Requirements for a Fully-deliberative Architecture
(Or component of an architecture)

This is a contribution to the study of the Space of Possible Minds:
http://www.cs.bham.ac.uk/research/projects/cogaff/sloman-space-of-minds-84.html

Aaron Sloman

(With thanks to Dean Petters and other colleagues.)

This discussion paper should remain accessible here:

TITLE: Requirements for a Fully Deliberative Architecture
(Or component of an architecture)
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/fully-deliberative.html (HTML)
It was originally a CoSy Project discussion paper REF: COSY-DP-0604
http://www.cs.bham.ac.uk/research/projects/cosy/papers
A PDF version (which may slip out of date if this paper is revised) is here:
http://www.cs.bham.ac.uk/research/projects/cogaff/misc/fully-deliberative.pdf

Originally installed: 2006 (After a conversation with Dean Petters)
Some major updates occurred on: Jan 2009; 1 Sep 2011; 4 Jan 2014
(see also: recent update list below, and updates before 2009 at end of document.)

Previous location:
http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0604 (HTML)
http://www.cs.bham.ac.uk/research/projects/cosy/papers/fully-deliberative.pdf (PDF)

CONTENTS

[0] Updates since 2009

Update 3 Sep 2016: Re-formatted
Update 3 Jan 2014:
Re-formatted, made some minor corrections and improvements, and attempted to clarify connection between discretisation and invariants.
Update 6 Jan 2010:
Early versions of this document completely ignored cases where the results of actions can be uncertain (e.g. instead of turning the key definitely making it possible to open the door, it may merely reduce one obstacle, if the paint of the door has stuck, or there is furniture behind the door.) For a reference see below.

Updated: 24 Jan 2009 -- online/offline creativity Two biologists, Jackie Chappell and Susannah Thorpe drew my attention to this paper by Karen Adolph

http://www.psych.nyu.edu/adolph/PDFs/MinnSymp2005.pdf
Learning to Learn in the Development of Action.
In *Action As An Organizer of Learning and Development:* Vol 33 in the Minnesota Symposium on Child Psychology Series, 2005
Eds. John J. Rieser, Jeffrey J. Lockman, Charles A. Nelson

The paper emphasises the ability of infants and toddlers to learn to cope with ever more complex and demanding physical situations by learning new ways to take creative decisions by extending what they have previously learnt. She calls this ’online creativity’, dealing with ’online novelty’.

This is very important work, but apparently she has not noticed the possibility of children learning to be creative ’offline’, e.g. learning to reason several steps ahead, about what would happen, or reasoning hypothetically about the past, what could have happened, or what would have happened if something had occurred, or making a plan for action involving several future steps.

Dealing with online novelty involves performing a complex action incrementally: as each sub-step is actually performed, a subsequent sub-step is selected from the possibilities that then become available (sometimes from a continuous range of possibilities rather than a discrete set).

[Paragraph updated 4 Jan 2014]
This does not require the information processing system to be able to represent **multiple** branching sets of future states and actions. There is only one set of (short) branches at any time, the set of possible next things to do -- where the agent is already physically poised to select and do one of them. Nothing further into the future than the next step is explicitly considered. Similar comments could be made about *continuous* control processes, where there are no next steps, e.g. in negative feedback loops -- homeostasis, such as temperature control, or steering towards a target by continually aiming at it. Homeostatic mechanisms, e.g. for controlling temperature, or osmotic pressure, occur in many organisms that would not normally be regarded as intelligent. Continuous control problems are also pervasive in control engineering. They were also discussed by J.J.Gibson, W.T.Powers, Norbert Wiener, and in Alain Berthoz’ book *The Brain’s sense of movement* (2000).

There are things to be said about *intelligent* vs *unintelligent* continuous control (e.g. different ways of using brakes when approaching a stopping point). There is also a difference between a system that uses the more intelligent form of control because it was programmed or trained to do it, and one that understands that there are alternative forms of control and understands why some are better than others. This might be described as *deliberate reactive* control.

The roles of intelligent meta-cognition, and the mechanisms that make it possible, need a much bigger document than this one.

The situations that led to the title of this paper, namely situations that require "fully deliberative" competences, are computationally more sophisticated than typical continuous control problems, because they require envisaging sequential choices in *branching possible future* situations, where the choices could lead to new choices. However, because the future possibilities are dealt with at a high level of abstraction, e.g. often ignoring details of physical interactions, the computations may be much less costly than detailed simulations would be, though not necessarily less costly than physical trial and error. Examples and further analysis are offered below.

Perhaps 'anticipatory' could be contrasted with, or combined with, 'online' in this context.

A book by developmental psychologists that does not make these distinctions, but is otherwise excellent, is:

> Eleanor J. Gibson, Anne D. Pick, *An Ecological Approach to Perceptual Learning and Development,* Oxford University Press, 2000.

Updates before 2009 were moved to the end of the document on 4 Jan 2014.

[1] Background (Updated 4 Jan 2014)

For several decades, researchers in AI and Cognitive Science have talked about animals or machines as having (or not having) 'deliberative' capabilities and using (or not using) deliberative mechanisms. In my own work, I have, since collaborating with Luc Beaudoin (whose PhD was completed in 1994), been contrasting 'reactive', 'deliberative' and 'meta-management' (sometimes referred to as 'reflective') capabilities (all of which are categories within which many further subdivisions are possible). Some related distinctions were made in my 1978 book, especially chapter 6 (deliberative and executive loops) and chapter 10 (including discussion of 'central administrative processes'). Students and colleagues on the Cognition and Affect project, and its successors have all contributed to these ideas. See:
http://www.cs.bham.ac.uk/research/projects/cogaff/phd-theses.html
http://www.cs.bham.ac.uk/research/projects/cogaff/crp/#chap6
http://www.cs.bham.ac.uk/research/projects/cogaff/crp/#chap10
Overview of the CogAff project (started 1991):
http://www.cs.bham.ac.uk/research/projects/cogaff/#overview

What is a deliberative system?

The key feature of a deliberative system is the ability to represent and reason about, and to compare and evaluate (for some purpose), possible situations that do not exist, or which could have existed but did not, or exist but are not known to exist, either because they are future possibilities, or because they are remote or hypothetical possibilities or because they occurred in the past.

Some deliberative systems are restricted to considering only possible next steps, while others can consider several possible future steps. Some of the latter can cope with branching possible futures, whereas others can only consider linear sequences of possible futures based on stored plans with built-in expectations about the consequences of each step. These plans may or may not be parametrised.

E.g. a fixed plan to get cheese from the refrigerator in the kitchen when starting in the living room may, possibly after some learning, of new plans, be replaced by a parametrised plan to get X from the refrigerator in the kitchen starting from room Y. The parametrised plan may be applicable in different situations even if it has fixed steps (e.g. go to hall, go to kitchen, go to refrigerator, open door, etc.)

