confuse the ability to understand and reason about spatial relationships and processes with the ability to *simulate them*, as is done in computer game engines. Our brains cannot perform similar simulations.

Conditional control. Organisms need to be able to generate motor control signals or sequences of signals partly on the basis of information about the environment and partly under the control of goals and plans. For this, information is needed about internal states, such as energy or fluid needs, and also predicted needs, so as to initiate actions to meet anticipated requirements. Such choices depend on information about both external states and internal states (e.g. desires, preferences). So requirements and uses for information processing can vary in ways that depend on static or changing factors, some within the organism (e.g. need for a particular sort of nutrient), some in the environment (e.g. the local or remote spatial relationships between various surfaces and objects), and some of that depend on the sensory-motor morphology of the organism, e.g. whether it has an articulated body with mobile grippers, and whether it has visual, olfactory, auditory, tactile, haptic, proprioceptive or other sensors.

Precocial/Altricial tradeoffs Additional information-processing requirements depend on how individuals change in shape, size, strength and needs, which depend on what parents can do to help offspring. Many carnivores and primates are born weak and helpless and as they grow, larger, heavier and stronger, they engage in forms of movement for which new kinds of control are required, not all encoded in the genome (for example manipulation of objects that did not exist in the evolutionary history of the species [34]).

In many species, development requires use of information about the environment in setting and achieving ever more complex goals, allowing cumulative development of forms of control required by adults. This process can include play fighting, using conspecifics of similar size and competence. Contrast larvae, that, after a phase of crawling and eating, pupate and transform themselves into butterflies that apparently do not need to learn to fly, feed or mate. Information for the later phase of must somehow have been present in the caterpillar stage where it was of no use. Some of the tradeoffs between nature and nurture found in animals and likely to be relevant to future robots are discussed in [31, 5]. Not using those biological forms of representation may explain why our robots, impressive as they are in limited ways, lack the generality and flexibility of pre-verbal humans and many other animals.

On-line *vs* off-line intelligence. The simplest known organisms are surprisingly complex.[5] All require information-based control for growth and reproduction, unlike sediment layers that simply accrue whatever external physical processes provide. Informed growth requires selection of nutrients outside the organism. If not everything in the environment is suitable, microbes can use sensors that react differently to chemicals in the surrounding soup, ingesting only nutrients (except when deceived). Such organisms have information-

---

[4] http://tinyurl.com/BhamCog/talks/wonac/

[5] https://en.wikipedia.org/wiki/Archaea

processing needs that are highly localised in space and time: so that transient sensing and control suffice – perhaps even just a fixed set of triggers that initiate responses to different types of contact. Complex online control uses continuously sensed information, e.g. about directions, about changing gaps, about local chemical gradients, used in deciding whether to modify motor signals, e.g. so as to increase concentration of nutrients or decrease concentration of noxious substances, or towards or away from light, etc. Using direction and magnitude of changes requires more complex mechanisms than detecting presence or absence, or thresholding. Feedback control using "hill-climbing" requires access to recent values, so that new ones can be compared with old ones in order to select a change.

On-line intelligence involves using information as it is acquired. *Off-line* intelligence acquires information usable later, in combination with other information, and for several different purposes. Off-line mechanisms transform sensed or sent information into new formats, stored for possible uses later, if required. Storing more abstract information can be useful because very precise details may not be relevant when one is thinking or reasoning about a situation that one is not in at the time, and also because information in a more economical and abstract form may allow more useful generalisations to be discovered, and may be simpler to combine with other forms of information.

Combining on-line and off-line intelligence. Doing something and understanding why it works requires parallel use of on-line and off-line intelligence. Some tasks, for instance mapping terrain while exploring it (SLAM) combine online and offline intelligence, as new sensor information is integrated into an multi-purpose representation of the large scale structure of the environment, where useful spatial/topological relationships and spatial contents are stored, not sensor readings. However, it is useful sometimes to store "summary sensory snapshots" for comparison with future snapshots, or to allow information to be derived from the low level details at a later time.

All this requires specific mechanisms, architectures, and forms of representation. Their uses will depend on what the environment is like and on previously evolved features of the species. We need more detailed analyses of the different functions and the mechanisms required for those functions, and how their usefulness relates to various environments and various prior design features.

Duplicate then differentiate *vs* abstraction using parameters A common pattern of change leading to more complex biological structures or behaviours starts by duplicating an already learnt or evolved specification, then allowing one, or both, copies to change, either across generations or within a lifetime. Without this a single fertilised cell could not grow into a complex organism with varied parts competences. That is also a common pattern in the development of engineering design knowledge. Another common pattern in mathematics and engineering inserts gaps into something learnt, to form a re-usable specification whose instances can take many forms that depend on the gap-fillers, e.g. algebraic structures defined in terms of types of operators and types of objects, which take different forms for different instances. This can also be a powerful form of individual learning. I suspect evolution also found ways to use it, speeding up evolution by allowing new complex sub-systems to be created by instantiating existing patterns (as opposed to duplicating old instances). This can support learning in diverse environments. It is a core feature of mathematical discovery. We need to study more biological examples.

