Evolved Cognition and Artificial Cognition: Some Genetic/Epigenetic Trade-offs for Organisms and Robots

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

April 16, 2011

DRAFT (shorter version under review): likely to be updated

Abstract

This paper attempts to characterise a variety of attitudes to nature-nurture relationships in natural or artificial intelligence, and defends and elaborates a version presented by McCarthy. Some researchers assume that a machine can acquire human-like intelligence if it initially has (a) a large but empty information store, (b) a very powerful general-purpose learning mechanism, (c) a rich environment in which to learn, and possibly also (d) a teacher to guide the learning; with learning occurring at speeds comparable to learning in humans, rather than requiring evolutionary time-scales, despite the absence of any specific innate knowledge about the environment initially, nor any innate concepts (an ontology) specific to the environment. This assumption is closely related to the ancient empiricist "tabula rasa" theory of knowledge acquisition. That theory can be contrasted with alternative hypotheses regarding starting points for various kinds of learning about the world in diverse animals and, by implication, future intelligent robots, including the approach proposed in (McCarthy, 2008), making use of a Design-based, environmentally informed, nativist meta-knowledge theory. An extended version of McCarthy's approach, applied across species, can lead to deeper and more powerful explanatory theories of information processing in organisms than the alternatives, and can also provide new ideas about both *requirements* for future intelligent machines and also possible new *designs*, linking AI and Biology in new ways.

1 Introduction

It is sometimes suggested (e.g. Turing(1950)) that it is possible to design a machine that will acquire human-like intelligence if it initially has no prior knowledge about the environment, but has (a) plenty of information storage capacity, (b) a very powerful general-purpose (topic-neutral) learning mechanism, (c) a rich environment in which to learn, and (d) a teacher to guide the learning – or possibly only some innate exploratory dispositions (Kuipers and Modayil, 2004). This is closely related to the ancient empiricist theory of knowledge, also defended by John Locke, based on a "tabula rasa"[1] learning mechanism (TRLM). It is assumed that a TRLM can learn at speeds comparable to learning in humans, and will not require evolutionary time-scales, despite having no specific knowledge about the environment initially, nor even any concepts (an ontology) specific to the environment. However it is far from obvious that current working examples can "scale out" to include much richer 3-D environments without being defeated by combinatorics. (A challenge to TRLM ideas has been presented on my web site[2]. Comments welcome.)

TRLM is just one of several views on the relationship between nature and nurture in intelligent systems. I shall outline some alternatives to the TRLM hypothesis, hypothesising different starting points for learning about the world in animals and, in future intelligent robots. The hypotheses differ in (a) what they assume the products of biological evolution available from birth or hatching are, and (b) how they assume we can find out what the innate systems are. Although this is somewhat oversimplified, the main alternatives can be grouped roughly into the following categories explained later (mnemonic labels are provisional):

> (RNT) The rationalist nativist thesis (one of the old ideas in philosophy, going back to Plato and probably earlier);
> (ENT) The empirical nativist thesis (currently manifested in a lot of work in developmental psychology, e.g. Spelke);
> (EMN) Empirical meta-knowledge nativism (closely related to Chomsky's ideas about linguistic universals);
> (DNM) Design-based nativist meta-knowledge (studied using what McCarthy (2008) calls "The Designer Stance").

This is not meant to be an exhaustive classification of types of theory about relationships between genetic influences and learning. After writing a first draft of this paper I found that the theory in Karmiloff-Smith (1992) is somewhere between EMN and DNM, as is Piaget's theory of development.

I attempt to show that an extended version of McCarthy's DNM approach, linking research in biology and AI, applied across species, can lead to deeper and more powerful explanatory theories of information processing in organisms than the alternatives, and may also provide new ideas about both *requirements* for future intelligent machines and possible new *designs*.