> *A fully-deliberative system is able to construct representations of possible states of affairs of varying structure and varying complexity, using at least one formalism with compositional semantics, in mechanisms that allow two or more such structures to be constructed, analysed and compared, where the result of comparing them may be another complex structure describing the pros and cons; and that, in turn, may or not be capable of being used to take a decision to select one of the options.*
>
> *The purposes for which alternative possibilities are compared and evaluated include selecting an explanation for something observed, selecting a future plan of action, choosing between alternative ways of interpreting evidence, making a prediction, designing something.*

This is not intended as a formal definition -- just a rough indication of a complex kind of functionality, described in more detail below. One of the important points is that there is a wide spectrum of competences and not all researchers have noticed the diversity documented here. For example some use the word ’deliberative’ to refer to very simple examples, which I have called ’proto-deliberative’ in contrast with with ’fully-deliberative’ systems at the other end of the complexity range. I suspect most people who use the word ’deliberative’ refer to some intermediate point in the complexity spectrum. So there is great confusion in current terminology.

In view of the variety of uses of ’deliberative’ among researchers, mentioned below, the labels

- *’selective deliberation’*
  (evaluation and selection from a set of available options),
  and
- *’constructive deliberation’*
  (selection from a set of options constructed as part of the deliberation process: i.e. with interleaved construction, evaluation, comparison, selection throughout the process of deliberation)

might be more appropriate than ’proto-deliberative’ and ’fully deliberative’.

However ’constructive deliberative’ is a bit of a mouthful. We could use ’s-deliberative’ and ’c-deliberative’ as abbreviations for the selective and constructive types of deliberation. For now the important task is not to define terms, but to understand the space of possibilities: the variety of types of capabilities that may or may not be present in various types of systems that people have been inclined to call ’deliberative’.

> Very little of what follows is new: all the main ideas go back to work by Minsky, McCarthy, Green, Simon, Evans, Winston, and many others during the 1960s and early 1970s.
>
> Apart from early AI researchers there have been others who understood at least the general points made here, e.g. Russell A. Barkley in his 1997 book *ADHD and the nature of self-control.*
>
> Also Arnold Trehub in his book *The Cognitive Brain* (MIT Press, 1991) reviewed here (by Luciano da Fontoura Costa). This book is noteworthy for its elaborate attempts by a neuroscientist to specify realistic mechanisms to support the capabilities of a deliberative system as part of a larger design. (As far as I know nobody has ever tried to implement Trehub’s specification, and I suspect that some parts of it will need to be extended to accommodate all the requirements given below. However it is fair to say that nobody has attempted to implement the full set of features listed below.)
>
> Although I have taken all that for granted for many years, gradually I have come to realise that the ideas are not all widely understood, and the word ’deliberative’ is used in different ways, partly because people have not analysed the variety of cases in a deep way that is widely shared.

In the rest of this document I’ll contrast what I have been calling ’fully deliberative’ (c*(onstructive)*-deliberative) systems with much simpler kinds of ’proto-deliberative’ (s*(elective)*-deliberative) systems, while allowing for many cases in between (including intermediate states through which evolutionary trajectories have passed).

In a complex architecture with many components there are different kinds of subsets, and in my work I have characterised three (partly overlapping) main subsets, which differ in their evolutionary history, in their spread amongst other animals besides humans, and in their functionality (though they may overlap in the kinds of mechanisms they use). Marvin Minsky introduced further subdivisions, leading to six layers, in his 2006 book *The Emotion Machine.* (Draft online on his web page: http://web.media.mit.edu/~minsky)

There are other ways of dividing up the components of an architecture, using different criteria, as we'll see. This document does not introduce all the main components and subdivisions. For example linguistic competences require mechanisms in all the layers, interacting with other mechanisms.

Some Related Distinctions

A related but different distinction can be made between

- systems that use ontologies that refer only to patterns and correlations (conditional probabilities) involving combinations of sensory and motor signals at various levels of abstraction (somatic ontologies)
- systems that use ontologies that refer to entities in the environment independently of how they are sensed or acted on (exosomatic ontologies).

For a tutorial introduction to these ideas and some observations and speculations concerning the development of the ability to experience exosomatic features of the world, see this PDF slide presentation http://www.cs.bham.ac.uk/research/projects/cosy/presentations/assc10-poster.pdf

and the associated discussion of the somatic/exosomatic distinction in relation to the notion of sensorimotor contingencies here.

Another related distinction is between use of a Humean concept of causation and Kantian concept of causation, discussed in the above PDF tutorial and in this presentation:
Two views of child as scientist: Humean and Kantian (PDF)

NOTES:

1. After writing most of this, I discovered that the phrase 'fully deliberative' which I thought I had coined as an unambiguous label for a system with a particular collection of competences (listed below) has been quite widely used in another way. There appear to be people who study varieties of distributed decision making who use the term 'fully deliberative' to refer to decision making that is completely centralised. E.g. in this paper 'RAVE: A Real and Virtual Environment for Multiple Mobile Robot Systems' (IROS 1999) Dixon et al define 'fully deliberative' thus:

   > Fully-deliberative systems exercise centralized control using a detailed world model, whereas fully reactive systems expect overall system properties and behaviors to emerge from individual robot behaviors that are not explicitly coordinated with one another.

   They also allow intermediate cases. Anyhow that is *not* what I mean by 'fully-deliberative', as explained below.

2. The term 'meta-management' was coined by Luc Beaudoin, while working on his PhD thesis, *Goal processing in autonomous agents (1994)*, Available online. I preferred it to the word 'reflective' which others had used because the latter often suggests something more passive, whereas what we call *meta-management* includes both internal self-observation and also control activities. Moreover, you can *reflect* on anything, including ongoing or recently performed external behaviours, which I regarded as part of the function of deliberative mechanisms -- which need to be able to monitor and learn from plan execution, for example. So Beaudoin's word usefully indicated the *inwardly* focused function.

   However, it might be argued that the label 'meta-management' has a connotation that is too narrow for our purposes, since we use it to refer not only to self-observation and control of things that would normally be described as 'management', but also self-observation of perceptual processes that would not normally be regarded as management processes, e.g. noticing that you have started hearing a noise that you had not previously heard, or noticing that what you are looking at has figure-ground ambiguity. An ambiguous figure might cause a perceptual system to flip between two interpretations without anything detecting that it is flipping. The ability to detect what is going on would be one of many functions of meta-management in its broadest sense.

(Compare McCarthy on 'Making robots conscious of their mental states'. He concentrates mainly on self-observation, as opposed to self-control or self-modulation, which is part of meta-management.)

One argument for using 'meta-management' to include monitoring and control of perceptual contents is that the monitoring of management should include monitoring influences on management (e.g. in order to evaluate responses to those influences) and that would include monitoring perceptual contents.

Hardly anyone outside the Cognition and Affect project in Birmingham uses 'meta-management' in the context of AI or Cognitive science, though it seems to be the name of a company and to be widely used in management theory! As far as I can tell by sampling examples thrown up by google, the management theory use of the word 'meta-management' is very similar to ours, but applied to organisations rather than individual animals or robots.

