Dynamic metaphysical grounding of consciousness in evolution[*]

Aaron Sloman
http://www.cs.bham.ac.uk/~axs

---

[*] (Not to be confused with "symbol-grounding" theories discussed by Harnad and others, reviving the old philosophical theory of concept empiricism, refuted by Kant in 1781, as discussed in http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#models)

[Note]
This paper makes use of Alastair Wilson's ideas about grounding as "metaphysical causation" [Wilson 2015]. However, he has not read this and cannot be held responsible for any errors or confusions.

This is work in progress. Details may change in later versions.

Introduction
Interplay between metaphysical and physical structures and processes

There is a view, at least as old as Plato, of metaphysics as a domain of timeless truths about entities that do not exist in space and time. According to that view of metaphysics if A and B are metaphysical entities (or types of entity) then if A is a ground of B then that must be a timeless relation between timeless entities.

An alternative view is that besides the physical changes that occur in the universe, or in a region of the universe, there are also metaphysical changes that produce (i.e. ground?) new *kinds* of existent.

A philosopher who believes that existence of timeless metaphysical entities cannot depend on the existence of physical structures or processes would deny that physical processes or events could produce new examples, or new types of examples, of such metaphysical entities.

Attribution of timeless metaphysical reality seems to be plausible as regards existence of certain *properties*, e.g. roundness, triangularity; *relations*, such as containment, being straighter; numbers, or mathematical *truths*, e.g. containment is transitive and the sum of three and five is eight, or the division of formal proofs into valid and invalid proofs. The existence of those properties, relations and truths cannot be affected by which physical objects actually exist or which physical events, processes, or states of affairs exist or have occurred. Additional candidates for timeless existence would be propositions themselves, whether true or false, such as the proposition that three fives make thirty five or the proposition that elephants enjoy catching and eating turkeys. I shall ignore most of those cases in what follows. The process or state of such a proposition being considered or expressed or believed by some individual would not be timeless.

It can also be argued that all the **types** of minds and all the **types** of mental phenomena that ever have existed or will exist or could exist, must have existed continuously as *abstract types* throughout the history of the universe, even when they had no instances. On this view the **types** of mental phenomena existed even when no arrangements of physical matter existed that supported **instances** of those types. The types exist independently of which physical processes have or have not occurred. For example, a type of thought being considered by a type of person is a type of mental event that exists

independently of any actual persons or thinking processes, just as the relation of containment or the property of being elliptical exists whether or not anything contains anything else or anything is elliptical.

In contrast, timelessness is not a feature of **instances** of various metaphysical (or at least non-physical) entity involved in the existence of minds, including thoughts, sensory contents (qualia, or sense-data) desires, beliefs, intentions, and mental processes such as noticing, learning, reasoning, wanting, enjoying, making a mathematical discovery, planning a route, forgetting, trying to remember or understand, etc. Your wanting something, planning a way to get it, and later trying to remember your conclusion are all mental occurrences with temporal bounds, and with causal links to spatially bounded parts of your environment.

Whether or not the types of which they are instances could be said to exist timelessly, the instances, such as a particular human being wanting something, could not have existed in the solar system when it first formed, since the conditions required for human life were not satisfied then, and presumably no such thing could exist at the centre of the sun as it is now, since the conditions are still not satisfied there. The fact that instances of those types can exist and do exist on this planet now is a consequence of the variety and creativity of the mechanisms and processes of biological evolution and the ways they have operated over several billion years.

This paper is about what makes such instances possible, e.g. what grounds the possibility of your having philosophical or mathematical thoughts now which you could not have had as a new-born infant, and what grounds the possibility of a species whose members can make and discuss the sorts of geometrical discoveries reported in [Euclid's Elements], which could not have been made by older species.

I do not believe it is possible to produce any deep philosophical theory about what grounds the corresponding *timeless* truths without basing it on a theory about what needed to change between the times when mathematical minds could not have existed on Earth and the earliest times when mathematical discovery processes became possible. What made them possible was a *dynamic* grounding process, whose general form is discussed in [Wilson 2015].

Many of the oldest metaphysical problems, e.g. problems about the nature of and relations between mind and matter are directly related to products of evolution. When the earliest forms of matter first came into existence there were no minds, nor mental contents, nor the sorts of vehicles for mental contents provided by biological information-processing mechanisms such as nerves and brains. Neither were there opportunities, risks, benefits and dangers for individual organisms, or mechanisms for detecting and avoiding the dangers and achieving the benefits.

Yet now Earth swarms with minds of many kinds and degrees of complexity, and their products -- resisting any classification as good or bad. Moreover, new kinds of minds became possible on Earth at different stages of biological evolution, and later on at different stages of cultural and technological change. It is still impossible for instances of any of those types of life to exist at the centre of the sun. So what is grounded in one part of the universe need not have been grounded there at earlier times, and may still not be grounded in another location. So any theory of *static* grounding of the mental by the physical will not fit the facts of our universe: we need a theory of *dynamic*, *situation-relative*, grounding that explains how various physical and evolutionary changes made new forms of life possible.

In particular, it is an essential feature of all life forms that they not only include physical and chemical structures and processes but also make use of *information* in controlling what they do. (Ganti's "chemoton" was proposed as a minimal design known to be capable of being implemented in the physics and chemistry of this universe [Ganti 1971/2003], [Fernando 2008].)

Organisms need to make use of changing information about current needs, dangers, and possible benefits as well as factual information about what exists, what is possible, and what the consequences of realizing certain possibilities will be. Use of information can be based on many types of information-processing *mechanisms* each of which will typically work in certain conditions but not others. So different mechanisms will be useful (or required) in different physical locations and different evolutionary lineages.

The contribution of information processing mechanisms to evolution does not depend on the use of *optimal* decision making, as sometimes assumed, as long as the mechanisms work well enough to support continued reproduction of species using them, by supporting continued life, growth, and reproductive functions of sufficient numbers of individuals of those species. What suffices for a particular species in a particular geogrphical location can change over time, as shown by geological records of extinct species that once flourished.

To contrast with notions of optimality, meeting such *sufficient* requirements, using "bounded rationality", was labelled "satisficing" by Herbert Simon (https://en.wikipedia.org/wiki/Satisficing). Many philosophers, psychologists, economists and biologists attempt to base their theories on a more restricted notion of rationality based on a notion of "optimality", which in turn requires a measure of *utility* or *value*, as discussed, for example in [Russell(2013)]. Both the existence and the need for such a measure are debatable, both for species and for individuals [Sloman, 2009]. (Similar criticisms can be made of widely assumed requirements for a probability metric.)

With these details and constraints in mind we can ask: What features of certain parts of the universe can ground the possibility of certain sorts of minds and mental contents at those locations? Different answers will be relevant to different places at different times: an uncomfortable fact for philosophers seeking "the one correct theory" of how physical matter can ground life, or mental phenomena.

Minds are complex entities. They include things like memories, experiences, desires, concepts, knowledge, memories, skills, mistaken beliefs, puzzlement, wonderment, joy, disappointment, grief, longing, drifting attention, excited anticipation, dreams, mathematical insight, mental arithmetic, and theory creation. So the coming into existence of such a mind involves creation of new instances of many such metaphysical types. How all that is possible is a (multi-faceted) question whose (multi-faceted) answer involves reference to metaphysical causation: what sorts of things can cause new instances of metaphysical types to come into existence, possibly for the first time?

More importantly, for the discussion of grounding, what physical changes make the existence of those new types of entity possible, if they are possible at a certain time, but were in some sense not possible earlier, e.g. when the planet first formed, before there was any form of life here? During biological evolution physical changes produce metaphysical changes. How is that possible?

In part, answering that question is a scientific challenge obviously linking physics, chemistry, biology, psychology and the socio-economic sciences. Less obviously what we have learnt about the science of information processing systems is also relevant.