Use of virtual machinery. Use of virtual machinery instead of physical machinery often facilitates extendability, monitoring, de-bugging, and improving designs and re-using them in new contexts. Conjecture: biological evolution "discovered" advantages of use of virtual machinery long before human engineers did, especially in self-monitoring and self-modifying systems, with many important consequences. Some virtual machines merely provide new implementations of functionality previously provided in hardware, whereas others are non-physically specified, for example, virtual machines for performing operations like forming intentions, detecting threats, evaluating strategies, extending ontologies. Describing these requires use of concepts like information, reference, error, perception, trying, avoiding, failing, planning, learning, wanting, and many more that are not definable using concepts of the physical sciences. When chess virtual machine runs we can describe what it does using concepts like pawn, threat, detect, fork, mate, plan, attempt, fail, but those descriptions cannot be translated into the language of physics, even though the chess machine is fully *implemented* physically. A translation would have to summarise all possible physical implementations using different technologies, including future ones about which we currently know nothing, so our concepts cannot presuppose their physical features [26].

Such virtual machinery is *fully implemented* in physical mechanisms (some of which may be in the environment) and cannot survive destruction of the physical computer, though a running VM can sometimes be transferred to a new computer when a physical malfunction is imminent: an option not yet feasible for biological virtual machinery. Mechanisms for supporting a class of virtual machines can enormously simplify the process of producing new instances, compared with having to evolve or grow new instances with new arrangements of physical matter. This could speed up both evolution and learning, as it speeds up engineering design.

Besides *single function* virtual machines (or application machines, e.g. a spelling checker) there are also *platform virtual machines* that support development of a wide range of additional machines implemented on the platforms, sharing the benefits of previously developed VM components with multiple uses. Platform VMs include programming language systems (e.g. a python VM) and operating systems (e.g. a linux VM). Contrary to the common notion of computation as inherently *serial* (as in a simple Turing Machine) many VMs inherently include *multiple concurrently active subsystems* interacting with one another and with things outside the machine (e.g. information stores, sensors, robot arms, displays or other networked systems).[6] Perhaps evolution of new platform VMs sped up evolution of new information-processing functionality.

These ideas raise deep unanswered questions about how specifications for different sorts of development and learning capabilities are encoded in a genome, and what needs to change in decoding processes to allow changes from mechanisms specified by their hardware (e.g. chemical implementation) to mechanisms encoded in terms of a previously evolved virtual machine.

Specifying functions rather than behaviours or mechanisms Human engineers and scientists have increasingly used virtual machinery to achieve more sophisticated design goals, driven by new engineering requirements, including the need for programs too large to fit into physical memory, the need to be able to run a program in different parts of physical memory without altering addresses for locations and the need to use novel forms of hardware. Design machines specified in terms of information processing *functions*

---

[6] The CogAff schema allows diverse highly concurrent VMs of varying complexity and functionality http://tinyurl.com/BhamCog/#overview

rather than their physical *structures and behaviours*, postpones the task of producing physical implementations and allows different solutions. Many computing systems are specified not in terms of the behaviours of electrons or transistors, etc., but in terms of operations on numbers, strings, arrays, lists, files, databases, images, equations, logical formulae, mathematical proofs, permissions, priorities, email addresses, and other notions relevant to providing a computing service. Programmers attempting to debug, modify, or extend such programs, normally do not think about the physical processes, but about the structures and processes in the running VM. Explaining how the program works, and what went wrong in some disaster typically involves reference to events, processes and causal interactions within the VM, or in some cases relations between VM processes and things in the environment.

Some philosophical functionalists define mental phenomena in terms of how they affect input-output mappings, e.g. [4], but this ignores designs for complex virtual machinery specified in terms of structures, processes and causal interactions in the machine, not input-output relationships – "virtual machine functionalism".