Minsky discusses similar ideas in his online book 'The Emotion Machine', e.g. in chapter 5. He distinguishes more kinds of 'reflective' subsystem than we do: that is not a disagreement, just an interest in different principles of subdivision. (E.g. I want to relate human architectures to architectures in other animals and to different evolutionary stages of development.)

3. **NOTE (6 Jan 2010) -- Dealing with uncertainty.** Early versions of this document completely ignored cases where the results of actions can be uncertain (e.g. instead of turning the key definitely making it possible to open the door, it may merely reduce one obstacle, if the paint of the door has stuck, or there is furniture behind the door.)

   For a pioneering book on adding considerations of uncertainty, in the form of probabilities, along with considerations of utility, to both deliberation about what to do, and meta-deliberation about how to think about what to do, see

   > S. J. Russell, E. H. Wefald, 1991, *Do the Right Thing: Studies in Limited Rationality,* MIT Press, Cambridge, MA,
   >
   > (Thanks to Richard Dearden for drawing my attention to this work.)

   Much has been written on that topic since then. A more complete version of this document should include discussions of uncertainty.

   I have discussed ways of avoiding dealing with probabilities by reasoning and planning at a higher level of abstraction in a discussion paper:

   > http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0702 Predicting Affordance Changes (Alternatives ways to deal with uncertainty) Nov, 2007

[2] Different uses of 'reactive'

It has been obvious for some time that the word 'reactive' is interpreted differently by different users. For example some people have restricted the label to systems that have no changeable internal state, so that at every moment the current input determines the current output and outputs are always the same for the same inputs. It is very surprising that anyone has ever believed that there are animals like that, given the extent to which living things are adaptive.

Others allow reactions to include changes of internal state, which in turn can cause future reactions to the same input to be different. Systems with hysteresis, would be examples. In my own work I have tended to use the word 'reactive' simply to refer to the absence of any deliberative capabilities. That allows reactive systems to have states, and to have internal cycles of reactive behaviours, as many dynamical systems do. But that sense of 'reactive' excludes systems that consider branching future possibilities, hypothesise about what might have been the case, or formulate hypotheses about what exists in some part of the world that is not currently being perceived.

[3] Different interpretations of ’deliberative’

Despite being aware of different uses of ’reactive’ I used to think everyone used ’deliberative’ in the same way, roughly as a label for many of the kinds of things symbolic AI systems have been doing at least since the early 1960s.

However, over the last 5 years or so it has become clear that the label ’deliberative’ is used with different meanings by different people.

In his online draft book, *The Emotion Machine* (Simon & Schuster, 2006), Minsky suggests (in chapter 5) that ’deliberative’ refers to the ability to select the best from a collection of alternatives -- a process that has been the subject of study in the theory of games and decisions for many decades. But he then goes on to make clear that it can involve much more than simply choosing from some fixed set of future actions, such as options in a payoff matrix, or values in a continuous interval. Rather, it may be necessary to explore a space of possibilities of varying complexity, e.g. possible future actions of varying complexity. Likewise a search for an explanation or a proof of something typically involves searching a space of possible explanations or proofs of varying complexity.

Indeed the ability to do that is exactly what was taken in the early days of AI to be characteristic of intelligence, and the need for it showed up in many problems, including planning, mathematical reasoning, interpreting complex images, finding explanations and playing various kinds of competitive games. Since then we have learnt that that is just one kind of intelligence, or one aspect of human and animal intelligence, but an important aspect with multiple facets.

For a long time I thought everyone used ’deliberative’ in the same way, namely to refer to this complex collection of capabilities, but it turns out that not everyone who uses the word has worked on or read about symbolic AI techniques concerned with planning, problem-solving, or explaining observed phenomena, so people often interpret the label ’deliberative’ differently -- often assuming a much simpler type of capability.

For example, I was surprised to hear Michael Arbib say at a conference in 2002 that a frog is capable of deliberation because in some situations it has two kinds of reactions R1, R2, triggered by two kinds of stimuli S1, S2, and when a stimulus Sm, intermediate between S1 and S2 is sensed, that can cause both R1 and R2 to be activated, requiring a competitive brain process to select one of them, which is then produced. E.g. S1 and S2 could be two kinds of moving objects, and R1 could be trying to catch the object while R2 is trying to escape from the object.

Likewise when Matthias Scheutz was working with me in 2000-2001 he wished to describe some simulated agents used in some of his evolutionary experiments as being ’deliberative’ because sometimes instead of heading straight for a target they could detect an obstacle and make a detour.

Another example turned up around 2001, when I met the biologist Nigel Franks, who studies ants and other insects. He claimed that a bee colony used deliberative capabilities when looking for a new place to start a hive. Many bees would go out hunting for good places to start the new hive, and if a location had good features several bees would be attracted to it. However, if other locations attracted more bees because they were more suitable, then bees would move from less popular to more popular locations. In the end almost all of them would be in one place which would be the location of the new hive.

> The portia jumping spider seems to be capable of something approaching planning capabilities, including working out a complex roundabout route in order to get to its prey (other spiders). There’s a more up to date very readable summary by John McCrone in the New Scientist (27 May 2006).
>
> Also M. Tarsitano, ’Route selection by a jumping spider (Portia labiata) during the locomotory phase of a detour’, *Animal Behaviour,* Vol 72, Issue 6, December, 2006 pp. 1437--1442, http://dx.doi.org/10.1016/j.anbehav.2006.05.007,

> However, no other non-human animal seems to come close to the capabilities of Betty the hook-making, puzzle-solving New Caledonian Crow whose movies are available here. (Alas Betty died recently.)

Another example is

> Holk Cruse 'The evolution of cognition--a hypothesis' *Cognitive Science* 27 (2003) 135-155

In which he claims

> In a reactive system the motor output is exclusively driven by actual sensory input{*}. An alternative solution to control behavior is given by "cognitive" systems capable of planning ahead. To this end the system has to be equipped with some kind of internal world model.{**}

{*} Note that this differs from my use of 'reactive' which allows changeable internal state (e.g. need for food or drink) to contribute to conditions triggering behaviours.

{**} In what follows I try to show that being 'capable of planning ahead' can cover a wide spectrum of cases, from very simple proto-deliberative systems to fully-deliberative (c-deliberative) systems. The paper by Cruse is interesting, but if 'cognitive' is taken to include fully-deliberative architectures then he discusses only a very small and relatively simple subset of cognitive systems. There's nothing wrong with that as long as it is made clear what is being left out.

A rich variety of competences in non-human animals is surveyed briefly in this review article:

> Nathan J. Emery and Nicola S. Clayton, The Mentality of Crows: Convergent Evolution of Intelligence in Corvids and Apes *Science,* 10 December 2004: Vol. 306. no. 5703, pp. 1903 - 1907 DOI REF.

However, like many other behavioural scientists they describe differences in capability at a vague common-sense level rather than in terms of design differences.