We do not have complete detailed answers yet (some gaps will be indicated below), but recent advances in our knowledge allow us to give new partial answers that build on achievements in computer science, software engineering and artificial intelligence in the last 60 years or so. Such a claim seems obviously false to some philosophers, e.g. John Searle in this lecture at google https://www.youtube.com/watch?v=rHKwIYsPXLg. However such objectors seem to be unaware of all the entirely new things we have learnt from our engineering successes and failures, about varieties of and properties of man-made information-processing systems. I'll try to show how these successes and failures provide clues directly relevant to metaphysical questions about the nature of mental contents and their relationships to physical machinery. However, gaps in our knowledge remain, mentioned below.

Most of what has been learnt about information processing systems seems to have been ignored by most philosophers, many of whom think that if they know about Turing machines they understand everything there is to know about computation. That view is often based on scientific and technological ignorance (including a misunderstanding of the meaning of Turing Universality Sloman(1996)), and a failure to reflect on requirements for functionality of the technology now widely used by a steadily increasing subset of the human population, including email, online banking, online shopping, social networks, distributed video games, online tutorials, automatic aircraft landing systems, computer driven cars, and many more. These facilities could not be provided by Turing machines alone. They depend on multiple virtual machines running on networked, unsynchronised, physical computers, interacting concurrently with one another and with many different parts of their environment, including physical machinery, weather systems, human minds and socio-economic systems. Debugging such a system when it doesn't behave as expected can sometimes be more like understanding and correcting a student's error than like replacing a fuse. This often requires special tools that give access to the machine's beliefs, inferences, priorities (i.e. its mind) rather physical recordings of its hardware or a program code listing.

Such a deeply embedded semantically rich virtual machine would be nothing like a Turing machine. An isolated fragment of such a network of systems, removed from all its normal connections could be equivalent to a Turing machine, but it would no longer be performing its normal functions. It is possible to view an isolated Turing machine as a purely syntactic engine, but the computers we use every day would not work without rich semantic capabilities -- including abilities to refer to parts of their memories, to the contents of their memories, to instructions to be performed, to the results of conditional tests, to other computers, to physical interfaces to the environment, and to other virtual machines and their contents (e.g. remote email handlers or flight reservation systems) and to physical objects, states and processes in their environment.

This semantic richness has been ignored by philosophers like John Searle for several decades, as demonstrated in the google presentation mentioned above. (I expect he has never tried to debug a complex running virtual machine. Sometimes sheer intelligence is not enough for good philosophy.) Some of the arguments against artificial intelligence are analogous to attempts to prove that brains cannot provide intelligence because they are made of atoms and an atom cannot be intelligent. Or replace "atom" with "neuron".

What we have learnt in the last half century about varieties and uses of virtual machinery able to support a wide variety of types of mental phenomena enables us to ask new questions about the processes and products of biological evolution, including questions that Darwin and most of his contemporaries (friends and foes) had not thought of and could not have thought of, though there is some evidence that at least Ada Lovelace, Babbage's assistant, had begun to think about them a century before Turing [Sloman (2002)].

These new ideas have not yet been, and will not easily be, absorbed into neuroscience without radical new ideas about brain functions and the supporting physical/chemical mechanisms, for example, ideas about how brains can acquire and use not only information about current nearby and remote states of affairs, but also information about what is possible and what is impossible. (For fairly simple examples of such requirements see [Sloman (impossible)].) It will also be necessary to explain how brains can acquire, hypothesize and use meta-cognitive information about their own contents, including information about gaps in knowledge, and also meta-cognitive information about contents of other individual minds, including their intentions, problems, knowledge-gaps, preferences, feelings, etc., as well as developing theories about minds of other sorts, e.g. dog minds, toddler minds, struggling student minds, and the huge variety of minds portrayed in stories and plays.

Such ideas will not be absorbed effectively into academic cognitive psychology until the Popper-inspired fetish with falsifiability and the over-zealous demands for publications based on statistical analysis of data have been tempered or abandoned, and research enriched by appreciation of deep explanations of complex sets of capabilities that cannot be summarised in statistics gained from behavioural experiments. Compare how the depth of chemical theories based on structures (some of them summarised crudely in the periodic table of the elements) and structural interactions of molecules and their parts contrasts with summaries of statistics collected from chemistry experiments. See also [Sloman, 2014, in progress]. For an example early draft theory of how certain kinds of metacognition (self-directed or other directed) might be implemented in working mechanisms see John Barnden's ATT-Meta project, summarised here: https://www.cs.bham.ac.uk/~jab/ATT-Meta/

There's a lot more to be said about the achievements of biological evolution and its metaphysical creativity, but first a few comments on grounding and explanation.

Explaining how minds are possible

The role of construction kits in evolution

The suggestion in [Wilson 2015] that grounding is metaphysical causation ("G=MC") is at least consistent with, and more likely supports, my claim that the study of evolution of life, consciousness, intelligence, etc. is an activity in which science, engineering, and metaphysics overlap. In particular, they need to be combined to answer questions of the form "How are entities (states, events, processes, causal interactions, capabilities) of type X possible?" -- a type of question I first encountered in Kant's writings, many years ago, e.g. in his inspiring but incomplete attempt to explain how synthetic apriori knowledge of geometrical and arithmetical facts is possible [Kant, 1781]. (A topic I'll return to.)

It was argued in Sloman(1978)[Ch 2] (direct link) that, contrary to the views of Popper and others, questions about what sorts of things are possible and what makes them possible are among the most important for science, even though statements about what is possible are not empirically falsifiable, and therefore, at least according to Popper(1934), belong to metaphysics, not science.

Popper partly changed his mind later, especially as regards Darwin's theory of evolution, which he came to regard as a major contribution to science, despite its unfalsifiability [Popper, 1977], though he sometimes resisted describing it as such. In [Popper, 1976] he wrote (p.168) *"I have come to the conclusion that Darwinism is not a testable scientific theory, but a metaphysical research programme--a possible framework for testable scientific theories".*

Anyhow, from the standpoint of this discussion, Darwin's theory of evolution by natural selection is an example of a scientific attempt to explain a collection of *prima-facie* surprising possibilities. Theories in Linguistics, whose adequacy (observational, descriptive, and explanatory adequacy) were

discussed in Chomsky, 1965 and theories in Artificial Intelligence can also provide scientific theories explaining collections of possibilities, though many things are still unexplained. The astounding successes of AI so far (e.g. in search engines, playing GO, automated theorem proving, planning and scheduling systems, and many other applications) conceal huge gaps in its achievements, for example the enormous differences between AI vision systems and biological visual capabilities, the abilities of humans and other animals to bootstrap sophisticated capabilities (including creation of languages) during development, and in particular the inability of current AI reasoning systems to make the sorts of mathematical discoveries that were made by ancient mathematicians like Euclid and Archimedes, some of which seem to be partly replicated by pre-verbal toddlers [Sloman (impossible)] (section on "Toddler topology").

Immediately after our planet formed there was an enormously rich space of **possibilities** for further development, vastly more rich than the variety any human has conceived of. This depended on the existence (then and now), as part of the constitution of the physical universe, what could be called the Fundamental Construction Kit (FCK) whose features must have had the potential to explain all the possibilities that have since been realised, including possibilities for myriad forms of life and all known forms of intelligence. All those possibilities are *grounded* in the FCK. There will always be more possibilities explained in principle, or potentially grounded, by the FCK than have actually been realised/grounded.

As a result of processes of evolution there is also a (largely unnoticed, but huge) variety of *Derived* Construction Kits (DCKs) all based on the FCK, but each capable of supporting or facilitating the realisation of some subset of the possibilities -- for example construction kits supporting evolution of plants that grow upwards out of soil, construction kits for building various sorts of digestive system, construction kits for various kinds of brain mechanism, construction kits supporting evolution of abilities to acquire and use information about extended terrain, and (possibly overlapping) construction kits involved in evolution of minds like Euclid's. (For a more detailed discussion of fundamental and derived, concrete, abstract and hybrid, construction kits and the possibilities they explain/ground see [Sloman (Kits)].) I'll give examples below of phenomena that have been claimed by philosophers to be incapable of being grounded physically, and show (in outline) how they could be grounded as a result of use of construction kits for building increasingly complex kinds of information-processing mechanisms. Examples include "first-person perspectives", and, more generally, *qualia*. The importance of mathematical qualia (as used in ancient mathemetical discovery processes) has gone largely unnoticed in this context though they are closely related to the perceptual affordances discussed by [Gibson 1979].