Meta-semantic competences and ontologies A semantic competence is the ability to refer to things. A *meta-semantic competence* involves being able to think about, reason about, make use of, or detect something that refers, or intends, or perceives (including possibly oneself). Such competences can take many forms. Some are shallow, while others are deep. Abilities to detect aspects of X's behaviour that indicate what X perceives, or what it intends, or whether it is annoyed or fearful, etc. can feed into decisions about how to act towards X. In the shallowest forms this can involve only evolved or learnt reactions to shallow behaviours (e.g. running, snarling), etc. Deeper meta-semantic competences include representing specific contents of percepts, intentions, preferences, beliefs, etc. of others, and possibly hypothetical reasoning about such states (what would X do if it knew that A, or desired B?). Dennett, in [7], and elsewhere, refers to this as adopting "the intentional stance", but seems to be reluctant to accept that that can involve representing what is going on inside the individual referred to. Developmental psychologists have studied "mind-reading" abilities, e.g. [2], but we still lack a comprehensive theory of the varieties of forms of semantic competence, their biological roles, which organisms have them, how they evolved, how they develop in individuals, how they can vary from one individual to another, and so on. The more sophisticated meta-semantic competences require abilities to refer to virtual machine events, states and processes. How this is done, including handling "referential opacity" is still a matter of debate: some researchers emphasis use of special logics (modal logics), while others (rightly!) emphasise architectural support for meta-semantic reasoning.

Re-usable protocols Recent history of computing included development of many specifications of re-usable protocols including networking protocols, protocols for communication with peripheral devices (screens, sensors, keyboards, etc.) and protocols for inter-process communication (among many others). Use of DNA and a set of transcription mechanisms can be viewed as a biological version of a multi-function protocol. There may be many others worth looking for, perhaps not shared universally, but perhaps shared between species with a common heritage, or between different functions within individuals or within a species. I conjecture that the advantages of use of VMs for specifying new functionality, for debugging, for modifying, extending, analysing processes were "discovered" by evolution long before human engineers. This suggests that much mental functioning cannot be understood as brain functioning, and research into minds and brains, what they do, and how they work, needs to be informed by what can be achieved by VMs whose relationship to the physical machinery of the brain may be very complex and indirect. How and when this first occurred, and how specifications for virtual implementations are encoded in genomes are unanswered questions. Some new biological competences initially developed using VMs might later use more efficient, but more inflexible, physical implementations. Sometimes the reverse might occur: competences implemented in brain mechanisms are later be replaced by VMs that provide more flexibility, more extendability, and more diversity of use [5].

Self-monitoring at a VM level. Programs that monitor and modify running systems (including themselves) can benefit from focusing on VM structures and processes as well as the underlying physical machinery. I suspect biological evolution found many uses for VMs long before there were humans on the planet. If machines or animals can introspect enough to find out that they create and manipulate non-physical entities, that could lead them to invent muddled philosophical theories about minds and bodies, as human philosophers have done [26, 28].

Representing the actual and the possible (i.e. affordances). Information-processing functions so far described involved acquiring, transforming, storing, combining, deriving, and using information about what is or has been the case, or what can be predicted: types of *factual* information. Some organisms can also represent and use information that is not about what exists but rather about what is, was, or will be *possible*. This may require new architectures, forms of representation, and mechanisms. The ability to acquire and use short-term information about possibilities for and restrictions on physical action, and restrictions on action was referred to by Gibson [8] as the ability to perceive and use "affordances", where the affordances can be either positive (enabling or helping) or negative (preventing, hindering or obstructing). There are many more ways of detecting, reasoning about, producing, or using possibilities for change in the environment or restrictions on possibilities [20, 30], included in competences of particular individuals or particular types of organism. These include representing proto-affordances (possibilities and constraints involving physical objects), vicarious affordances (for other agents - including predators, prey, collaborators, offspring, etc.), epistemic affordances, deliberative affordances, and others described in [30, 25]. For organisms with meta-semantic competences (summarised above) types of affordance that can arise will be much greater than for animals that can represent or reason only about physical/spatial possibilities.

Yet more complexity in the ontology used, the forms of representation, and the information processing arises from the need not only to represent what actually exists, at any time, but also what is and is not possible, what the constraints on possibilities are, and how those possibilities and constraints can depend on other possibilities.

People can use information without being able to answer questions about it, e.g. human syntactic competences. So tests for meta-semantic competences in young children can be misleading if the tests require explicit meta-knowledge.[7] When and how all these information-processing capabilities arose in biological organisms is not known. There are many intermediate cases between the simplest uses of grippers and the competences of human engineers. We may not be able to understand the latter without understanding more about

---

[7] One of the forms of "representational redescription" discussed in [11] is the transition from having a competence to being able to articulate its features.

the intermediate capabilities on which they depend.