[4] This is not a debate about definitions

Note that I am not claiming that my use of the word 'deliberative' is right and others wrong: it is almost always silly to argue about the 'correct' definition of any word in the context of scientific debates. What is important is to get an overview of the range of phenomena that we are trying to understand and to find ways of dividing them up that lead to deep explanatory theories. (Unlike: earth, air, fire and water, or Lewis Carroll's walrus' list: 'shoes and ships and sealing-wax and cabbages and kings'.) The philosopher Gilbert Ryle used to describe this as exploring the 'logical geography' of a set of concepts. We can contrast this with analysing the 'logical topography' supporting a space of possible concepts.

Analysing logical geography and topography is far more important than arguing about definitions. (For an introductory overview of Ryle's ideas see Julia Tanney's paper Rethinking Ryle.)

It is perhaps surprising that nobody has presented me with an argument that water is deliberative because if a stream of water starts flowing down a gully and comes to a fork in the gully, where one branch soon meets a dead end (where ground rises) whereas the other branch goes on down the hill, the water will 'try' both branches, and most of it will select the better branch to flow down. Perhaps cases like that explain why Aristotle thought water had the goal of getting to the centre of the earth?

Cases like those mentioned above (though perhaps not the water example) led me to realise that within a system of the sort that I had been describing as 'purely reactive' it was clear that some things could occur that many people would describe as 'deliberative', undermining the supposed distinction between reactive and deliberative mechanisms. Now it is generally utterly pointless to argue about the 'correct' definition of a loosely used pre-theoretical (non-technical) concept that is inherently under specified, or the 'correct' definition of technical terms that have been used by different research communities in different ways. What is important is to understand the space of possible designs and what sorts of distinctions may be made within the space, and why such distinctions are or are not of theoretical or practical importance.

NOTE: This is connected with the distinction mentioned above between "logical geography" (relationships between concepts actually in use) and "logical topography" (features of the underlying reality that allow different logical geographies, or which lead to changes in the logical geography -- set of concepts used -- when the logical topography is better understood). We don’t yet understand the logical topography underlying our current set of concepts related to deliberation. (I was surprised to discover from Ursula Coope that Aristotle had related ideas about different levels and kinds of skill.)

[5] Proto-deliberative vs fully-deliberative

So I started using ’proto-deliberative’ to refer to mechanisms that could make selections between alternatives, without requiring what I had previously called a deliberative architectural layer. I contrasted proto-deliberative systems with what I began informally referring to as ’fully-deliberative’ systems.

It was clear to me that between the simplest forms of reactive architectures (microbes of various sorts, or even water running down a hill) and what I called ’fully-deliberative’ systems there must have been many intermediate cases during biological evolution. That is because fully-deliberative systems (as shown by AI research over several decades) may require many different components of varying complexity, and it is not likely that all of them evolved at once.

Studying all the kinds of transitions that are possible in such an evolutionary process would be an interesting research project for another occasion. For now I merely wish to explain what I mean by a fully-deliberative system. Thus I do not claim that there is an exhaustive dichotomy between proto-deliberative and fully-deliberative systems. Neither is there a *continuum* of cases: there may be many discontinuities in biological development, and anything involving changes in DNA *must* be discontinuous -- a point that is sometimes forgotten by critics of evolutionary theory.

It is not clear to what extent pioneering thinkers who first presented some of these ideas, e.g. (Kenneth Craik *The Nature of Explanation* (1943) and Karl Popper who wrote about the need for some organisms to be able to simulate possible futures so that their mistaken hypotheses could be killed instead of themselves -- see *Objective Knowledge* page 244) understood all the possible versions of the ability to plan ahead.

[6] Fully-deliberative systems

One of the most important features of the kind of deliberative competence that AI researchers have been investigating for several decades, which is often ignored by others, is the ability to explore branching futures. Hence all the work on search-spaces, depth-first, breadth-first and various kinds of heuristic search.

What is not always noticed is that our ability to do that depends on our ability to *discretise* the environment. To a first approximation the environment, at a low level, is full of things that vary continuously, as do sensor signals. Even when a signal jumps discontinuously that is typically just a rapid change in a signal that takes various values in a continuous range. (Ignore for now the question whether ’ultimately’ physical reality is continuous or discrete. Chemical processes obviously include discontinuous changes.)

The point is that if our environment has myriad features that can vary continuously and we can select how some of them vary, then, if we wish to explore branching futures involving many decisions, we must impose discrete boundaries between options and between time-steps: otherwise we would have to explore continuously branching continua, which requires exploding information processing capabilities that could not fit on the planet, or even the whole universe.

Types of discretisation We can distinguish ’on the fly’ discretisation from ’enduring’ discretisation, as follows:

- ’On the fly discretisation’ could be used by an architecture in which all states are represented by vectors of continuously varying components, where chunking of sets of values is done dynamically on demand.

- ’Enduring discretisation’ would be used in an architecture that partitions phenomena into discrete categories (and relationships), and the same partitions are used over extended time-periods.

An advantage of the latter is that if categories are reused then it is possible to learn ’laws’ relating them across different times and different contexts, such as the law that unsupported heavy objects move to the ground, that certain types of actions cause pain, that certain sorts of objects taste good, that a certain type of tree can be climbed, etc. Such laws can be used in planning processes, both in order to derive possible actions in future situations, and to predict consequences of the actions in those situations. Without such discretisation the only forms of control would be mechanisms that use function optimisation (maximisation or minimisation of values of functions of continuous variables). It’s not clear that that would be a good way to plan the construction of a building or machine made of many discrete parts.

NB: The claim that intelligent systems need to discretise and form categories is not new. It can be found in many statistical packages, in Kohonen nets, in Zadeh’s notion of ’fuzzy chunking’ (sometimes unfortunately described as ’computing with words’ in contrast with numbers), the work of Gardenfors on ’Conceptual spaces’, and no doubt many other partial reinventions and extensions of the idea.

What is not always realised is that besides discrete sets of *categories*, organised in multiple hierarchies, an intelligent system also needs many types of *relations* some with and some without parameters, e.g. contains, touches, moves towards, Xmm away from, taller, Xmm taller, etc. There are also causal and functional relations, e.g. supports, prevents, protects, cuts, and many more, including geometrical and topological relationships that may have causal roles in some contexts and not others, e.g. being symmetrical, having complementary shapes (allowing two things to fit together snugly), etc.

> The evolution of architectures that need to be able to discretise for different purposes is discussed in an incomplete draft paper on vision and varieties of representation here. (What the mind’s brain tells the mind’s eye.)