It is sometimes claimed that selection processes suffice to explain all products of biological evolution. But that ignores the need to explain what produces viable new options between which selections are made. The option-generating mechanisms that existed when the very earliest forms of life existed on earth are very different from the option-generating mechanisms available now, some of which are themselves products of natural selection, while others are physical chemical conditions that are by-products of long past evolutionary developments: e.g. the developments that provided an oxygen-rich atmosphere, which is now part of the essential scaffolding for many life forms.

A deep analysis of some of the requirements for the FCK and some of the DCKs can be found in [Schrödinger 1944] (which Popper described as "A work of genius" in [Popper, 1976] p 137). Schrödinger shows that despite the randomness that is a feature of quantum mechanics there are other aspects that achieve the opposite of randomness: long term structural stability, without which reproduction could not be reliable and complex organisms with many parts derived largely from the genetic material could not exist. The section on evolution and epigenesis below addresses the need for that developmental predictability in individuals to be parametrised so that it can adjust to and build on

information acquired from the environment at various developmental stages. Development of spoken language is one of the best known examples. However, our theory implies that rich *internal* languages for encoding information about percepts, questions, intentions, plans, learnt generalisations and many more must have evolved long before the communicative uses of human languages (signed or spoken) evolved.

Characterising that space of possible biological forms in terms of developments describable in the language of physics (a language that may change in important ways in future decades, or centuries, as it has changed in the past) is a challenge for physicists. Not all the realised and unrealised possibilities are describable in the language of physics: for example physics does not include concepts like "percept", "desire", "prefer", "thought", "interest", "learn", "discover", "reason", and others used to describe features and activities of human and non-human minds.

So the biological developments that made possible organisms with percepts, desires, preferences, etc. may be describable only in a language that is richer than the language of physics. Developing a language rich enough to describe all the actual and possible minds that can exist on earth, and to explain their abilities and behaviours, will require use of concepts that are not part of physics.

This presents a challenge for metaphysics -- perhaps impossible for human metaphysicians to address completely. Characterising the subset of life forms that has been realised on this planet is a part of that task. Identifying the mechanisms and processes involved in that realisation, and explaining how those mechanisms were able to produce new phenomena is a combined challenge for scientists and metaphysicians, each guiding and constraining the other. As illustrated below, we now have conceptual tools for that task that were not available to earlier philosophers -- though they are still ignored by most philosophers. Scientists and engineers are still extending them as they use them.

Increasingly complex *physical* structures, mechanisms and process have come into existence since the planet formed. In addition, as a result of evolution of a multitude of forms of life making use of information there has been a steadily expanding variety of types of non-physical products of evolution, including new types of information, new types of information-based control, and increasingly complex varieties of life-forms, with both physical and non-physical aspects. Clarifying the boundaries between physical and non-physical phenomena is a task for collaboration between physicists and philosophers aided by biologists, psychologists, social scientists, and computer systems engineers with philosophical expertise.

Examples of non-physical products of evolution are the increasingly rich and complex abilities to perceive or infer things, to want or plan things, to predict and theorise about things, and to reflect on and try to understand these changes. They are non-physical insofar as the concepts used to describe them are not part of the language of physics, i.e. not explicitly definable in terms of the concepts required to express fundamental physical theories. A way of thinking about this is suggested by what we learned in the last half-century about virtual machinery: a VM (possibly composed in part of multiple interacting VMs distributed in space) can be fully *implemented* in physical machinery without the VM states and processes being *describable* in the language of physics [Maley and Piccinini 2013] [Sloman, (VMF)]

At a certain stage in the history of our planet nothing on it had any of those abilities. However, through processes that have not yet been identified, new sorts of information, new sorts of information processing mechanisms, and new uses of information somehow emerged, using mechanisms that have not yet been discovered (especially intermediate mechanisms that are parts of evolved "construction kits" [Sloman (Kits)].

Example: first-person perspectives

The phenomena that are explicable in terms of biological virtual machinery, include many things that have been thought by some to be beyond the realm of what science can describe and explain, for example "first-person perspectives" as discussed by metaphysicians, at least since Descartes, and probably earlier, though other labels are used. An example is:

> Lynne Rudder Baker [2011]:
> "There are different kinds of agents, and there are different kinds of first-person perspectives. On the one hand, all persons are agents, but not all agents (e.g., chimpanzees, dogs) are persons; on the other hand, all rational and moral agents are persons, but not all persons (e.g., human infants) are rational and moral agents."
> ...
> "My aim here is two-fold: First, to offer an account of action that emphasizes first-personal aspects of human agency; and second, to suggest a view of intentional causation that allows reasons to cause actions without being neural events."

Not everyone would agree that chimpanzees and dogs are any less persons than human infants, though this may be a simple difference of terminology. Clearly they have some things in common with human persons and also differences. Are the differences metaphysically important?

A challenge for the claim that science and engineering can help metaphysics would be to show how the phenomena involved in a first person perspective (of a human or a non-human animal) -- henceforth FPP -- could be products of certain kinds of (virtual) machinery produced by biological evolution, with aspects common to human and non-human animals and future intelligent robots, as discussed in Sloman (Affordances), and demonstrated in the videos in that paper.

On this view, although the FPPs and other features of consciousness are not neural phenomena, that does not prevent them being *implemented* partly in neural phenomena, and partly in a network of relationships (including causal relationships) to aspects of the environment.

In principle, that could also be done for the changing FPP of an intelligent robot moving around a cluttered environment and performing various tasks, although at present as far as I know there is no robot that can make use of the sorts of changing FPP features demonstrated in the above videos. Merely moving a video camera around does not produce the relevant FPPs because the camera merely records changing illumination values, without doing any grouping, interpretation, reasoning, or controlling. It has no awareness of the relationships referred to in the video, and therefore cannot perform any of the tasks for which animals use vision. Like an eye, it can provide raw data for the first-person perspective, but more is required for the potential to be realised.

Exactly what needs to be done with the video data to enable them to be used for intelligent control of actions and ability to reason about changes in the environment is a non-trivial research problem. Current robotics researchers, and brain scientists whose work I have encountered mostly misdescribe the challenges in terms of finding statistical regularities and using them to make probabilistic predictions, instead of doing the sort of qualitative and relational reasoning illustrated in the video.

As you approach and move around objects in the environment your first person perspective is experienced as constantly changing, with various percepts changing in shape, relative size, and visibility to you even though there are no changes in the size shape and visibility of the objects. The fact that evolution has been able to produce brains that can make use of these changes seems to be part of the explanation of kinds of intelligence observed in many non-human species as well as humans. I don’t think brain scientists are as yet able to explain in detail how those perspectives are used in

intelligent perception, reasoning, and action in humans and other animals -- but that is simply one of many gaps in current brain science.

A more subtle fact about FPPs is that, at least in *adult* humans, evolution has produced an additional remarkable capability, namely the *meta-cognitive* capability to notice, reason about, remember, and ask questions about FPPs. This creates what could be called Meta-First-Person Perspectives (Meta-FPPs). Meta-FPPs can be used by one individual to help or instruct other individuals.

For instance, learning to draw or paint what you see requires the ability to notice changing visibility relationships in your own experience and project them onto paper or canvas. The original process of production of Disney cartoon films required artists to produce large sequences of pictures simulating the changing visibility relationships in FPPs of imagined viewers of imagined situations and processes. These artistic uses of FPPs are recently invented uses of a kind of cognitive functionality (Meta-FPP functionality) that originally served other biological purposes, including intelligent control of location and gaze direction in order to gain practically useful information about the environment (as illustrated briefly in the video -- when change of viewpoint is used to resolve uncertainty about whether a pencil point does or does not project into the hole formed by the handle of a mug).

One of the remarkable abilities that humans and some other animals have is using a mirror to obtain visual FPPs of normally invisible body parts, like the contents of one's own mouth. Another more subtle biological function of meta-cognition of FPP is the ability to reason about another individual's FPP and use that either to deceive or to help the other, e.g. arranging plant matter to make the entrance of a den hard to see, or showing a child how to move in order to get a better view of something. We could label that a use of a **vicarious** Meta-FPP.