1.1 The "tabula rasa" theory

For some who do not study human infants closely[3] it seems plausible that human learning is based on a TRLM, since neonates appear to be almost completely incompetent and ignorant, to

[1] See URL http://en.wikipedia.org/wiki/Tabula_rasa

[2] `http://www.cs.bham.ac.uk/research/projects/cogaff/misc/simplicity-ontology.html.html`

[3] As (E. J. Gibson & Pick, 2000) (Rochat, 2001) (Karmiloff-Smith, 1992) (Spelke, 2000) and Piaget, Vygotsky and many others have done

start with. A clue that something is wrong with this view comes from the observation that so many species produce offspring that either have to fend for themselves completely, or at least have to keep up with adults.

For instance, without being taught: birds extricate themselves from the egg when ready to hatch, young ducklings identify a nearby moving object, with some adult-duck-like qualities, to imprint on and follow around, many invertebrates live two entirely different lives, e.g. one larval and one as flying insect, and do not have to *learn* how to live either type of life, nor how to transform themselves from one form to another, the young of some grazing mammals are able, within minutes or hours, to get up, walk to the mother's nipple, suck and, in some species, soon after that run with the herd, e.g. to escape predators: displaying visual and other competences no current robot comes even close to.

If so much competence assembled over evolutionary time-scales can be transmitted (presumably mostly encoded in genomes, or genomes plus their epigenetic environment, e.g. features of the mother's womb) to members of so many "precocial" species, is it possible that people who think humans are born incompetent and ignorant are missing something very important? Some alternatives have been proposed.

2 Alternatives to a Tabula Rasa

Various alternatives to the TRLM theory have been proposed, a subset summarised here very briefly, including the one I believe is of most importance for both AI and biology. The hypotheses differ according to what they assume the products of biological evolution available from birth or hatching are, what those products are thought to do after birth or conception, and what they assume about how we can find out – a topic in philosophy of science. The descriptions below are necessarily over-simple.

The rationalist nativist thesis (RNT): There is a very old view among philosophers, including Plato, Descartes (and perhaps Kant), that nothing can have experiences and acquire knowledge empirically if it does not have certain kinds of concepts and knowledge to start with. Moreover, unlike the more recent empirical nativists, these rationalist nativists believe it is possible to work out from "first principles" what the innate knowledge must be, e.g. starting from conceptual analysis, without doing empirical research to find out what the innate knowledge actually is.

The empirical nativist thesis (ENT): During the last two decades an empirical hypothesis strongly opposed to the "tabula rasa" theory has emerged among developmental psychologists, namely that there is a considerable amount of *innate* knowledge in humans though its presence is not immediately obvious, and its content cannot be discovered without empirical research (contrast RNT). So an "empiricist nativist" industry, led by Elizabeth Spelke, among others, has been probing infants at various stages in their early development for signs of "core concepts", innate competences, etc.

Empirical meta-knowledge nativism (EMN): A more sophisticated theory-driven type of research, often inspired by some of Noam Chomsky's ideas[4], assumes not that the neonate's

[4]E.g. see http://en.wikipedia.org/wiki/Principles_and_parameters

knowledge displayed in successful behaviour is innate, but that some highly schematic meta-knowledge, about a class of environments, is innate, which is instantiated to specific knowledge and competences through interaction with the environment after birth or hatching. The knowledge the child actually needs in its life is derived by transforming the schematic (though not environmentally neutral) forms to substantive instances by acquiring parameters from the environment to produce the required instantiations. A special case of this is the Chomskian thesis that individual human languages differ in detail but all are instantiations of a universal *language schema* specific to humans, and encoded in their genome.

The diversity of human languages, all presumably products of some shared innate human-language learning competence (a set of hypothesised "language universals") is sometimes taken to support this, though not all researchers agree with Chomsky that the genetic component that gives rise to language is language-specific. (Results of research on innate language universals remain controversial.)