[7] Criteria for fully-deliberative (constructive-deliberative) competence (To be completed)

The strong requirement for a system (or part of a system) to be "deliberative", i.e. the conditions for being fully-deliberative (or c-deliberative) are, to a first approximation, as follows. When attempting to answer a question, solve a problem, find an explanation or form a plan, the system can

- gradually build up a solution in stages, where

- at each stage there is typically a set of discrete options from which one (or several) should be chosen, leading to a new stage with more choices,

- some of what is represented is relationships between objects or parts of objects or processes, not just lists of features

- in some cases relationships between relationships are needed, as first shown in a working program by T.G. Evans’ program to solve geometric analogy problems, around 1962. See his article in Minsky’s 1968 collection *Semantic Information Processing.*

- so that successive choices build representations of partial solutions of increasing complexity

- which requires the system to be able to build information structures of different complexity, including hierarchical structures of different depth, where there is no upper bound to the complexity (apart from contingent facts such as shortage of available space)

(As discussed in ’The Irrelevance of Turing Machines to AI’)

- and in some cases non-hierarchical networks or graphs (e.g. as required for the design of a machine whose parts and sub-parts have many relationships other than ’part of’).
- where the interpretation of complex information structures is based on compositional semantics (systematicity)
- the system can simultaneously hold and compare two or more solutions
- the comparison between solutions is not just some sort of numerical evaluation or total or partial ordering of options, but can be expressed in terms of how the solutions differ and and what the implications of the differences are
- where the descriptions do not have a fixed format but can vary according to the task (e.g. planning-trees, theories, explanations of an observed event, predictions of behaviour of other things, etc.).

Additional requirements can be derived from the above, including:

- having the ability to generate goals of varying complexity and varying structure
- having the ability to process goals in various ways including
  - telling whether they have been achieved,
  - telling whether one solution is better than another,
  - telling what remains to be done when they have not yet been achieved,
  - evaluating the importance of a goal (in various ways, e.g. short term benefits, long term benefits, obtaining approval from others, as an end in itself, as a means towards other things, etc.)
  - detecting conflicts between goals, or goals and values, principles, ideals
  - resolving conflicts, deciding which alternative is better than the others, or simply deciding what to do without having reasons (e.g. other than that something has to be done),
  - aborting, suspending, resuming goals
  - generating sets of possible actions relevant to the goals,
- having the ability to explain or justify deliberative decisions (including saying why something was NOT done), by selecting appropriate subsets of the information used in taking the decision
- having the ability to learn and deploy generalisations (associations) of the form ’in situation S, when there is a goal G, possibly relevant actions are A1, A2, A3’
- having the ability to learn and deploy generalisations of the form

  > ’in situation S, if action A1, is performed possible consequences are some combination of C11, C12, C13, ..., whereas if action A2 is performed possible consequences are C21, C22, etc.’

  The ability to ask and answer questions about what things (plural) are possible in a situation, and what things (plural) could happen if one of the possibilities occurs is a requirement for associative mechanisms used in a fully-deliberative system. I don’t know if there are neural mechanisms in brains that have been shown to be capable of supporting such processes in which a question gives a set of alternative answers as opposed to just the most likely or best answer. It might require a conventional associative net to be modulated by an external device that makes it produce a series of answers to a question (e.g. by perturbing some 'hidden' nodes that are in multi-stable states.)

In symbolic AI systems there are many ways of doing this, e.g. when running a system like STRIPS in breadth first or depth first mode, or some combined 'best-first' mode.

- having 'garbage collection' facilities.

  Since most of the kinds of structures described here as being created during planning, predicting, explaining, or solving problems are temporary structures capable of being modified or discarded the mechanisms within which they are constructed need either to have very large supply of memory for constantly building new structures, or else a garbage collection mechanism that allows old structures to be replaced by new ones, using the same limited memory. There are different kinds of garbage collection mechanisms, including at one extreme requiring everything to be discarded with each new structure built from scratch after every change, and at the other extreme allowing piecemeal-replacements to be made to a growing structure. The latter is a requirement for most kinds of searching, e.g. for a multi-step plan.

- having mechanisms to save temporary structures and return to them.

  If you are interrupted when in the middle of making a plan or thinking about a problem you are sometimes able to return to what you were thinking about and continue with it. This can happen even if the intervening activity also involved making plans, solving problems or performing other tasks requiring the use of temporary structures.

  This suggests that instead of just having one fixed (small) workspace in which temporary structures are created we need additional temporary but longer term workspaces in which things are constructed that we temporarily abandon and then return to. Of course, no claim is made that in humans this process or suspension and continuation of tasks is always reliable. On the contrary we often find it difficult to restore such a context in our thinking and can make mistakes in the process. That may be one of many clues as to how these mechanisms are implemented in humans, since that kind of unreliability would not afflict a typical computer based implementation of such a mechanism.

- and many more, including various kinds of learning.

  Humans do not have a fixed set of goals, preferences, values, etc. No child is born loving the music of Bach, or wanting to be a brain surgeon, or wanting to avenge his country's suffering or humiliation. There are many different ways in which the affective subsystems involved in generating, evaluating, comparing, goals etc. develop over time, both within an individual and in society. Hardly any of this has been modelled in AI, and it is likely that most of it is not yet understood.

  There are some suggestions in Minsky's 'The Emotion Machine'. Popper suggested (in his autobiographical book *Unending Quest*, 1976) that during evolutionary changes in which a species acquires modified physiology, used for new sorts of behaviours, leading to new goals, what will evolve first in the species are new goals aimed at, which in turn can lead to evolution of behaviours to support those goals, which then leads to evolution of physical or physiological mechanisms to support the behaviours.
  Notice that some of these capabilities are not implemented entirely in brains, because

humans can use many external aids e.g. diagrams, sentences, tables, lists, partial proofs, plans, represented on paper or some other external medium. I don’t know if any other animal has this ability to use an external medium as an extension of its short term deliberative memory. (Of course pheromone trails and other externally generated physical changes can perform control functions for other animals.)

It should be clear that many aspects of information processing described here depend on the ability of the robot or organism to discretise (or ’chunk’) possible states of the environment so that it can use concepts that are relevant to considering and combining alternatives. How that capability arose and how it was refined over time is an interesting question about evolution (probably with answers pointing to multiple re-inventions of approximately the same solutions to similar problems, as well as diverse solutions).

The word ’discretise’ does not imply the use of sharp boundaries with determinate and consistent classifications for all objects. Rather it is more common to have what Lotfi Zadeh has called ’fuzzy chunking’ (not to be confused with fuzzy logic).

Clearly some of these requirements will overlap with requirements for meta-management capabilities (e.g. telling whether a goal has been achieved or not). However in a deliberative system that sort of ability may simply be compiled into aspects of the internal decision making, whereas in a system with meta-management there is a *separate enduring process* ’observing’ what is going on, recording it, and drawing conclusions, including making control decisions. (This could be implemented as a ’thread’ in a computing system.)

This discussion needs to be expanded with notes on varieties of meta-management, distinguishing proto-meta-management found in simple organisms and machines with hierarchical control systems from full meta-management of types hinted at in this paper, though not described in any detail. One of the requirements for full meta-management is having meta-semantic capabilities: the ability to refer to something that refers to something else. For some purposes meta-meta-semantic capabilities may be needed.

Full meta-management makes use of a fully-deliberative competence, including the ability to consider how things might have been, or might in future be, different from the way they are currently perceived to be.