A puzzling feature of FPPs that was noticed long ago by philosophers and others and caused much confusion is the possibility of distorting FPPs without changing the objects looked at. For example if you shut one eye and place a finger on the lower lid of the other eye and jiggle it while looking at an object (e.g. a pen) in front of you you will see that object apparently moving. The perceived object (the pen) is not in motion, but something inside you is: an information structure changes its experienced spatial relationships, e.g. distances, directions, relative size, to other information structures. If you keep both eyes open while disturbing one eye as described you may see two pens where you previously saw only one: a static pen and a moving pen. (A widely believed theory of binocular rivalry would predict that only one could be experienced at a time, but that's because the theory is based on an impoverished collection of experiments, like many psychological theories.)

Since the physical pen is not moving the perceived "jigging" pen leads some people to the conclusion that some mysterious non-physical thing is moving, whereas what is actually happening is that a real visual information structure in a perceptual virtual machine is being perturbed in an abnormal way. There are visual contents that are moving relative to other visual contents, and normally when that happens it is caused by perceived objects in motion. But the jigging of an eyeball is not a normal part of visual perception. At some future date we'll be able to give our robots similar experiences, which may cause some of them to become philosophers.

I have talked about changes in the content of a viewer's experience, where a particular part of the content changes its experienced spatial relationships to other parts. But some philosophers (e.g. Dennett?) object to talk of such non-physical entities in relative motion and seem even to want to deny their existence, claiming that they are useful fictions. However anyone familiar with complex information processing systems in which contents of virtual machinery can behave in unintended ways (requiring the designer to engage in debugging activities) should have no problem regarding the experienced moving pen as an entity in a virtual machine that is changing its relationships within the

virtual machine, as could happen in an AI vision system if there is a flaw in the transfer of spatial information from one part of the system to another. That would indicate either a hardware fault or a need for some software to be debugged.

However, in our case, there is no debugging of software needed, merely care to be taken when fingers come into contact with eyes, and recognition that when an eyeball is moved or its shape distorted some of the projective relationships generating internal data-structures can be changed in misleading ways. Anyone who has been through an oculist's examination to asses requirements for new spectacles will have experienced cases where the oculist deliberately interferes with the patient's FPP.

There are other cases where contents of the FPP change without any physical perturbation causing the change, e.g. when you look at an ambiguous figure such as a Necker cube, or a vase-faces picture. These "flips" are often misdescribed as cases where only one view can be experienced at a time, e.g. only two faces or only a vase, whereas it is easy to see such a picture as depicting two faces with a vase wedged between them, though that does not normally happen spontaneously. It is much harder to see different views of a necker cube simultaneously though a few people can (with a struggle) see two inconsistent part-cubes joined up in an impossible configuration.

I see no reason why future research should not enable all these phenomena to be replicated in machines with visual capabilities combined with the Meta-Cognitive abilities mentioned previously in connection with Meta-FPPs. I don't know whether current physical computing technology will suffice or whether new kinds of information-processing mechanisms will be required, if robots are to replicate the geometrical and topological reasoning capabilities, and mathematical experiences, of humans. (New mechanisms may be required to replicate the abilities to discover geometrical and topological theorems reported in Euclid's *Elements*.)

NOTE: Baker's paper does not mention robots or artificial intelligence, nor the possibility of scientific explanations of how FPPs come into existence and interact causally with other things. However she does deny that FPPs and their contents are neural phenomena, so perhaps she might allow that they could exist in intelligent machines but would not be electronic phenomena, or chemical phenomena if the machines used chemical computations (as brains seem to do). That denial would be correct: the contents of virtual machines engaged in complex information processing are not identical with physical parts or processes in the implementation ([Maley and Piccinini 2013] [Sloman, (VMF)]).

Our ability to attend to, ask questions about, and theorise about the contents of FPPs has caused much philosophical puzzlement and confusion, and has led some to regard the contents of such FPPs (variously labelled sense-data, qualia, percepts, experiences, ideas, seemings, ...) as having a kind of existence that cannot be explained in terms of physical processes because their contents are not physical (no physical object moves in the manner perceived). But a good education in philosophical software engineering should help to counter such confusions.

To be continued

If we consider the *information processing* requirements of different sorts of organisms, from the very simplest through increasingly complex physical forms, behaviours and environments, we can compare these two sorts of products of evolution:

> (a) organisms that merely detect and react to states of the immediate environment and their own states, responding directly only to internally and externally sensed information without any concern about the source of the information, and without regard to ways in which the information might change, and without explicitly taking steps to get the information or improve on it, etc. and

(b) organisms that can discover, learn about, and think about entities, states and processes that endure and interact in the environment, sometimes independently of any perceiver or thinker, sometimes not.

(This is not intended to be an exhaustive distinction -- many more sub-types of information-users can be distinguished, some of them contrasted in [Sloman (2006)] and also here http://www.cs.bham.ac.uk/research/projects/cogaff/misc/meta-morphogenesis.html#focus)

Organisms of type (b) will need to have an ontology (explicit or implicit collection of information categories) that includes enduring locations, and enduring occupants of locations some of them able to move between locations, where the existence of the occupants and locations is independent of the perceiver. [McCarthy (1996)] As McCarthy notes the ontologies can depend not only on the physical structure of the environment or the properties and behaviours of the occupants, but also the observer's abilities to move. Animals that move only along supporting surfaces and animals that can fly will require different ontologies for perceived structures and changes of appearance, types of obstacle, types of route between locations, etc.

It should be obvious how such abilities of type (b) would be of immense importance to creatures capable of moving around in an extended, partially visible, environment with various nutrients, shelters, dangers, obstacles and possibly other creatures -- some competitors some not.

Being able, in addition, to think about how the contents of what you know might be incomplete and what can be done to fill some of the gaps could be essential for survival in a partly unfriendly world. A more complete version of this story would need to provide evidence based examples, and explanatory mechanisms of varying kinds and degrees of complexity. Tracing their evolution would be an exercise combining science and metaphysics: the science explaining how and why new metaphysical kinds come to exist, or have instances, on this planet.

Useful varieties/layers of meta-cognition

Some control mechanisms can use currently sensed information to control changes of state or location, for example the Watt governor that uses centrifugal force to control a steam valve, or a windmill that uses a vane or a secondary rotor to produce changes in direction to maximize the rate at which energy is acquired from wind. Biological evolution has produced a wide variety of functionally similar homeostatic control mechanisms using negative feedback, as in thermostats, More complex designs can use comparisons between information sensed at different times as part of homeostatic control -- using a memory for previous sensory states. Detecting rate of change, e.g. velocity, requires two records to be compared. Information about acceleration (increasing or decreasing velocity) would require at least three time-labelled records.

Some organisms use changes in sensed information of type (b), above, to control motion through a complex enduring environment, for example using chemical or other sensors to detect that they are approaching a desired target. More sophisticated organisms with multiple, changing, needs and opportunities involving different parts of the environment at different times, may combine stored past information and currently sensed information to derive new information about their current location and direction of motion in a structured environment that extends far beyond the current reach of their sensors.

It is likely that many of the organisms with such capabilities, including very young human children, acquire and use the information, but lack the kind of meta-cognitive information-processing architecture required to notice what they are doing. The requires evolution of additional layers of information processing, as does the ability to acquire and use, or to speculate about, the information

available to others, which can be useful when stalking prey, hiding from predators, collaborating with conspecifics and educating offspring. The ability to record one's sensory experiences and the decisions based on them can be useful when things go wrong: triggering attempts to find out what previously unnoticed difference between two situations could have cause a decision making process to fail (or succeed) which in an apparently similar previous context succeeded (or failed).

These are among the many cases in which evolutionary opportunities could contribute to biological changes from organisms that are less like humans to organisms that are more like humans in their information processing, including what they are and are not conscious of. Similar changes can occur within an adult human under expert guidance, e.g. learning how to paint, how to control posture to avoid stress, how to design good software, how to control bow movements to improve the tone when playing a violin, how to detect errors in one's mathematical thinking, and many more. Often that requires use of introspection (e.g. focusing attention on how a particular action feels) as part of the control process. (Compare [Spener 2015].)