Generalising Chomsky's idea to include visual competences, motor competences, learning competences, reasoning competences, social competences, and others, leaves open the possibility that human language learning does not have its own genetic specification, but is based on a combination of other more general generic schemata.

Generalised EMN can inspire empirical research on humans and other animals to find out just what the innate, parametrisable, meta-knowledge is, as opposed to ENT research aimed at finding out what the innate concepts and knowledge of a neonate are.

On this sort of view (Sloman & Chappell, 2005; Chappell & Sloman, 2007) the innate knowledge, though abstract and missing detailed parameters is not totally general: it is suited to development by certain forms of interaction only in a restricted class of environments, a class that generalises features of the environments in which the species evolved. Neisser (1976) seems to share this view, since he chastises psychologists for ingoring the environment in which our brains evolved. For example, the innate learning mechanism in humans might be incapable of driving learning in a baby whose sensory input from the environment is restricted a 2-D TV display of the view from a mobile camera moved by someone else. Compare the experiments on visual learning in kittens in the classic paper (Held & Hein, 1963).[5]

Moreover innate *meta*-knowledge may be shared between species with shared evolutionary history despite current morphological and environmental differences.

Investigating details of the EMN mechanisms and innate meta-knowledge would require extension of the techniques of Spelke and others so as to expose not the *specific* innate concepts and knowledge, but the innately specified *patterns of development* of concepts and knowledge, as in Piaget's research.

Design-based nativist meta-knowledge (DNM): After learning of Spelke's work and reading (Pinker, 1994.), John McCarthy presented a related approach (originally 1996), sharing the assumption that "Evolution solved a different problem than that of starting a baby with no a priori assumptions." I'll describe features of his DNM approach that he did not make fully explicit, but

[5]"These findings provide convincing evidence for a developmental process, in at least one higher mammal, which requires for its operation stimulus variation concurrent with and systematically dependent upon self-produced movement. This conclusion neither denies nor affirms that other processes, such as maturation, occur concomitantly. The results demonstrate the complementarity of studies of adult rearrangement and neonatal deprivation."

are suggested by what he wrote.

He starts from the observation that products of biological evolution must have been shaped to a considerable extent by features of the environments in which they evolved – which may have had some constant and some changing characteristics. So, instead of (a) trying to derive the nature of the innate component from totally general abstract principles (e.g. principles of rationality, as in RNT), or (b) merely trying to discover it empirically by observing as many as possible of the exercises of competence in very young children (as in ENT and EMN), McCarthy proposes (c) that we should *also* study the environment or environments in which a species and its precursors evolved, trying to **work out**, on the basis of general features that are common to a range of relevant environments in which the evolutionary process occurred, the requirements for being able to do various things in those environments, e.g. feeding, learning, reproducing, etc. This process has much in common with the process of *requirements analysis* in engineering, which needs to precede and feed into the design process, though once begun each can influence the other.

On the basis of analysis of requirements for functioning in a range of environments, we can develop explanatory mechanisms designed to meet those requirements and test them by building working models. In practice, of course, the discovery of requirements and the process of design and testing have to be done in parallel with much mutual influence: since some requirements only become apparent after partially working designs have failed. Requirements analyses, like designs, can include bugs!

The scientific investigation based on DNM therefore overlaps with the engineering activity of trying to design various machines that perform and learn in such an environment. McCarthy used the label "the designer stance" to characterise such research: it is related to, but slightly different from what Dennett called "the design stance", which does not require actually doing design and implementation.

DNM differs from EMN insofar as it relies more on analysis of information-processing demands of features of the environment and (initially) less on empirical observation of very young humans and other animals, while not discarding empirical observation, since that is needed to test the theories developed. Building *working* instances of theories to check whether they can in principle do what is claimed provides a much deeper test than formulating and testing predictions from a theory specified only using words and diagrams. Moreover, creative designers will usually come up with alternative, competing, theories about innate meta-competences requiring empirical research to choose between the theories. DNM therefore leads to a mixture of creative design and empirical research, required for building machines that work in complex, varied, and changing environments. Some of the empirical research feeding into this work will investigate features of the environment that both influenced biological evolution and also support the developmental processes driven by evolved mechanisms.