Added: 20 Feb 2009 Some related work:

http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0803
Varieties of Meta-cognition in Natural and Artificial Systems
Workshop on Metareasoning, AAAI’08 Conference
pages = "12--20",

http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0604
Long Term Requirements for Cognitive Robotics
Cognitive Robotics Workshop, AAAI’06, pp 143--150",
http://www.aaai.org/Library/Workshops/ws06-03.php

Give to search engines: "meta-semantic competence"

It should be evident that what I have been calling 'fully-deliberative' systems have features some of which were developed in research on symbolic AI systems during the decades following the early 1960s.

However those systems generally had limitations which meant that some of them were fragile and difficult to control. Many people argued, wrongly, that the only way avoid the problems was to start all over with different mechanisms. Such people generally failed to notice that all they were doing was focusing attention on a different class of problems rather than producing better ways to solve the original problems.

This point has been appreciated by researchers who design hybrid systems instead of assuming that an AI system must simply use one kind of mechanism.

Some of these points were made in 'The Computer Revolution in Philosophy' (1978) especially in Chapter 6 and to some extent in subsequent chapters.

> A distinction was made there between 'executive sub-processes', involving normal, straightforward, cases of controlled behaviour, and 'deliberative sub-processes', dealing with 'kinds of things that can happen when new planning is required, so that a question has to be answered, or an unexpected new obstacle or resource has turned up: the kinds of things which may require further intelligent deliberation and decision-making, using the agent's full resources'. In the terminology developed in the CogAff project, the former would mostly be reactive competences and the latter a mixture of deliberative and meta-management competences.
> http://www.cs.bham.ac.uk/research/projects/cogaff/#overview

[8] A difference between depth first and breadth first search

(Added 6 Sep 2008)

Two standard search procedures are depth first and breadth first search.

Depth first search is trivially implemented using a stack mechanism (last in first out) while breadth first search is trivially implemented using a queue mechanism (first in first out). There are many other search strategies including ones that use knowledge about the domain, an evaluation function, or a learning process to modify the searching process.

[Changed: 4 Jan 2014]
There are many web sites explaining some of the differences:

> A web page explaining the difference:
> http://www.programmerinterview.com/index.php/data-structures/dfs-vs-bfs/
>
> A video demonstrating depth-first search, suitable for beginners:
> http://www.youtube.com/watch?v=iaBEKo5sM7w
>
> A video demonstrating breadth-first search, suitable for beginners:
> http://www.youtube.com/watch?v=QRq6p9s8NVg

Older web sites: Alison Cawsey and Paul Brna

One of the main differences is that breadth first search is guaranteed to find the shortest route to a goal state, but at the cost of storing a lot of different routes waiting to be expanded in different directions, while depth first search always has just one route that is currently being explored, and if it hits a dead-end it backtracks to the latest branch point and starts down another route.

An important fact that is not usually pointed out is that whereas at the time a breadth first search finds a route to a goal it retains a collection of alternative routes that have so far not been successful, whereas a depth first search has only the route to the goal, though it may record unexplored branch points on that route.

This means, for example, that if you are searching for a way of building something out of meccano parts, the depth search strategy means that at any time you have one partially or fully constructed model which you can extend or undo in many ways, and if you reach a stage where you cannot extend the model and it is not yet what you need, you start undoing the model until you get to a stage where you can try a modification that you have not yet tried.

In contrast, with breadth first you would have many partially constructed models and each time you go back to one of them you copy it and modify it in order to explore a new variant.

A consequence is that with depth first you may have a solution, but no record of the alternative partial or failed solutions that you have tried, whereas with breadth first you have a number of partial solutions (e.g. partially built models) including the one you have selected as the best so far.

An important difference implied by all this is that with breadth first search you can explicitly compare one of the partial solutions, or a complete solution, with others that have been explored in order to be able to explain why the one selected is best and what was wrong with the others, whereas with depth first you never have even two coexisting partial solutions that can be compared.

In many contexts intelligent planning or searching for a solution to a complex problem requires you to be able to answer questions, e.g. about what was wrong with one of the alternatives, and this is typically not possible with a simple depth-first strategy that throws away failed alternatives.

Perhaps more importantly, when you are searching not for a unique entity or a structure with a definitely identifiable property, but for something that meets a variety of criteria which can change as your needs or ambitions change, then it may be necessary sometimes to go back to a partial solution that was previously abandoned and continue developing it (like going back in the 20th century to the particle theory of light which had been abandoned after Young had demonstrated interference effects supporting a wave theory).

So a planning or problem-solving or searching mechanism that can only maintain a single version of a possible solution, like simple depth-first search, is inferior to one that, like breadth-first search, can maintain several different alternatives which can be explicitly compared, experimented with and extended. Systems whose explorations involve only one workspace in which at most a single solution or partial solution at exist will therefore be inferior to systems that support parallel workspaces.

The latter is one of the requirements for a fully deliberative system. It may be that a compromise mechanism is one that can store abstract specifications in a descriptive formalism, for the contents of a single workspace and can switch between alternative solutions by switching between descriptive specifications that can be used rapidly to reconstruct the alternatives to enable them to be further elaborated, generating more specifications.

It would be interesting to find out which solutions are adopted by brains, since it is clear that humans can and do reason about, talk about, and learn from things learnt in alternative branches of a search space. Sometimes they use external stores, like notebooks or different laboratory experiments running in parallel. But in some cases they clearly do not need external memories to do parallel searches, just as a chess master can play parallel games of chess blindfold.

[9] The need for temporal competence

As pointed out in this book:

Russell A. Barkley, *ADHD and the nature of self-control*, The Guildford Press, (1997).

humans have an apparently unique(?) ability to think and reason in multiple ways about times other than the present, an ability that is sometimes impaired in genetic or other brain disorders. A full survey of requirements for temporal competence would be very lengthy as they are many and subtle but for now we can list a few.

- The need to think of time both as continuous and as divisible into discrete events and intervals at different levels of abstraction
- The ability during planning to represent events and processes as ordered, without necessarily having any metrical relations: e.g. do A, then do B, then do C. Telling stories requires a similar ability.
- The foregoing generalises to allow overlaps and partial orders: e.g. A should be done before B, and C should be done before D.
- The ability to relate times either to natural subdivisions (today, tomorrow, the next day), or to conventional subdivisions (next week, next month, October 2002, etc)
- Understanding that time intervals are indefinitely divisible.
- The ability, implied in previous sections, to consider *possible* temporally related actions which may or may not ever exist.
- Understanding notions like 'speed', 'acceleration', as involving comparisons with changes in other things and changes in times, and being able to reason about consequences of speeding up/slowing down different sorts of processes (continuous and discrete).

- Being able to synchronise some concurrent processes, e.g. pointing and counting, or reaching forward to grasp something and altering the grasp concurrently with moving towards it.

- Being able not only to do various things involving temporal concepts and relationships, but also to think about what is involved in doing them. E.g. being able to reason about things being in and out of phase, or about degrees of urgency of various task, or being able to think about cyclic opportunities -- if you don’t sow the seeds today (or this summer) the opportunity may come again tomorrow (or next year).