Motivational and deliberative competences Organisms have changing needs that influence behaviours. Some changes directly trigger reactions that can reverse, or make use of the change: for instance shivering can be triggered by mechanisms detecting a drop in temperature. Evolution discovers some conditions under which such "reactive" responses are beneficial, and encodes genetic information producing the mechanisms in new individuals. But evolving reactions to needs can be very slow. It can take many generations for arrival of a new danger or a new form of food making a new response useful to lead to evolved behavioural reactions. Instead, between the mechanisms that detect needs and the mechanisms that produce behaviours, evolution interposed mechanisms that select goals triggered by detected needs, which in turn trigger planning mechanisms to select actions to achieve the goals [15]. Much AI research has been concerned with ways of achieving this. From a biological standpoint, the use of such mechanisms provides opportunities for novel evolutionary or development processes concerned with (a) selecting new goals, (b) finding plans for achieving them and (c) using plans to control actions. Many variants of these patterns are relevant to the meta-morphogenesis project. A type of evolution that generates new kinds of rewards is described in [16]. Another possibility is adding mechanisms that generate goals not because they will satisfy some need or provide some reward, but merely because there are currently no important tasks in progress, and an opportunity for generating a certain sort of goal has been detected. In [24] it is argued that reflex triggering of such goals along with mechanisms for achieving goals, will sometimes cause useful new things to be learnt, even if achieving the goal has no reward value. Failing to achieve goals often provides more valuable learning than succeeding.

Factorisation of the link between needs and actions introduces modularity of design, allowing opportunities for separate types of improvement, with benefits shared between different needs – perhaps permitting evolution and/or learning to be speeded up through sharing of benefits.

"Peep-hole" vs "Multi-window" perception and action. Although it would take up too much space to explain fully here, there is a distinction between architectures in which there is limited processing of perceptual input and the results of the processing are transmitted to various "more central" mechanisms (e.g. goal formation, or planning subsystems), which I call "peep-hole" perception, and architectures using "multi-window" perception in which perceptual subsystems do several layers of processing at different levels of abstraction in parallel, using close collaboration with the layers and with more central mechanisms (e.g. parsing, searching for known structures, interpreting). Multi window perceptual processing is crudely illustrated in this figure http://tinyurl.com/BhamCog//crp/fig9.6.gif Likewise a distinction can be made between peep-hole and multi-window *action control* subsystems. For example a multi-window action could include, in football, concurrently running towards a goal, dribbling the ball, getting into position to shoot, avoiding a defender and eventually shooting at the goal. Linguistic production, whether spoken, handwritten, or signed always has multiple levels of processing and reference. (Compare Anscombe's analysis of intention in [1].)

The use of multi-window perception and action allows a wider range of information processing at different levels of abstraction to be done concurrently with sensory inputs and motor outputs, permitting more powerful and effective perception and action subsystems to evolve or be developed. I conjecture that the multi-window solutions are used by far more species than have been noticed by researchers, and are also well developed in pre-verbal human children, though yet more development occurs later.

Transitions in representational requirements. Even in this overview of a tiny subset of evolutionary processes we find requirements for different information structures: binary on/off structures in a detector, scalar values varying over time used in homeostatic and "hill-climbing" control processes, information about spatial and topological relationships between surfaces and regions that are not currently being sensed, that are needed for planning routes, and information about possibilities for change, constraints on change, and consequences of possible changes, needed for selecting and controlling actions manipulating physical structures, along with use of meta-semantic information about information users and information-bearing structures. These requirements are related to old philosophical problems, e.g. How is information about possibilities and impossibilities be represented? Can young children, or non-human animals, make use of modal logics, and if not what are the alternatives?

Often it is not obvious how a particular type of information will be most usefully represented for a particular type of organism. Many researchers, whether studying animal cognition or attempting to design intelligent robots, assume that the representation of spatial structures and relationships must use something like global 3-D coordinate systems, forgetting that such forms of representation were a relatively late discovery in human culture. Humans made tools, machines, houses, temples, pyramids, aqueducts and other things requiring a deep understanding of spatial structures and processes before geometry had been arithmetized by Descartes, so it is possible that they were using some other form of representation.

Re-representation and systematisation. The main motive that originally got me into AI was the hope of showing that Immanuel Kant's theories about the nature of mathematical knowledge [10], were superior to the opinions of most other philosophers, including Hume, Mill, Russell, and Wittgenstein. I hoped to show this by building a robot that started off, like infants and toddlers discovering things about spatial structures and motions empirically and later finding ways of reorganising some of the information acquired into theories that allowed it to *prove* things instead of discovering them empirically, e.g. using diagrammatic proofs of the sort used in Euclidean geometry [23]. This task proved far more difficult than I initially hoped, in part because of the great difficulty of giving robots animal-like abilities to perceive, understand, and use information about structures and motions in the environment, in order to predict or explain their behaviours, as suggested by Craik [6]. Perhaps something like the processes Karmiloff-Smith labelled varieties of "Representational Redescription" [11], are needed, though there's more than re-description going on, since architectural changes are also required. I suspect these mathematical competences in humans build on precursors found not only in pre-verbal children, but also in other animals with powerful spatial reasoning capabilities required for using complex affordances, as in such as some nest-building birds.[8] This remains an important task for the Meta-morphogenesis project, which may enhance research in AI and psychology on learning and creativity.