3 DNM Research in AI and Biology

Compared with the purely empirical approaches of ENT and EMN, and the purely philosophical arm-chair approach of RNT, using the designer stance seeking DNM theories can be a powerful source of new, testable, ideas about what evolution might have provided, especially if researchers doing that research have the benefit of real experience of designing, building, testing and debugging working systems that provide insights into unobvious features of the problems evolution may have solved. (Most philosophers, linguists, psychologists, neuroscientists and ethologists have not had any such experience.)

Some psychologists, e.g. Ulric Neisser and James Gibson, also emphasise the need for research on cognition to take account of details of the environment. Gibson's ecological approach analysed features of the environment that provide opportunities and problems for perceivers who act in that environment, but ignored requirements for planning and reasoning, and he did not build working models. Piaget's research on development is not concerned with innate competences but with *developmental processes*, as in EMN. He attended to detailed, but generic, features of the environment which pose challenges and opportunities for learners (and by implication, evolution). Very late in life he recognised that AI could have made a major contribution to his work.

In the terminology of Lakatos (1980), I conjecture that DNM is more likely than the others to produce strongly *progressive* (as opposed to *degenerating*) research programmes.

The implications of McCarthy's DNM approach are not all obvious, and carrying out the project is difficult. We can distinguish at least the following aspects, some of which go beyond what McCarthy himself wrote.

• Different species can share environments, including species that interact with one another (e.g. predators and prey, or animals that have a symbiotic or competitive relationship). This can give clues as to how features of the environment pose information processing problems whose solutions generalise across species of different forms. (E.g. birds, primates, elephants and octopuses can perceive and manipulate objects in their environment, in order to achieve goals.) Animals with very different morphologies and competences can share evolutionary histories, may share common evolutionarily old sub-mechanisms, forms of representation, and ontologies, all extended in different ways by more recent evolution and individual development. Environmentally driven convergent evolution can also occur.

Looking at similarities and differences between such species may help researchers to separate out design *requirements* from design *solutions*. Without that cross-species investigation theorists may tend to conflate the specifics of the designs they propose with the problems solved by those designs, failing to think of alternative designs. So DNM-inspired research should study and compare more different species than researchers in AI and psychology normally do.

• DNM requires specification not only of *actual* features of environments, but also what *can* exist, about which animals or robots may need to acquire information, including: things that might need to be perceived, thought about, acted on, created, destroyed, used, eaten, avoided, communicated with, etc., even if few individuals encounter them. We may need to develop analogues of grammars, for structures and processes. E.g. relations can hold not only between complete objects but also parts of objects (within and between objects) like hands and things they are close to. When things change, these "multi-strand relationships" between complex objects can change in parallel, producing "multi-strand processes".

• Most organisms move much of the time and are surrounded by things in motion as well as things that are static. Consequently it is likely that from the very beginnings perceptual subsystems had the function of acquiring information about *processes* occurring in the environment rather than just *objects* and *situations*, unlike much research on perception in AI and psychology, which starts from perception of static entities, with the intention (if the need is considered) of adding motion later. For an organism a static scene may just be a special case of a process!

• One of the features of the environment that some animals need to be able to cope with is the fact that spatial and spatio-temporal structures and processes exist on very different scales, where the smaller scale entities are embedded within larger structures. For example, someone peeling a banana may or may not need to relate that action, and its consequences to other things in the field of view, or to other things in the immediate environment that can quickly be brought into view by small movements, or to other things in the larger environment that can only be perceived by large scale motion of the whole agent (e.g. to another part of the building, perhaps where garbage goes, or a fruit shop in another town). In addition to relating objects to larger containing structures it is often necessary to relate events and processes to larger containing temporally (or spatially and temporally) extended events and processes, for instance when relating the current situation to possible future plans, or possible explanations (causal histories). Such spatio-temporal nesting can often be represented without use of a global coordinate systems. (See below.)