[10] The need for modal competence

There are several kinds of modal logic. The oldest kind (sometimes called ’alethic modal logic’) is concerned with notions like ’possible’, ’impossible’, ’necessary’, ’contingent’. There are subtly different versions of these notions that have been formalised by logicians. For now all we need to note is that an animal or robot considering what to do, or how to do things, or when to do things, may find it useful to distinguish among things that it can represent as possibilities those that are really impossible, those that it can do, those that are necessary conditions for others, and so on. For instance an animal that is unable to distinguish a chasm that it can jump over from one that it cannot, may not live long in certain kinds of terrain.

There is a lot more to say about modal competence, including the ability not only to think about what *is* possible or impossible, or about what *would be* possible or impossible in situations that don’t exist but could exist if some action or sequence of actions were performed.

A paper that may be of interest in this context is Actual Possibilities (1996)

Placeholder (added 6 Jan 2010): Material needs to be added regarding consideration of sets of possibilities, with or without use of probabilities, in contexts where the agent does not have enough information to be sure which of a set of possibilities will be the result of an action, or which of a set of possibilities is the best way to interpret sensory/perception information. Earlier work in AI mostly ignored these issues (though not in all contexts -- e.g. games against an opponent whose moves could not be predicted were investigated). More recent work using probabilistic mechanisms often ignores the non-probabilistic structure of a problem (e.g. effects of rigidity of objects being manipulated, or walls, etc). Work on so-called "hybrid" planning and reasoning systems attempts to address this. (Compare the comments on discretisation and invariants, above.)

[11] The need for Affective/Evaluative mechanisms/competences

There is much confusion and also much wishful thinking about the relationship between intelligence and emotions, some of it based on wide-spread acceptance of a fallacious argument in a book by Damasio, as discussed in this presentation on emotions.

What is true, as Hume pointed out long ago, is that no amount of knowledge and competence of itself determines that one should do anything at all: some motivation is required to select between all the possible things to do, including doing nothing at all (’Reason is, and ought to be, the slave of the passions’ in *A Treatise of Human Nature,* (2.3.3.4)).

Of course, most physical objects behave without having reasons, like a leaves blown about in the wind, an avalanche triggered by a change in temperature, water evaporating in strong sunlight. Such behaviours are not based on motivation, perception, reasoning, planning, or selection between options. Organisms and robots are different. They have access to a source of energy (usually chemical energy in plants) that can be deployed in different ways, to achieve different end results. So there is a need for choice.

Therefore an animal or intelligent machine needs to have something like goals, motives, preferences, ideals, values, hopes, fears, desires, likes, dislikes. Moreover, since having a fixed set of goals or motives would be inappropriate for something embedded in a rich and changing environment, something else is needed, which I have elsewhere referred to as ’motive generators’.

For example, the perception of an approaching dangerous animal can generate motives like *finding somewhere to hide* or *running away*, possibly both at the same time.

As that example shows, if two or more incompatible motives arise, then mechanisms are required for *evaluating* them and choosing between them. Sometimes that can be automatic, where the advantages of one are great and obvious (as a result of prior learning or the operation of some innate evaluation mechanism), whereas in other cases it may require arbitrarily complex investigations of prerequisites, difficulty, costs, consequences, benefits, etc., in order to decide which option is better.

It is often assumed by AI theorists and psychologists, and even some philosophers, that all such conflicts are resolved by reference to some fixed evaluation function or utility measure that can be assigned to all possible outcomes of actions.

However, it seems that in the case of humans there is no such thing. Rather people go on learning or developing new ways of evaluating and comparing alternatives of many different kinds throughout their lives, with many different influences on the process. A full theory of how human-like deliberative systems work, therefore, would have to include an account of those processes. For more on this see slides 96-100 of the IJCAI’01 tutorial on philosophical foundations of AI, and Luc Beaudoin’s PhD thesis (1994).

[12] Further requirements related to using results of deliberation

A more detailed analysis would have to consider what an intelligent system does with the *results* of deliberation, be it a proof, an explanation, an interpretation of something, or a plan.

Early robotic research (in the 1960s) demonstrated clearly that the process of plan-creation could not be relied on to produce a plan that can simply be executed in order to achieve the plan’s goal or goals. Reasons for this include such facts as

- in a dynamic environment conditions assumed during plan construction may no longer hold and
- actions may not produce their expected effect because of slippage, faulty mechanism, imprecision of control, etc.
- perceptual data may be incorrect or insufficiently precise to form the basis of ballistic actions: hence feedback loops (e.g. visual servoing) may be required during plan execution and frequent testing and re-planning may be required (as in the hybrid reactive/deliberative sheepdog demonstrated here.)

> **NOTE** Some youngsters criticising symbolic AI a couple of decades later thought they (or their supervisors) had discovered these limitations of ballistic control. What they did not appreciate was that the technology available in the early days of AI made visual servoing and most other forms of online control of actions totally impractical. That does not mean that the need was not understood. However, it is possible that there were some unimaginative individuals who believed that a ’sense-plan-decide-act’ cycle could work for intelligent robots.

Besides short-term plan and action modifications that could be required during plan execution there are sometimes deeper, more long term, changes that result from monitoring what happens during plan execution. In particular, flaws in the plan (e.g. unnecessary steps, unexpected interactions between plan steps) may be discovered during execution, leading to the discovery of ways of improving the planning procedures so that future plans do not have similar flaws.

As far as I know the first working demonstration of this was the PhD thesis of

> G.J.Sussman, describing the ’HACKER’ program (Sussman, G. (1975), *A computational model of skill acquisition,* American Elsevier.)

More recent techniques include explanation-based and case-based learning. In the long term such developments will require deep advances in meta-management mechanisms, including the ability of such systems to extend their own ontologies. Mechanisms for introducing non-trivial extensions, i.e. introducing new symbols that are not definable in terms of the pre-existing concepts, such as may be needed for explaining some new observation or failed actions, raise many problems including the problem of controlling the search for good extensions. That is a problem that has been faced in the history of science and mathematics, and more recently in the history of programming languages.

It may also be a problem faced by infants and toddlers learning about the world.

I suspect that one of the strong drivers for ontology extension is the process of debugging, of which simple examples are in Sussman’s work mentioned above.

There is much more to be said about requirements for mechanisms and forms of representation involved in processes related to using the results of deliberations of various kinds.

[13] Some implications

There are many obvious and unobvious consequences of these ideas.

One of the consequences is that tasks that can be achieved by a fully deliberative system will not map onto most of the kinds of mechanisms that have been explored by the recent wave of researchers who reject symbolic mechanisms, whether they study neural nets, or more general dynamical systems. In part this is related to the fact that it is not at all clear what sorts of mechanisms in animal brains are able to satisfy all the above conditions.

But it is clear that at least human brains support all the kinds of functionality described here, though to different extents in different humans, or in different stages of development of the same individual.

Whether anything known about brain mechanisms explains *how* brains are capable of supporting such functionality is another question. I suspect new brain mechanisms will have to be discovered to account for these human capabilities.