These requirements will have to be met by future intelligent social robots, though at present all the ones I know about have are very primitive "canned" or "learned" capabilities based on the theories of intelligence of their designers, which may be good or bad theories. The hope that good social competences can be created merely on the basis of statistical learning mechanisms based on positive and negative rewards ignores much of what we have learnt in the last half century, e.g. about the limits of behaviourist psychology.

Among all those products of evolution on earth, a subset that I find especially interesting (partly inspired by [Kant, 1781]), are the mathematical capabilities that evolved in our ancestors, leading up to the discoveries reported in [Euclid's Elements], based on special kinds of introspection required for finding proofs in geometry and other branches of mathematics.

These are not merely new and increasingly complex configurations of matter or processes of physical change. They include increasingly complex, abstract, and powerful mechanisms and processes that create, manipulate, derive, use and communicate information contents, some without, and some with, manipulations, derivations and uses of information have causal consequences, some of them, on this planet, truly awful, others wonderful products of mathematics, science, art, and human kindness and concern for other things.

The fact that such changes can occur on a planet, or in a solar system, that initially has nothing but physical/chemical matter and interactions is quite staggering, and that surely involves a host of metaphysical changes in what goes on in this part of the universe.

Moreover, it does not seem to be the case that this was an inherently *linear* process, like the consequences of a fixed collection of equations applied to a fixed collection of numerical measures. For everything that actually happens there are many alternative things that could have happened, varying in kind and complexity, including major qualitative changes that are capable of being triggered by a few small variations in physical states and processes. (I leave open whether the "could have" refers to possible slight variations in initial conditions on the planet, or some kind of intrinsic creativity of later processes of evolution and development.)

> *It can't all just be determined by initial physical conditions if the products of evolution include mechanisms for identifying and studying sets of alternative possibilities, eliminating some and then selecting a remaining possibility at random (e.g. using a quantum-mechanical randomiser), or if selection processes use preferences, values, principles, standards, or goals that were not fully determined, or whose interactions were not fully determined, by an earlier physical state of the universe.*

Among the many things that need to be explained are changes that are tightly related to human or non-human consciousness, including making mathematical discoveries, some of which allow broad swathes of knowledge to be transformed later.

What can be learnt from other disciplines?

Philosophical discussions of metaphysical "grounding" of consciousness that I have encountered tend to focus on what other philosophers have said or written, and on common sense and everyday experiences, but generally ignore what might be learnt from scientific and and technical investigations and discoveries such as:

(a) Investigations of how minds of various kinds (from microbe-minds onwards) could have evolved on an initially lifeless, mindless planet, including evolution of varieties of self awareness, self control and mathematical meta-cognition. This is a major theme in the Turing inspired Meta-Morphogenesis project. Sub-topics are included in [Ganti 1971/2003], [Dennett 1996], [Sloman 2013b)], [Sloman (Kits)] (and many more).

(b) Progress in various aspects of computer science and computer systems engineering concerned with design, implementation, testing, debugging, and uses of *virtual machinery*, including virtual machine architectures supporting self-monitoring, self-modification, various kinds of self-control, and "horizontal" causation between virtual machines as well as upward and downward causation [Sloman (VMF)], [Maley & Piccinini 2013]. (Daniel Dennett sometimes comes close to acknowledging these phenomena, but usually backs away from the existence of virtual machines and virtual machine states with causal powers, suggesting instead that virtual machine talk is some kind of useful fiction, possibly because he, like most philosophers, has never had to design test and debug a complex running virtual machine, which should be a requirement for advanced study in philosophy of mind).

(c) Attempts to design and implement artificial minds of various kinds, with qualia, various kinds of meta-cognition, self-control and introspection discussed in [Minsky 1968], [Sloman 1978]. [Franklin 1995], [BICA Architectures], [Minsky 2005], [Franklin LIDA], [Rescorla 2015] and other more recent developments. (For an extended, thorough, but partly out of date, survey see [Boden 2006]. Her more recent (very short) introductory overview is suitable for non-experts [Boden 2016]);

(d) Computational mechanisms that explaining the possibility of incommunicable, ineffable, qualia based on causal indexicality (as explained in [Sloman & Chrisley 2003]) such as occurs in "self organising networks" (e.g. Kohonen nets). In 2012 Marcel Kvassay wrote a detailed tutorial commentary on this idea comparing and contrasting the ideas with the anti-reductionism of David Chalmers: http://marcelkvassay.net/pdf/machines.pdf

(e) Requirements for the sorts of *mathematical consciousness* discussed by Immanuel Kant, involved in geometrical and topological discoveries made by Euclid, Archimedes and many others, long before the development of modern logic and the arithmetisation of geometry by Descartes. Those ancient mathematical competences are not explained or modelled by current theories or computational models of mathematical reasoning, vision, and learning, or current models in neuroscience (that I know of). The unexplained human mathematical capabilities seem to be related to intelligence in pre-verbal children and other animals [Kant 1781], [Sloman 1962], [Sloman (impossible)], [Karmiloff-Smith 1992]);

Philosophy that is detached from developments in science, mathematics and technology can lead to theories that have already been overtaken or implicitly refuted by non-philosophers. It also risks ignoring scientific and technical concepts relevant to the philosophical project, e.g. concepts developed in decades of research and engineering related to increasingly sophisticated virtual machines used in information processing systems, and their powers, whose details have eluded most philosophers, despite widespread everyday use of the technology. [Maley & Piccinini 2013] give some examples. Compare [Pollock 2008], and [Sloman (VMF)].

Branching grounding

A static theory of metaphysical reality may be explicitly or implicitly accepted by some philosophers because they ignore both the richness and creativity of biological evolution, whose complexity and diversity threatens to defeat all human thinkers and some of the results of computer systems engineering and AI that ought be far more widely known

The notion of "Grounding as metaphysical causation" in [Wilson 2015] usefully shifts philosophy away from contemplation of static structures and possibilities to consider grounding of *branching* collections of possibilities possibilities. At any time various alternative developments are possible, e.g. in a game of cricket. If one of the possibilities is realised, e.g. the batsman is bowled out, that changes the possibilities for further developments. It may also be the case that a particular possible situation can be realised via (grounded by) different possible histories: e.g. a situation in which at the lunch break, team members X and Y are still batting, and the score for their side is 99 could be reached via many sequences of states and events. In that sense the type of grounding specified by Wilson is a partial ordering: transitive, anti-symmetric, and irreflexive. (Such a structure may or may not be a mathematical lattice.)

The same kind of branching set of possibilities related by partial orders of grounding is also relevant to abstract structures. For example, in a language with a phrase structure grammar, a complex linguistic structure, e.g. phrase or clause or sentence, with several parts that have parts, so that the parse tree has multiple layers, could be constructed in various different orders. (For anyone who is unfamiliar with this idea there are many online tutorials on parsing, parse trees, and grammars.) The same thing is true of complex arithmetical or algebraic expressions: the parts, at various levels of complexity, can be assembled in different orders, with changing sets of possibilities for continuing on the basis of what has been constructed at any time. When such abstract structures stand in grounding relations, their instances can also. E.g. each of the sentences in this document is an instance of an abstract grammatical pattern, including this one. The grammatical components of the sentences are instances of abstract grammatical structures. However, actual construction processes (in the mind of a speaker or writer) can sometimes proceed from abstract structures that are instantiated piecemeal, complicating the notion of grounding. That complication will not be discussed further here, though it is also relevant to Biological evolution, our main topic.

Mathematical truths, e.g. Pythagoras' theorem, can also have alternative branching grounds, as well as grounding many other mathematical truths via branching "upward" derivation possibilities. Moreover, mathematical reasoning, like sentence construction and evolutionary development can sometimes involve derivation of an instance, or sub-type, from a more abstract structure. Again, this complication will be ignored here.

The relevance of all this to consciousness may not be obvious. In part that's because the noun "consciousness", like many other abstract nouns, including "goodness", "efficiency", "relevance", "legality", "reliability", and many more, has a deceptively complex meaning on account of being polymorphous, as will be explained below.