• A thorny issue concerns how metrical aspects of spatial structures and relationships should be represented. Most researchers (unaware of Piaget's research results) assume that visual mechanisms have access to the kind of global cartesian coordinate system that engineers and scientists now take for granted, forgetting that this was a relatively late development in human culture. Most organisms don't have access to a global measurement system, and integrating local measures is notoriously unreliable – often forcing researches to rely on manipulation of probability distributions.

It is not surprising that confusions in children suggest that architectures for perceiving and using spatial information develop in stages, possibly creating coherent global systems of spatial relationships only at a relatively late stage. Most humans don't seem to use global cartesian coordinate systems with locations, motions, orientations, sizes, etc., represented by vectors of real numbers – except after extended education in modern schools and universities. Can any other species do that? What alternatives exist?

• We need to study animal environments to see what alternatives could be useful, e.g. perhaps mixtures of topological structure (containment, overlapping, touching, and various kinds of discontinuity in structures and processes) augmented by networks of partial ordering (of relative size, relative distance, relative angular magnitude, relative curvature, relative speed, etc.) The details get very messy, especially if the need to represent multi-strand processes is taken seriously, but I conjecture that this approach is capable of avoiding some of the fragility and other problems caused by imprecision and noise in sensory mechanisms. For example, imprecise information about boundaries of a pair of objects may not matter when the question is whether one of the boundaries encloses the other object. Perhaps brains evolved with mechanisms for propagating constraints in networks of partial orderings. Turning this idea into a demonstrable functioning design will be difficult.

• Pursuing these ideas may demote the importance of probabilistic inference in perceiving and acting in the environments encountered by many animals. That's not to deny that metrical precision and probabilistic mechanisms are never important. Metrical precision is required for throwing things at small targets, for grasping small objects quickly, and for a cat jumping from the ground to land on the top of a narrow wall. However, in many other cases the need for metrical precision in perception and motor control can be eliminated by continuous visual and other servoing, and use of abstract qualitative (e.g. topological) relationships.

• These capabilities are relevant to perception as well as to thinking, reasoning, planning, and remembering. A great deal of work on visual perception focuses on perception of objects (and how they are represented) ignoring perception of processes (including multi-strand processes) and perception of possibilities for processes and constraints on such possibilities. (Gibson, like many researchers on embodied cognition, focused only on a small subset of these, namely the affordances for the perceiver.)

Some perceptual information is about transient, fast-changing, continuously changing, states and processes used in servo-control, and and some is about relatively enduring structures and relationships on various scales (e.g. a particular cave, the immediate environment of the cave, the larger terrain containing that environment, etc.) and some will be a mixture of changing details while some high level process structure persists (e.g. a lion chasing a deer). Not all organisms need all of this.

Likewise, some organisms need only information about how things actually are, to which they can react immediately, while some will be about what is possible (what changes can occur in any given configuration), some will be about constraints on the possibilities (e.g. laws, necessities), and some about branching sets of chained sequences.

• One of the tasks for a designer is to specify the various forms of representation (types of syntax, or types of variability, for information-bearing structures) in which the information can be acquired, stored, manipulated, combined, derived, analysed, and used, as well specifying innate and learnt ontologies capable of generating the required diversity of information contents, using those forms of representation.

Most researchers have experience of far too few forms. Many don't even realise this is a problem, either because they have never implemented a working system or because they have been taught to accept only a narrow range of forms of representation (e.g. only vectors of numerical values and probability distributions over them, or only logical forms, depending on their background).