Empirical learning *vs* working things out Many forms of learning investigated in AI, robotics and psychology make use of mechanisms for deriving taxonomies and empirical generalisations

---

[8] See also http://tinyurl.com/BhamCog/talks/#toddler

from collections of examples. The evidence used may come from the experiences of an individual (animal or robot) exploring an environment, finding out what can and cannot be done in it, and what the consequences are, or they may make use of data-mining techniques applied to much larger externally supplied sample sets.

Humans, and many other species, are clearly capable of discovering useful empirically supported patterns, for example linking actions, circumstances and consequences. However, human mathematical knowledge shows that humans are also capable of a different kind of learning – by *working things out*. Collecting empirical generalisations may eventually trigger a switch to another process, which instead of merely using more data to extend known generalisations, takes what is already known and attempts to find a "generative basis" for it. A special case is switching from pattern-based language use to syntax-based language use, a common transition in child development. Syntax-based competences use generative rules and compositional semantics that allow new, richer forms of communication, and also new richer forms of thinking and reasoning – one type of "representational redescription".

I conjecture that the linguistic case is a special development of a more general biological capability, that evolved earlier and in more species, which allows a collection of useful empirical generalisations to be replaced by something more economical and more powerful: a generative specification of the domain. The creation of Euclid's elements appears to have been the result of a collective process of this sort, but that collective cultural process could not have happened without the individual discoveries of new more powerful generative representations of information previously acquired empirically piecemeal [27].

In simple cases the new generative (e.g. axiomatic) representation may be discovered by data-mining processes. However in the more interesting cases it is not sufficient to look for patterns in the observed cases. Instead it is necessary to *extend the ontology used*, so as to include postulated entities that have not been experienced but are invoked as part of the process of explaining the cases that have been experienced. The infinitely small points and infinitely thin, straight and long lines, of Euclidean geometry are examples of such ontological extension required to create a system with greater generative power. This process of reorganisation of knowledge into a new, more powerful, generative form, seems to be closely related to the hypothesis in [6] that some animals can create models that they use to predict the results of novel actions, instead of having to learn empirically which ones work and which ones don't, possibly with fatal costs. The ability of human scientists to come up with new theories that explain old observations, making use of ontological extensions that refer to unobservable entities (e.g. atoms, sub-atomic particles, valences, gravity, genes, and many more) also illustrates this kind of process replacing empirical generalisations with a generative theory.

I suspect that similar transformations that have mostly gone unnoticed also occur in young human children, discovering what could be called "toddler theorems". (See http://tinyurl.com/TodTh) Such transformations could occur, both in humans and some other species, without individuals being aware of what has happened – like children unaware that their linguistic knowledge has been reorganised. Later, as meta-semantic competences develop, individuals may come to realise that they have different kinds of knowledge, some of it empirical, derived from experience, and some generated by a theory. Later still, individuals may attempt to make that new knowledge explicit in the form of a communicable theory

These conjectures about different bases for knowledge about the world are closely related to the main ideas of [11], but came from a very different research programme based on the idea of using AI techniques to solve problems in philosophy of mathematics [23]. I suspect this is closely related to Kant's theories about the nature of mathematical knowledge [10]. Such discoveries are very different in kind from the statistics-based forms of learning (e.g. Bayesian learning) that now dominate much research. The mathematical reasoning shows what *can* be or *must* be the case (given certain assumptions) not what is highly probable: e.g. working out that the angles of a triangle must add up to a straight line, or that 13 identical cubes cannot be arranged in rectangular array other than a 13x1 array, is very different from finding that stones thrown up normally come down: the latter discovery involved no mathematical necessity (until Newtonian mechanics was developed). At present I don't think there are any good theories about either the biological basis of such knowledge or how to provide it for robots.

Enduring particulars For many species the only environmental information relevant to control decisions is information about the *types* of entity in the immediate environment. E.g. is this a place that provides shelter or food? Is that a dangerous predator? Is this conspecific friendly or aggressive? For a variety of different reasons it became useful to be able to re-identify particular individuals, places, and objects at different times (e.g. is this the tool I have already tested, or do I need to test it before using it?). However, as philosophers have noted there are enormous complications regarding tracking individuals across space and time (e.g. is it the same river after the water has been replenished; is this adult the same individual as that remembered child?). This is not the place to go into details (compare [32]), but analysis of the many types of particular and the means of referring to or re-identifying them and the purposes that can serve, can give clues regarding evolutionary and developmental transitions that have so far not been studied empirically and also have not been addressed in robot projects except in a piecemeal, *ad hoc* fashion, with much brittleness.