This implies that the notion of "grounding of consciousness" is also polymorphous, and different sorts or examples of consciousness will require different sorts of grounding, for example consciousness of being hungry, consciousness of blurred vision, consciousness of being unpopular, consciousness of mathematical impossibility or necessity, and consciousness of being lost in a new town.

Like Kant [Kant 1781], I am particularly interested in mathematical consciousness, a topic that is usually ignored by those who write about consciousness, and about which I don't believe there are any good theories in psychology or neuroscience. But I shall not go into that in detail here. (It was the topic of my thesis [Sloman 1962].) I think this sort of mathematical consciousness is related to some cases of consciousness of positive and negative affordances almost, but not quite, noticed by James Gibson [Gibson 1979]. For examples of consciousness of impossibility and necessity see [Sloman (impossible)].

Grounding of products of biological evolution

Since biological organisms, and their information processing are also complex structures, with relationships between parts at many levels of complexity, there are similar branching possible realisations of sets of possibilities, both in the development of each individual (a sort of 'growth game') and also in the evolutionary history of various biological *types*: types of organism, types of biological mechanism, types of competence, types of group organisation, types of ecosystem, etc.

Moreover, as explained in [Sloman (Kits)] the possibility of a particular biological mechanism or species may depend on the prior existence of a set of *construction-kits* required during the process of assembly, or growth, or evolution, even though the construction kits are not parts of the organism.

An extreme case is the complex apparatus in the womb of a female mammal that makes possible the development of a foetus. In other cases parts of construction kits may be portions of the environment used for support, or chemicals that play a crucial role at certain stages of development, or a control mechanism that manages the coordinated development of parts of the organism's body, or brain, or immune system but is discarded after development.

The idea of a system of branching possibilities for the development of each organism of a species was summed up in [Waddington 1967] using the label "Epigenetic landscape", which could be thought of as fixed for a particular species, or a particular individual. This idea was generalised as shown in Figure EVO-DEVO to allow the landscape determining subsequent options to be constructed and repeatedly modified during an individual's development, under the influence of the environment, which itself may depend in part on prior actions of the individual.

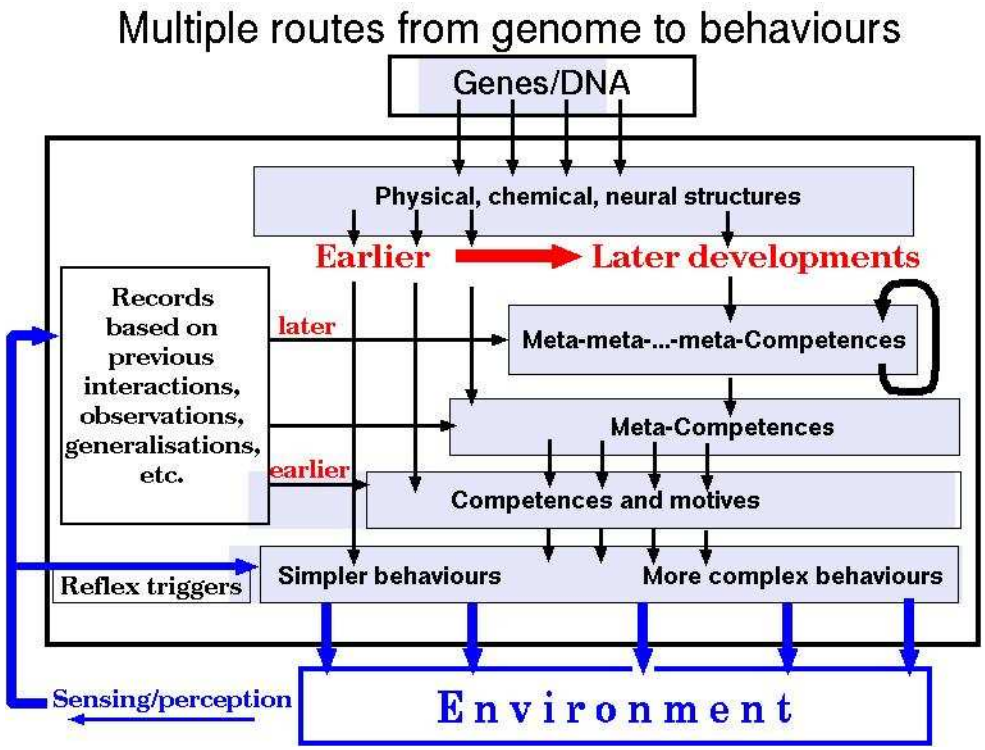

Figure EVO-DEVO:

A particular collection of construction kits specified in a genome can give rise to very different individuals in different contexts if the genome interacts with the environment in increasingly complex ways during development, allowing enormously varied developmental trajectories. Precocial species use only the downward routes on the left, producing only "preconfigured" competences. Competences of members of "altricial" species, using staggered development, may be far more varied within a species. Results of using earlier competences later interact with delayed products of the genome, producing new "meta-configured" competences shown on the right. This is a modified version of a figure in [Chappell&Sloman 2007], intended to replace Waddington's famous depiction of a fixed "epigenetic landscape". Genetic abnormalities and malign environmental influences can interact in hugely varied ways at many different stages during processes of development (as emphasised by Annette Karmiloff Smith in online lectures and [Karmiloff-Smith 1992]).

Without this sort of flexibility in the developmental process the huge variety of languages, physical artefacts, dwellings, machinery, and cultures arising out of a (presumably) common genetic specification would not be possible. This explains how certain sorts of development (e.g. expert use of a particular language) may be possible only at certain stages of development. I am claiming that development by the individual of a type of mind with particular forms of consciousness, e.g. fluent comprehension of a spoken or signed language may also be associated with a branching set of possibilities grounded, in part, in the stages in the individual's prior history.

These ideas invite a cross-disciplinary bridge-building venture, relevant to a universe in which there are metaphysically distinct types of phenomena at different times, and different places, e.g. different kinds of minds and different kinds of mental contents, made possible at different times and places by evolutionary developments. These are explanations of possibilities in the sense discussed in [Sloman 2014(in progress)], expanding on Chapter 2 of [Sloman 1978].

It is clear that many major scientific discoveries are about *possibilities* not *laws* or their instances, and some major theoretical advances provide *explanations* of previously unexplained possibilities. New possibilities extend the *kinds* of things and processes that can come into existence, or make them more readily reachable: extending what can be grounded, like adding hinges to Meccano. Compare the "descriptive metaphysics" of P.F.Strawson, which implicitly (despite Strawson's disclaimers) allows the scope of metaphysics to change as cultures evolve [Sloman (Strawson)].

The Darwin/Wallace theory of natural selection is generally thought to explain the possibility of qualitative changes of form, behaviour, and information processing in life forms e.g. [Dennett 1995]. (Great care is needed in specifying what natural selection is and does, but for now I'll merely observe that it does not support or depend on the truth of "Only the fittest survive" as shown by hugely different levels of fitness among humans.) What grounds the selectable possibilities on which natural selection depends? What makes them possible, i.e. available for selection?

[Schrödinger 1944] showed how quantum mechanics explains/grounds the possibility of changeable but very stable sub-microscopic information structures, which, in turn, make possible robustly inheritable biological changes. These work only in appropriate portions of the universe (e.g. on the earth but not at the centre of the sun). He also anticipated some of Shannon's ideas about digital information structures, for example in his discussion of the possibility of reliably encoding huge amounts of information in very small aperiodic physical/chemical structures that, despite the inherent non-determinism of quantum mechanisms, are predictably highly stable even in normally disruptive environments -- a crucial requirement for grounding of evolution by natural selection.

QM also grounds branching layers of possible biological "construction kits" [Sloman (Kits)]. Increasingly complex evolved "construction kits" of many sorts (including concrete, abstract, and hybrid construction kits) produced in successive evolved layers, constantly extend the immediate reach of natural selection - while suppressing some alternative possible evolutionary trajectories. (Un-grounding?) Compare the "radical emergence" of [Longo, Montévil, Kauffman 2012].