We know very little about the forms of representation that allowed humans to investigate Euclidean geometry long before modern logic and algebra, or the cartesian mapping between geometry and arithmetic, were developed. Further investigation of problem solving and reasoning in other animals, e.g. nest-building birds, dam-building beavers, carnivores that not only hunt (sometimes collaboratively) and kill their prey, but also dismember them in order to get to edible parts, may or may not suggest connections with human spatial reasoning. McCarthy was mostly interested in logical (Fregean) forms of representation, though he briefly mentions the possibility of forms of representation whose instances are specially tailored to, but not isomorphic with, spatial structures and processes which they are used to represent. Impossible objects drawn by Reutersvaard, Penrose and Escher (which young children don't see as impossible) may give important clues.

• There are especially deep and difficult problems with a long philosophical history concerning the nature of causal knowledge of various kinds. Jackie Chappell and I have argued that animals sometimes need Humean (correlational) forms of causal information and sometimes

Kantian forms, where causal reasoning is structure based and non-probabilistic, e.g. like mathematical reasoning in geometry (Sloman, 2007).

• Different "kinds of stuff" can occur – varieties of material of which things can be composed. Human-like agents will need an ontology that includes not only object parts, surface fragments etc., but also materials constituting clothing, food, furniture, nappies, towels, tissues, cotton wool, sponges, water, then later things like mud, sand, plasticine, elastic bands, different sorts of paper, stiff and bendable metal, and various parts of plants and other animals. Materials affect both perceptual features and dispositions to produce or resist *processes* in which things change their properties and relationships in different ways. How can different kinds of stuff and their properties (including "invisible" dispositional properties, like brittleness) be represented in the minds of (novice or expert) humans and other animals, and how should they be represented in future robots? What innate mechanisms could use physical exploration to drive construction of such an ontology? (Pat Hayes' "naive physics" project may be due for a revival in a dynamic form.)

Some researchers assume that the methods of simulating physical processes in computer games or in rendering simulated scenes can be used. But those forms of representation are not designed to enable an active agent to perceive, think about, act in and learn from the environment: their functions are much more restrictive (graphical display functions and predictive simulation). Forms of representation that are useful for generating synthetic movies are not necessarily useful for representing possibilities for action and reasoning about consequences. Simulation tools will often produce results at the wrong level of detail. When deciding how to travel to a conference, I cannot, need not, and usually should not, create videos of all the detailed steps in the processes of travel.

• **Other agents:** For some organisms, the environment includes not only inert or passively moved physical entities and processes, but also things that process and use information to select and execute actions – i.e. *agents*. So, besides the semantic competences required for coping with the former, some animals (and future robots) will need meta-semantic competences for representing and reasoning about semantic competences and their uses in the latter – i.e. for perceiving, thinking, reasoning about, and interacting with other thinking agents, while handling referential opacity.

This may, or may not, include self-referential meta-semantic competences, e.g. the ability to be self-conscious regarding one's own experiences, thoughts, plans, decisions, etc. (Which comes first, self or other representation, may be a chicken and egg question: perhaps they developed in parallel, serving different but overlapping functions.)

• Virtual *vs.* physical machinery: Some people who are not familiar with, or have not understood, advances in tools and methods for representing information and controlling processes over the last 60 years or more, assume that all the information-bearing structures must be physical: and seek evidence for them from neuroscience. (Compare: Newell and Simon's "Physical symbol systems".) This ignores the possibility of increasingly abstract and flexible *virtual* machinery containing more powerful forms of representation implemented in, but quite different from, the chemical and neural structures and processes in brains: both "application virtual machines", that have some specific information processing function (e.g. parsing sentence structures) and "platform virtual machines", capable of supporting a wide variety of application VMs and possibly also new platform VMs.

• For handling complexity human engineers find it useful to design monitoring and control mechanisms operating at the virtual machine level. Perhaps biological evolution "discovered" that trick much sooner, using it for important control functions and also allowing idle introspection. That might account for phenomena generating debates about consciousness and qualia – in humans and future robots.