Meta-management. As information-based controlling processes become more complex, across evolutionary or developmental time-scales, the need arises for them also to be controlled, in ways that can depend on a variety of factors, including the changing needs of individual organisms, their bodily structure, the types of sensorymotor systems they have, their developing competences, and the constraints and affordances encountered in their environments, some of which will depend on other organisms. New forms of control of controlling process are also examples of meta-morphogenesis.

Evolving new mechanisms for turning on each new kind of functionality, without harmfully disrupting other functions, is less useful than using a pre-existing, extendable, mechanism for handing control from one subsystem to another.[9] This can also support centralisation of major decisions, to ensure that all relevant available information is taken into account, instead of simply allowing strongly activated sub-systems to usurp control. Using scalar strength measures, like scalar evaluation functions in search, loses too much information relevant to comparing alternatives.

"Hard-wired", implicit control mechanisms, implemented using only direct links between and within sub-systems, can be replaced by newly evolved or developed *separate and explicit* control functions (e.g. selecting what to do next, how to do it, monitoring progress, evaluating progress, using unexpected information to re-evaluate priorities, etc., as in the meta-management functions described in [3, 35]). Such new control regimes may allow new kinds of functionality

---

[9] Compare the invention of a procedure call stack for computing systems.

to be added more simply and used when relevant, thereby expanding the opportunities (affordances) for evolution and learning.

From internal languages to communicative languages For some people languages are by definition a means of intentional communication between whole agents. But that ignores the vast amount and variety of types of *internal* information processing using structured forms of representation of varying complexity with compositional semantics e.g. to encode learnt generalisations, perception of complex structures, intentions to perform complex actions, questions, predictions, explanations, and plans – in both non-human animals and pre-verbal children. Philosophers and psychologists who have never thought about how to design a working animal usually never notice the requirements. As argued in [18, 19, 22], there is a natural correspondence between the contents of internal plans and behaviours controlled by the plans. I suggest that a series of evolutionary transitions allowed actions to become communications, initially involuntarily, then later voluntarily, then enhanced to facilitate communication (e.g. for cooperation) and then, using the duplicate and differentiate evolutionary strategy) "hived off" as a means of communication, which evolved into sophisticated sign languages. Later additional requirements (communication at night, and while using hands) might have led to evolution of vocal accompaniments that finally became spoken language. This conjecture has deep implications regarding structures of human and animal brains and minds that need to be explored as part of this project. Further variations in functions and mechanisms both across generations, between contemporary individuals, and between stages of development within an individual would include:
– genetically specified forms of communication (possibly specified in a generic way that can be instantiated differently by different individuals or groups).
– involuntary vs intentional forms of communication. It seems unlikely that the "begging" for food actions of fledglings and young mammals are intentional (in various meanings of that word). In other cases there are different kinds of intentionality and different levels of self-awareness when communication happens.
– other variations include whether there is explicit teaching of means of communication by older individuals (Compare [11])

Varieties of meta-morphogenesis Some examples of evolutionary meta-morphogenesis seem to be restricted to humans. We have a collection of mechanisms (closely related to some of the themes in [11]) that allow humans (a) to acquire novel capabilities by various processes of learning and exploration, including trial and error, (b) to become aware that we have acquired such a new competence or knowledge, (c) find a way to express its content, (d) decide to help someone else (e.g. offspring or members of the same social group) to acquire the competence – through a mixture of demonstrations, verbal explanations, criticisms of incomplete understanding and suggestions for improvement, and (d) to provide cultural artefacts for disseminating the knowledge.

Some previous results of information-processing morphogenesis can alter current processes of morphogenesis, for instance when learning extends abilities to learn, or evolution extends evolvability, or evolution changes abilities to learn, or new learning abilities support new evolutionary processes. Where morphogenesis produces new types of learning or development and new sorts of evolvability, that can be labelled "meta-morphogenesis". A deep explanatory theory will need to characterise the "evolutionary affordances" (generalising Gibson's notion [8]) made use of. In particular, evolved cognitive abilities may provide new affordance detectors, such as mate-selectors, accelerating evolution as agricultural breeding has done. Evolution starts off blind, but can produce new affordance detectors that influence subsequent evolution.

If every new development opens up N new possibilities for development the set of possible trajectories grows exponentially, though only a subset will actually be realised. Nevertheless, the cumulative effects of successive phases of meta-morphogenesis seems to have produced enormous diversity of physical forms, behaviours, and less obviously, types of biological information processing (including many forms of learning, perceiving, wanting, deciding, reasoning, and acting intentionally) making evolution the most creative process on our planet. The diversity may be essential for evolution of (e.g.) mathematicians, scientists, and engineers.