One of the clues in the search for explanatory mechanisms is probably going to turn out to be the fact that most computer-based implementations of these ideas have two striking features that differ from human competence

- The computer-based systems are much more narrowly focused than human capabilities: the computer systems do not ’scale out’ as humans do, allowing capabilities and knowledge of many different kinds to be deployed creatively in solving new problems.
- The computer-based symbolic systems typically outperform all or most humans in their narrow domains of competence, whether the field is numerical calculation, performing operations of sorting and permuting, solving constraint problems, finding plans, or playing games like chess or checkers/draughts and many other discrete board games, (but not GO). In short, computer-based symbolic systems often ’scale up’ much better than humans even if they don’t scale out.

It is possible that these two features: the ability to *scale out* and the ability to *scale up* are incompatible. If we understand why, we may get new kinds of clues as to what the brain mechanisms are that allow scaling out without scaling up.

One of the ideas that is often re-invented in this context concerns use of what some people have come to refer to as a ’global workspace’ (e.g. Bernie Baars, Stan Franklin, and Murray Shanahan) which other people have (since the mid 1970s) been referring to as the blackboard model. What they have in common is the use of some central mechanism where work is done sequentially combining expertise related to a lot of subsystems with narrow domain expertise which can to some extent operate in parallel, including monitoring and reacting to what is going on in the blackboard or global workspace. (Compare Chapters 6 and 10 of The Computer Revolution in Philosophy (1978))

Another important idea is that the powerful computer-based problem-solving systems typically operate with a uniform mode of representation, whereas humans, including mathematicians, seem to use a variety of different forms of representation for different purposes, and for different sub-tasks within a single task, and are are capable of inventing new ones to add to what they can do easily, as has been pointed out by many people.

Perhaps by combining the requirements that these different approaches explicitly or implicitly impose on mechanisms we may be able to constrain the search for biological mechanisms able to meet the requirements. My hunch is that nobody has so far brought all the requirements together, though there are many partial attempts (including recent work by Marvin Minsky and John McCarthy, Arnold Trehub’s book ’The Cognitive Brain’, MIT press, 1991, some probing analyses of work in infant or child development (e.g. Philippe Rochat), work on brain damage and varieties of genetic cognitive deficiency (e.g. Russell Barkley *ADHD and the nature of self-control*, 1997) and work on requirements in Birmingham. I know that there are many more that I don’t know of!)

NB:

> The fact that I have not discussed mechanisms and requirements that deal with continuous variation, continuous control, attractors, etc. does not imply that I regard them as unimportant or irrelevant. All animals need such mechanisms, and that includes humans. The same is true of future human-like robots acting fluently in a complex and changing 3-D environment will also need them, e.g. for posture control, visual servoing of many kinds, and a host of predictive or anticipatory competences that can superficially be regarded as overlapping in their functions with the functions of a deliberative system.
>
> Neither am I saying that the deliberative, or more generally discrete symbolic, mechanisms are more important than the continuously varying sub-mechanisms that may be better described by systems of differential equations than by computer programs. As J.L.Austin once said in response to an objection at a conference 'Truth is more important than importance'.

---

[14] Related papers and presentations

See this paper on 'Sensorimotor vs objective contingencies' and papers referred to in it.

I have not checked recently, but from my memories of Minsky's 1960 paper, 'Steps toward artificial intelligence' (published in 1964 in the Computers and Thought volume by Feigenbaum and Feldman), it makes several of the points presented here (and many more that should not be forgotten by philosophers, scientists and engineers interested in the nature of minds, natural or artificial).

I have not tried to define 'representation' here. The word is now used very widely and very loosely to refer to many different sorts of things that can occur in information processing systems, or in their environments. The important thing is not what a representation is but how it used. I have written extensively about the variety of forms of representation elsewhere, e.g. in 1971 and 1994

The first definition of the word 'representation' I heard from a computer scientist is still one of the best. It was expressed by Robin Stanton when he was a PhD student at the University of Sussex around 1970, something like this:

> *A representation is an addressable structure that facilitates computation.*

Whether that summarises all uses of 'representation' by scientists and engineers adequately or not, it fits most of what I have said here. A chapter written for a book on *Information and Computation* was published in 2011:
http://www.cs.bham.ac.uk/research/projects/cogaff/09.html#905 What's information, for an organism or intelligent machine? How can a machine or organism mean? This argues that words like "information" and "representation" cannot be explicitly defined. However they are implicitly partially defined by theories referring to information and representation and their role in the world.

> Jane Austen had an implicit theory of information, about 200 years ago. Evidence for that is presented here, using extracts from 'Pride and Prejudice'.
> http://www.cs.bham.ac.uk/research/projects/cogaff/misc/austen-info.html
>
> ---
>
> Other cases are mentioned in connection with information processing in invertebrates by Barbara Webb in 'View From the Boundary', *Biol. Bull.* 200: 184-189. (April 2001) and in Transformation, encoding and representation *Current Biology,* Volume 16, Issue 6, Pages R184-R185, 2006 (alas available only to subscribers).

---

[15] Updates up to 2008 (moved to end 4 Jan 2014)

NOTE: Early versions of this document ignored cases where the results of actions can be uncertain (e.g. instead of turning the key definitely making it possible to open the door, it may merely reduce one obstacle, if the paint of the door has stuck, or there is furniture behind the door.)

Updated: 6 Sep 2008 Added ability to explain decisions, and important difference between depth-first and breadth-first search in providing knowledge to searcher.

Updated: 12 Apr 2008 Further relevant material was in an invited talk at: Workshop on Meta-Reasoning at AAAI'08, Washington, July 2008. Now available, with a revised version in the book based on the workshop, here:

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

Updated: 30 Mar 2007 Added *example purposes* for which alternative possibilities may be compared, in the summary of key features of deliberation.

Updated: 19 Jan 2007 Insofar as the need to take account of deliberative competences of the sorts described here in humans, other animals, and intelligent robots is denied by some authors, this document is a contribution to the list of Controversies in Cognitive Systems Research on the euCognition wiki also accessible here. Comments and criticisms welcome.

Updated: 31 Dec 2006 Minor clarification. Reformatted updates section.

Updated: 21 Nov 2006 Added further details on the need for discretisation, and the distinction between 'on the fly discretisation' and 'enduring discretisation'. Stressed the importance of relations as well as categories or types of thing.

Updated: 13 Aug 2006 In response to a comment from Alasdair Turner, replaced the intrasomatic/extrasomatic distinction with somatic/exosomatic. This is partly because 'soma' is Greek, whereas 'intra' and 'extra' come from Latin, and partly because I think some other people have been using the word 'somatic' with roughly the meaning I gave to 'intrasomatic'.)

---

[16] Admin

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

Note on British/American spelling:
I tend to use British spelling, e.g. discretise, discretisation, and behaviour, not behavior. However sometimes I quote things written by other authors using American spelling, e.g. discretize, discretization, behavior, behaviors. This paragraph using both spellings should help search engines that don't know about the equivalences.