• As McCarthy notes in discussing planning, competences of both pre-verbal children and many non-human animals require the ability to make use of forms of representation that share some of the features previously thought to be unique to human language, including structural variability, varying complexity of forms, decomposability and recombinability of parts, compositional semantics (particularly context-sensitive compositional semantics), and the ability to support inference also mentioned in (Sloman, 1979).

Despite sharing those features with human communicative languages (and also logical, algebraic, and programming languages), these biologically useful internally used forms of representation could be very unlike human language both in function, since their main use is not for communication between separate individuals, and in their form, insofar as they are not all composed of linear sequences of discrete elements. (In the 1960s, for instance, some researchers on vision investigated "web grammars", for graph-like visual structures.)

There may be deep relationships between the forms of manipulability of the information structures and the forms of manipulability of objects in space and time. Yet the representations cannot be isomorphic with what they represent: since many uses of information are different from the uses of what is represented. (E.g. information about food and hiding places need not be edible, or provide shelter from rain.)

• For an animal or robot, not only factual information about the environment needs to be represented: There also need to be formulations of information gaps (questions to be answered), goals, plans, preferences, values, hypotheses, experiments, and many kinds of control information, some of it transient in servo-control systems, others more abstract and enduring, e.g. about dangers to be avoided, preferences to be followed, etc. (Some theories about ventral and dorsal visual streams fail to take account of these different functional requirements.)

Much of what needs to be controlled is not externally visible behaviour but internal information-processing behaviour, including controlling the perceptual interpretation of sensory inputs, controlling the competition between inconsistent goals or preferences, or sometimes selecting between planning options on the basis of unusual requirements for sensitivity to changing features of the environment during plan execution. We need to investigate mechanisms that can *grow* such capabilities.

• The design of mechanisms required for development and learning requires far more to be achieved than simply externally observable (and rewardable) changes in behaviour. Often there are deeper developments hidden from external view, including new forms of representation, new ontologies, new forms of storage and searching, new reasoning algorithms, new conflict resolution strategies.

Sometimes this requires changing the structure of an information-processing architecture, such as adding a whole new subsystem, or modifying connections between sub-systems. An example of architectural revision seems to be the transition from pattern-based language use to grammar-based use, which causes children to start making errors because they don't yet have an architecture that copes with exceptions to the rules – which requires yet another architectural change. I suspect there are many more such architectural transitions that have gone unnoticed in developmental psychology (though Piaget collected evidence for some).

• A problem for many researchers is understanding the possibilities for *development* of the forms of representation, algorithms, architectures, uses of information, and specific knowledge of the environment. Many AI researchers assume that the forms of representation available and the architecture are fixed from the start – which clearly is not the case in humans, as Piaget noted a long time ago, though he lacked the conceptual tools required to express good explanatory theories. Unfortunately, most researchers have not personally encountered the diversity that has been explored in AI in the last half century, partly because of factional warfare (and shortage of time) that leads senior researchers to teach only the types of mechanism and forms of representation that they happen to like or be familiar with. The DNM approach has important implications for education in AI.

4 Investigating requirements

Much more work is needed to separate discussions of *designs* for working systems from analysis of the *requirements* to be met by such designs – which may be partly similar and partly different for different species, and for a particular individual at different stages of development. Some are requirements for physical features and sensors, but the most subtle requirements concern information-processing capabilities, mechanisms, architectures, forms of representation and ontologies, including requirements for self observation and self control. (Compare McCarthy on "Making robots conscious of their mental states", 1995, Minsky "Matter mind and models", (1968), Sloman 1978 Chapter 6).

Often it is assumed that requirements are clear and can be stated briefly, leaving only the problem of producing a design. When we are discussing systems whose functionality is the result of evolution over long time scales in which many different kinds and layers of functionality have been provided in a single design, the requirements that led to that design are generally far from obvious.