3 Conclusion

I have tried to present a variety of transitions in kinds of information processing that seem to have occurred in the evolutionary history of humans and other species. This is merely a taster, which may tempt more researchers to join the attempt to build a systematic overview of varieties of ways in which information processing changed during biological evolution, with a view to implementing the ideas in future computational experiments. This will require much computationally-guided empirical research seeking information about social, developmental, epigenetic, genetic and environmental transitions and their interactions.

In his 1952 paper Turing showed how, in principle, sub-microscopic molecular processes in a developing organism might produce striking large scale features of the morphology of a fully grown plant or animal. This is a claim that if individual growth occurs in a physical universe whose building blocks permit certain sorts of spatio-temporal rearrangements, complex and varied structures can be produced as a consequence of relatively simple processes.

Darwin proposed that variations in structures and behaviours of individual organisms produced by small random changes in the materials used for reproduction could be accumulated over many generations by mechanisms of natural selection so as to produce striking large scale differences of form and behaviour. This is a claim that if the physical universe supports building blocks and mechanisms that can be used by reproductive processes, then the observed enormous diversity of forms of life can be produced by a common process.

Partly inspired by Turing's 1952 paper on morphogenesis, I have tried to show that there are probably more biological mechanisms that produce changes in forms of information processing than have hitherto been studied, in part because the richness of biological information processing has not been investigated as a topic in its own right, though some small steps in this direction were taken by [9], and others. Moreover it seems that the collection of such mechanisms is not fixed: there are mechanisms for producing new morphogenesis mechanisms. These can be labelled meta-morphogenesis mechanisms. The cross-disciplinary study of meta-morphogenesis in biological information processing systems promises to be rich and deep, and may also give important clues as to gaps in current AI research.

Can it all be done using computers as we know them now? We need open minds on this. We may find that some of the mechanisms required cannot be implemented using conventional computers. It may be turn out that some of the mechanisms found only in animal brains are required for some of the types of meta-morphogenesis. After all, long before there were neural systems and computers, there were chemical information processing systems; and even in modern organisms the actual construction of a brain does not (in the early stages) use a brain but is controlled by chemical processes in the embryo.

Biological evolution depends on far more than just the simple idea of natural selection proposed by Darwin. As organisms became more complex several different kinds of mechanism arose that are able to produce changes that are not possible with the bare minimum mechanisms of natural selection, although they depend on that bare minimum. This is not a new idea. A well known example is the use of cognition in adults to influence breeding, for instance by mate selection and selective feeding and nurturing of offspring when food is scarce. My suggestion is that we need a massive effort focusing specifically on examples of *transitions in information processing* to accelerate our understanding.

There must be many more important transitions in types of biological information processing than we have so far noticed. Investigating them will require multi-disciplinary collaboration, including experimental tests of the ideas by attempting to build new machines that use the proposed mechanisms. In the process, we'll learn more about the creativity of biological evolution, and perhaps also learn how to enhance the creativity of human designed systems. This research will be essential if we are to complete the Human Genome project.[10][11]

ACKNOWLEDGEMENTS

I am grateful to Alison Sloman for useful discussions of evolution and to Stuart Wray for comments on an early draft of this paper and his diagram summarising the draft.[12] The EU CogX project helped to shape my thinking.

REFERENCES

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

[2] I. Apperly, *Mindreaders: The Cognitive Basis of "Theory of Mind"*, Psychology Press, London, 2010.

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

[4] N. Block. What is functionalism?, 1996. (Originally in The Encyclopedia of Philosophy Supplement, Macmillan, 1996).

[5] J. Chappell and A. Sloman, 'Natural and artificial meta-configured altricial information-processing systems', *International Journal of Unconventional Computing*, **3**(3), 211–239, (2007).

[6] K. Craik, *The Nature of Explanation*, Cambridge University Press, London, New York, 1943.

[7] D.C. Dennett, *The Intentional Stance*, MIT Press, Cambridge, MA, 1987.

[8] J.J. Gibson, *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston, MA, 1979.

[9] E. Jablonka and M. J. Lamb, *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*, MIT Press, Cambridge MA, 2005.

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

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

[12] J. Maynard Smith and E. Szathmáry, *The Major Transitions in Evolution*, Oxford University Press, Oxford, England:, 1995.

[13] John McCarthy, 'From Here to Human-Level AI', volume 171, pp. 1174–1182. Elsevier, (2007). doi:10.1016/j.artint.2007.10.009, originally KR96.

[14] M. L. Minsky, 'Steps towards artificial intelligence', in *Computers and Thought*, eds., E.A. Feigenbaum and J. Feldman, 406–450, McGraw-Hill, New York, (1963).

[15] M. Scheutz and B.S. Logan, 'Affective vs. deliberative agent control', in *Proceedings Symposium on Emotion, cognition and affective computing AISB01 Convention*, ed., *et al.* C. Johnson, York, (2001).