Consciousness is polymorphous

Early life-forms had physical structures, physical behaviours and information-processing capabilities enabling control of: feeding, waste disposal, maintenance of temperature, osmotic pressure, etc., and reproduction possibilities grounded in physical/chemical conditions and resources [Ganti 1971/2003]. Was there any consciousness? This question cannot have a simple Yes/No answer because, as indicated previously, consciousness is polymorphous, like efficiency, reliability, stability, and many other concepts ([Ryle 1949, White 1967] *et al.*),

This implies that there will be no evolutionary stage at which "it" comes into existence. Rather, for various types of organism X and various types of objects of consciousness Y, there will be different earliest occurrences of an X being conscious of (or that) something of type Y. For instance, a microbe might detect a potentially nourishing, or harmful, molecule in contact with its membrane, and on that basis allow it to pass through the membrane, or not. Such detection is a very primitive form of consciousness: primitive in the simplicity of its content, in the simplicity of the mechanisms used, in the simplicity of its effects, and in the lack of any meta-cognition: no self awareness is involved. Later forms included much greater richness.

There is no stage at which the transition to organisms with consciousness occurred: there are many (discrete) stages at which different transitions to new kinds of consciousness occurred. It is often wrongly assumed that the alternative is a continuum: but chemistry cannot ground continuous evolution!

So grounding of different types of consciousness is a discrete multi-layered process with possible branches, like stages in development of a foetus, including various individual stages in the development of human linguistic competences, creative process that is often mistaken for a pure learning process Sloman 2015 (Lang).

(The type of grounding specified in [Wilson 2015] is a partial ordering: transitive, anti-symmetric, and irreflexive.)

Some transitions require more complex supporting mechanisms than others. A microbe's being conscious of (having, and being able to use, information about) something in the immediate environment is very much simpler than an elephant's consciousness of an extended enduring environment only parts of which are sensed at any time, and which includes different sources of nourishment, different dangers to be avoided, places to find mates, a sheltered location in which young can be fed, etc.

The evolutionary transitions required to bridge the microbe-to-elephant gap, making possible the later forms of consciousness, are mostly unknown. However, mobile robots already have extremely simple variants - some with simple introspective awareness.

There are many different consciousness-related evolutionary (and developmental) transitions, e.g. grounding abilities to use information about current needs, future needs, or past needs and whether those needs were met after various actions: grounds of affective consciousness.

Varieties of meta-consciousness (and meta-meta...consciousness) emerged later, including awareness of past decisions and their consequences, awareness of consequences of not satisfying various requirements, awareness of reasoning or planning processes that did or did not produce expected results, increasingly abstract and de-centred conscious evaluations of actions and possible states of affairs forming ethical judgements, and many more. (Compare the discussion of uses of introspection in [Spener 2015] and [Sloman (VMF)].)

For organisms that interact with other organisms, e.g. mates, offspring, parents, competitors, potential collaborators, etc. evolutionary transitions produced (newly grounded) mechanisms enabling acquisition, processing and use of information about the information states of others; e.g. being conscious that an infant does not know what to do in a potentially dangerous situation, or that a predator can/cannot see the location of one's nestlings [[Gibson 1979].

Among the still unexplained possibilities are those that enabled the proto-mathematical and mathematical discoveries leading up to the amazing mathematical achievements of [Euclid (Elements)] over 2,500 years ago, using consciousness of geometrical and topological possibilities and constraints, not yet replicated in robots or automated theorem provers (or explained by neuroscience)[Sloman (triangles)].

This viewpoint allows the possibility that new forms of technology, or new evolutionary developments, or some combination, may be able to ground new forms of consciousness with features that we are currently unable to conceive of.

We can distinguish "shallow grounding", illustrated by the much discussed "singleton Socrates", and Wilson's "deep grounding", illustrated by the enormous multi-layer multi-branch metaphysical creativity of evolution, which, among other things, demonstrates that creativity does not require intentionality. With luck, describing undiscovered transitions required to explain the grounding of human consciousness will not require super-human competences.

References

The Meta-Morphogenesis project was proposed as a tentative answer to the question: if Alan Turing had not died two years after publishing his paper on "The chemical basis of morphogenesis", what might he have worked on for the next 30-40 years? At present the project is represented by a large and steadily growing collection of papers and discussions of various aspects of evolution of biological information processing and the various "construction kits" required to support them: http://goo.gl/eFnJb1

Margaret Boden, 2006, *Mind As Machine: A history of Cognitive Science* (Vols 1--2) (2006) OUP

Margaret A. Boden, 2016, AI: Its nature and future Hardcover, OUP, Oxford, https://www.amazon.co.uk/AI-nature-future-Margaret-Boden/dp/0198777981

BICA Society Survey of architectures http://bicasociety.org/cogarch/.

Lynne Rudder Baker, 2011, First-Personal Aspects of Agency *Metaphilosophy* vol. 42, no. 3 (Wiley-Blackwell)

Jackie Chappell and Aaron Sloman, 2007, Natural and artificial meta-configured altricial information-processing systems, *International Journal of Unconventional Computing*, 3, 3, pp. 211--239, http://www.cs.bham.ac.uk/research/projects/cogaff/07.html#717

N. Chomsky, 1965 *Aspects of the theory of syntax* MIT Press, Cambridge, MA,

M. T. Cox and A. Raja, (Eds) (2011) *Metareasoning: Thinking about thinking*, MIT Press, Cambridge, MA

Dennett, D. *Darwin's dangerous idea: Evolution and the meanings of life.* London and New York: Penguin Press.

D.C. Dennett *Kinds of minds: towards an understanding of consciousness,* Weidenfeld and Nicholson, London, 1996, http://www.amazon.com/Kinds-Minds-Understanding-Consciousness-Science/dp/0465073514

Euclid & Casey, J. *The First Six Books of the Elements of Euclid.* Salt Lake City: Project Gutenberg. http://www.gutenberg.org/ebooks/21076 (Third Edition, Revised and enlarged. Dublin: Hodges, Figgis, & Co., Grafton-St. London: Longmans, Green, & Co. 1885.)

Chrisantha Fernando, 2008, Review of "The Principles of Life" by Tibor Ganti. [Ganti 1971/2003], in *Artificial Life*, 14, 4, pp. 467--470, Fall, http://dx.doi.org/10.1162/artl.2008.14.4.14404

Stan Franklin, 1995, *Artificial Minds,* Bradford Books, MIT Press, Cambridge, MA,

Stan Franklin and Colleagues (2008) IDA-LIDA Tutorial http://ccrg.cs.memphis.edu/tutorial/index.html

Ganti, T. 1971/2003 *The Principles of Life.* New York: OUP. (Eds. Eors Szathmary & James Griesemer, Translation of the 1971 Hungarian edition)

Gibson, J J. *The ecological approach to visual perception.* Boston, MA: Houghton Mifflin.

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

Annette Karmiloff-Smith, 1992, *Beyond Modularity: A Developmental Perspective on Cognitive Science*,
MIT Press, Cambridge, MA,

Kohonen networks, explained briefly here
https://www.ibm.com/support/knowledgecenter/SS3RA7_15.0.0/com.ibm.spss.modeler.help/kohonennode_general.htm
in IBM's Knowledge Center.