An example is the need to categorise different classes of use of information. Karen Adolph referred to the "on-line intelligence" involved in infants and toddlers interacting with some complex and changing situation, such as walking across a narrow bridge, putting on clothes, pursuing escaping prey. That can be contrasted with "off-line" uses of information about an environment with which the individual is not currently interacting and may not soon interact. The off-line uses of information about the environment tend to go unnoticed by researchers who emphasise the importance of embodiment (Sloman, 2009).

5 What next?

If we take the Design-based Nativist Meta-knowledge (DNM) hypothesis seriously, perhaps the "AI as science" research community can collaborate on systematically investigating the relevant features of environments in which evolution and development occurred, the requirements implied by various features of the environments and the possible types of solution that could work for various subsets of the requirements, producing a large-scale, systematic, collaborative project.[6] Research papers would then not merely describe a system and what it does, but relate it to a requirements-space and a design-space. I suspect this is a hitherto unnoticed requirement for completing the human genome project.

References

Adolph, K. (2005). Learning to learn in the development of action. In J. Lockman, J. Reiser, & C. A. Nelson (Eds.), *Action as an organizer of perception and cognition during learning and development: Minnesota Symposium on Child Development* (Vol. 33, pp. 91–122). Mahwah, NJ: Erlbaum.

Chappell, J., & Sloman, A. (2007). Natural and artificial meta-configured altricial information-processing systems. *International Journal of Unconventional Computing*, *3*(3), 211–239. (http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0609)

Gibson, E. J., & Pick, A. D. (2000). *An Ecological Approach to Perceptual Learning and Development*. New York: Oxford University Press.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

Held, R., & Hein, A. (1963). Movement-produced stimulation in the development of visually guided behaviour. *J. of Comparative and Physiological Psychology*, *56*(5), 872–876.

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

Lakatos, I. (1980). Falsification and the methodology of scientific research programmes. In J. Worrall & G. Currie (Eds.), *Philosophical papers, Vol I* (pp. 8–101). Cambridge: Cambridge University Press.

McCarthy, J. (2008). The well-designed child. *Artificial Intelligence*, *172*(18), 2003-2014.

Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Neisser, U. (1976). *Cognition and Reality*. San Francisco: W.H. Freeman.

Pinker, S. (1994.). *The Language Instinct*. Morrow.

Rochat, P. (2001). *The Infant's World*. Cambridge, MA: Harvard University Press.

Sloman, A. (1979). The primacy of non-communicative language. In M. MacCafferty & K. Gray (Eds.), *The analysis of Meaning: Informatics 5 Proceedings ASLIB/BCS Conference, Oxford, March 1979* (pp. 1–15). London: Aslib. (http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#43)

Sloman, A. (2007, Sept). *Understanding causation in robots, animals and children: Hume's way and Kant's way.* Paris. (Presentation at CoSy Project MeetingOfMinds Worksh)

Sloman, A. (2009). Some Requirements for Human-like Robots: Why the recent over-emphasis

[6]This web site `http://www.cs.bham.ac.uk/research/projects/cogaff/talks/` is an attempt to illustrate the diversity of this investigation.

on embodiment has held up progress. In B. Sendhoff, E. Koerner, O. Sporns, H. Ritter, & K. Doya (Eds.), *Creating Brain-like Intelligence* (pp. 248–277). Berlin: Springer-Verlag.
Sloman, A., & Chappell, J. (2005). The Altricial-Precocial Spectrum for Robots. In *Proceedings IJCAI'05* (pp. 1187–1192). Edinburgh: IJCAI. (http://www.cs.bham.ac.uk/research/cogaff/05.html#200502)
Spelke, E. (2000, November). Core Knowledge. *American Psychologist*, 1233–1243.
Turing, A. (1950). Computing machinery and intelligence. *Mind*, *59*, 433–460. ((reprinted in E.A. Feigenbaum and J. Feldman (eds) *Computers and Thought* McGraw-Hill, New York, 1963, 11–35))