[16] S. Singh, R.L. Lewis, and A.G. Barto, 'Where Do Rewards Come From?', in *Proceedings of the 31th Annual Conference of the Cognitive Science Society*, eds., N.A. Taatgen and H. van Rijn, pp. 2601–2606, Austin, TX, USA, (2009). Cognitive Science Society.

[17] A. Sloman, 'Interactions between philosophy and AI: The role of intuition and non-logical reasoning in intelligence', in *Proc 2nd IJCAI*, pp. 209–226, London, (1971). William Kaufmann.

[18] A. Sloman, 'What About Their Internal Languages?', *Behavioral and Brain Sciences*, **1**(4), 515, (1978).

[19] A. Sloman, 'The primacy of non-communicative language', in *The analysis of Meaning: Informatics 5, ASLIB/BCS Conf, Oxford, March 1979*, eds., M. MacCafferty and K. Gray, pp. 1–15, London, (1979). Aslib.

[20] A. Sloman, 'Actual possibilities', in *Principles of Knowledge Representation and Reasoning: Proc. 5th Int. Conf. (KR '96)*, eds., L.C. Aiello and S.C. Shapiro, pp. 627–638, Boston, MA, (1996). Morgan Kaufmann Publishers.

[21] A. Sloman. Why symbol-grounding is both impossible and unnecessary, and why theory-tethering is more powerful anyway., 2007.

[22] A. Sloman. Evolution of minds and languages. What evolved first and develops first in children: Languages for communicating, or languages for thinking (Generalised Languages: GLs)?, 2008.

[23] A. Sloman, 'The Well-Designed Young Mathematician', *Artificial Intelligence*, **172**(18), 2015–2034, (2008).

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

[25] A. Sloman, 'Some Requirements for Human-like Robots: Why the recent over-emphasis on embodiment has held up progress', in *Creating Brain-like Intelligence*, eds., B. Sendhoff, E. Koerner, O. Sporns, H. Ritter, and K. Doya, 248–277, Springer-Verlag, Berlin, (2009).

[26] A. Sloman, 'How Virtual Machinery Can Bridge the "Explanatory Gap", In Natural and Artificial Systems', in *Proceedings SAB 2010, LNAI 6226*, eds., S. Doncieux and et al., pp. 13–24, Heidelberg, (August 2010). Springer.

[27] A. Sloman, 'If Learning Maths Requires a Teacher, Where did the First Teachers Come From?', in *Proc. Int. Symp. on Mathematical Practice and Cognition, AISB 2010 Convention*, eds., Alison Pease, Markus Guhe, and Alan Smaill, pp. 30–39, De Montfort University, Leicester, (2010).

[28] A. Sloman, 'Phenomenal and Access Consciousness and the "Hard" Problem: A View from the Designer Stance', *Int. J. Of Machine Consciousness*, **2**(1), 117–169, (2010).

[29] A. Sloman, 'What's information, for an organism or intelligent machine? How can a machine or organism mean?', in *Information and Computation*, eds., G. Dodig-Crnkovic and M. Burgin, 393–438, World Scientific, New Jersey, (2011).

[30] A. Sloman. What's vision for, and how does it work? From Marr (and earlier) to Gibson and Beyond, Sep 2011.

[31] A. Sloman and J. Chappell, 'The Altricial-Precocial Spectrum for Robots', in *Proceedings IJCAI'05*, pp. 1187–1192, Edinburgh, (2005). IJCAI.

[32] P. F. Strawson, *Individuals: An essay in descriptive metaphysics*, Methuen, London, 1959.

[33] A. M. Turing, 'The Chemical Basis Of Morphogenesis', *Phil. Trans. R. Soc. London B 237*, **237**, 37–72, (1952).

[34] A A S Weir, J Chappell, and A Kacelnik, 'Shaping of hooks in New Caledonian crows', *Science*, **297**, 981, (2002).

[35] I.P. Wright, *Emotional agents*, Ph.D. dissertation, School of Computer Science, The University of Birmingham, 1977. http://www.cs.bham.ac.uk/research/cogaff/.

[36] L. A. Zadeh, 'A New Direction in AI: Toward a Computational Theory of Perceptions', *AI Magazine*, **22**(1), 73–84, (2001).

---

[10] Instead of regarding evolution as a "blind watchmaker", we can think of it as a blind theorem prover, unwittingly finding proofs of "theorems" about what sorts of information-using systems are possible in a physical world. The proofs are evolutionary and developmental trajectories. The transitions discussed here can be regarded as powerful inference rules.

[11] The concept of "information" used here is not Shannon's (purely syntactic) notion but the much older notion of "semantic content", explained more fully in [29] http://tinyurl.com/BhamCog/09.html#905

[12] See: http://www.cs.bham.ac.uk/˜axs/fig/wray-m-m-label-small.jpg