Longo, G., Montévil, M. Kauffman, S. No entailing laws, but enablement in the evolution of the biosphere. In *Proceedings of the 14th annual conference on genetic and evolutionary computation* (pp. 1379-1392). New York, NY, USA: ACM. http://arxiv.org/abs/1201.2069v1

Maley, C. Piccinini, G. 2013, Get the Latest Upgrade: Functionalism 6.3.1. *Philosophia Scientiae*, 17(2), 1-15.
http://poincare.univ-nancy2.fr/PhilosophiaScientiae/

John McCarthy, 1995 Making robots conscious of their mental states, *AAAI Spring Symposium on Representing Mental States and Mechanisms*, Palo Alto, CA, AAAI, Revised version:
http://www-formal.stanford.edu/jmc/consciousness.html

John McCarthy (1996). "The Well Designed Child"
http://www-formal.stanford.edu/jmc/child.html
(Later published in the AI Journal, 172, 18, pp 2003--2014, 2008)

John McCarthy, 2004, Notes on self-awareness Report related to the 2004 April DARPA Workshop on self-awareness. Stanford University

Marvin Minsky, 1968, Matter Mind and Models, in *Semantic Information Processing*, Ed. M. L. Minsky, MIT Press, Cambridge, MA,

Marvin Minsky, 2005 Interior Grounding, Reflection, and Self-Consciousness, *Brain, Mind and Society, Proceedings of an International Conference on Brain, Mind and Society*, Tohoku University, Japan, September, 2005, reprinted in *Information and Computation*, Eds G. Dodig-Crnkovic and M. Burgin, World Scientific Press, 2011.
http://web.media.mit.edu/~minsky/papers/Internal%20Grounding.html
PDF available here:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/minsky/interior-grounding.pdf
NOTE: Minsky is here referring to so-called "Symbol Grounding", but the paper is relevant anyway.

J.L. Pollock (2008), What Am I? Virtual machines and the mind/body problem. *Philosophy and Phenomenological Research*, 76(2), 237-309. http://philsci-archive.pitt.edu/archive/00003341
(Alas his work was cut short by cancer.)

K.R. Popper, (1934) *The logic of scientific discovery*, Routledge, London,

K. R. Popper, (1976), *Unended Quest*, Fontana/Collins, Glasgow,

K.R. Popper, (1977) Natural Selection and the Emergence of Mind (Darwin Lecture), Delivered at Darwin College, Cambridge, November 8, 1977. Available at:
http://www.informationphilosopher.com/solutions/philosophers/popper/natural_selection_and_the_emergence_of_mind.html

Michael Rescorla, (2015), The Computational Theory of Mind, *The Stanford Encyclopedia of Philosophy* Ed. Edward N. Zalta, Winter 2015,
http://plato.stanford.edu/archives/win2015/entries/computational-mind/

S. J. Russell and E. H. Wefald (1991), *Do the Right Thing: Studies in Limited Rationality,* MIT Press, Cambridge, MA,

S.J. Russell (2013) Rationality and Intelligence: A Brief Update Presented at PTAI-13 Conference, Oxford, 2013.
www.eecs.berkeley.edu/~russell/papers/ptai13-intelligence.pdf

G. Ryle (1949), *The concept of mind.* London: Hutchinson.

E. Schrödinger (1944), *What is life?* Cambridge: CUP.
(Annotated extracts from the book are available here:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/schrodinger-life.html

A. Sloman, 1962, *Knowing and Understanding: Relations between meaning and truth, meaning and necessary truth, meaning and synthetic necessary truth*, DPhil Thesis, Oxford University, (Digitised online version 2016). http://www.cs.bham.ac.uk/research/projects/cogaff/62-80.html#1962-01

A. Sloman, 1976, What are the aims of science, *Radical Philosophy,* 13, pp. 7--17, (Revised and extended in [Sloman 197](Ch 2).) Original version:
http://www.radicalphilosophy.com/article/what-are-the-aims-of-science

A. Sloman, 1978, Revised 2015, *The Computer Revolution in Philosophy, Philosophy, Science and Models of Mind*, Harvester Press (and Humanities Press),
http://www.cs.bham.ac.uk/research/cogaff/62-80.html#crp

A. Sloman, 1996, Beyond Turing Equivalence, in *Machines and Thought: The Legacy of Alan Turing (vol I),* Eds. P.J.R. Millican and A. Clark, The Clarendon Press, Oxford, pp. 179--219,
http://www.cs.bham.ac.uk/research/projects/cogaff/96-99.html#1

A. Sloman, 2002, The irrelevance of Turing machines to AI, in *Computationalism: New Directions,* Ed. M. Scheutz, MIT Press, Cambridge, MA, pp. 87--127,
http://www.cs.bham.ac.uk/research/cogaff/00-02.html#77

A. Sloman, R.L. Chrisley, (2003,) Virtual machines and consciousness, *Journal of Consciousness Studies,* 10, 4-5, pp. 113--172,
http://www.cs.bham.ac.uk/research/projects/cogaff/03.html#200302

A. Sloman, 2006
Requirements for a Fully Deliberative Architecture (Or component of an architecture), University of Birmingham, UK, May, 2006,
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/fully-deliberative.html
(This presents a variety of intermediate cases between simple reactive information-processing architectures and "fully deliberative" architectures with several concurrently active layers of processing, that evolved at different times, and develop at different stages in individuals.)

A. Sloman (Affordances) Predicting Affordance Changes (Alternative ways to deal with uncertainty), Nov, 2007, Unpublished discussion paper, COSY-DP-0702, School of Computer Science, University of Birmingham,
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/changing-affordances.html

A. Sloman, (Triangles) Hidden Depths of Triangle Qualia (Work in progress).
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/triangle-theorem.html

A. Sloman, 2009, Architecture-Based Motivation vs Reward-Based Motivation, *Newsletter on Philosophy and Computers,* American Philosophical Association, 09, 1, pp. 10--13, Newark, DE, USA.
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/architecture-based-motivation.html

A. Sloman, (VMF) Virtual Machine Functionalism (VMF):
The only form of functionalism worth taking seriously in Philosophy of Mind and theories of Consciousness. (Technical report, in progress)
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/vm-functionalism.html

Sloman A (2013b) Virtual machinery and evolution of mind (part 3) meta-morphogenesis: Evolution of information-processing machinery. In: Cooper SB, van Leeuwen J (eds) *Alan Turing - His Work and Impact*, Elsevier, Amsterdam, pp 849-856,
Expanded web site
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/meta-morphogenesis.html

Sloman, A (2013c) Meta-Morphogenesis and Toddler Theorems: Case Studies. Online discussion note, http://goo.gl/QgZU1g U. of Birmingham.

Aaron Sloman (2014, in progress) Using construction kits to explain possibilities
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/explaining-possibility.html

Aaron Sloman (Strawson) Meta-Descriptive Metaphysics Extending P.F. Strawson’s "Descriptive Metaphysics". (Work in progress.)
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/meta-descriptive-metaphysics.html

Aaron Sloman (Kits) 2015, Construction kits for evolving life
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/construction-kits.html

A. Sloman (Impossible), Some (Possibly) New Considerations Regarding Impossible Objects. Their significance for mathematical cognition, and current serious limitations of AI vision systems. (Discussion paper.)
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/impossible.html

A. Sloman, 2015 (Lang) What are the functions of vision? How did human language evolve? Online research presentation, University of Birmingham.
http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#talk111

Maja Spener, 2015, Calibrating introspection, *Philosophical Issues 25, Normativity,* pp. 300-321, Wiley,
DOI 10.1111/phis.12062,

Strawson, P F. *Individuals: An essay in descriptive metaphysics.* London: Methuen.

C. H. Waddington, 1957, *The Strategy of the Genes. A Discussion of Some Aspects of Theoretical Biology*, George Allen & Unwin.

White, A R. *The Philosophy of Mind.* New York: Random House. 1967

Wilson, A. Metaphysical Causation (Technical Report). University of Birmingham, Department of Philosophy. 2015 http://alastairwilson.org/files/mc120715.pdf
(Video lecture: https://www.youtube.com/watch?v=j2l1wKrtlxs)

Some related discussion papers

http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0803
Varieties of Meta-cognition in Natural and Artificial Systems (2008)

http://www.cs.bham.ac.uk/research/projects/cogaff/misc/trisect.html
How to trisect an angle. (Using P-Geometry)

http://www.cs.bham.ac.uk/research/projects/cogaff/misc/bio-math-phil.html
Biology, Mathematics, Philosophy, and Evolution of Information Processing.

http://www.cs.bham.ac.uk/research/projects/cogaff/misc/austen-info.html
Jane Austen's concept of information (As opposed to Claude Shannon's)

This document
This is http://www.cs.bham.ac.uk/research/projects/cogaff/misc/grounding-consciousness.html
A PDF version may be added later.

Installed: 15 Jun 2016
Updated:

Maintained by: Aaron Sloman
School of Computer Science
The University of Birmingham