FP6-004250

CoSy

Cognitive Systems for Cognitive Assistants

Integrated Project

Information Society Technologies

DR.2.1

Requirements study for representations

Due date of deliverable: Month 12
Actual submission date: October 2, 2005

Start date of project: September 1st, 2004 **Duration:** 48 months

Organisation name of lead contractor for this deliverable:
BHAM

Revision: Final, October 2, 2005
Dissemination Level: PU

Contents

Executive Summary

The full title of Work Package 2 is 'Representation of space, objects and context'. It is defined in section 8.2, on pages 54–56, of the project workplan.[1] There are three deliverables, the first one due at month 12, and the others at month 18.

DR.2.1 Requirement study for representations (This report)

DR.2.2 Requirement study: Analysis of roles of alternative forms of representation.

DR.2.3 Working definitions for information & information processing as a basis for a cognitively plausible theory of representation in embodied cognitive artificial agents

The introduction to Workpackage 2 states:

> Many AI researchers have been attracted by the claim that meanings have to be "grounded" which is very close to the old concept-empiricist theory that all concepts are derived by abstraction from experiences of instances. Such theories are, however, problematic in that they require concepts for having experiences, and that there are concepts that are either only indirectly or not at all related to experience (e.g., mathematical concepts).
>
> Our approach is to avoid dogmatic pronouncements about absolute requirements and instead investigate different ways in which an agent can acquire a grasp of meanings, e.g., through innate structures, through abstraction from instances, from combinations of structure and sensory motor connections or via boot-strapping processes that use exploration of the environment to instantiate high order schematic structures. (As noted in 6.1.1 we are not concerned with the debate over whether explicit representations of meaning are useful, only with investigating which kinds are useful for which purposes.) We shall use whatever mechanisms and forms of representation are appropriate to the tasks, including, if necessary, different forms of representation for fast fluent actions produced by highly trained reactive sub-systems from those needed for deliberative and reflective mechanisms. A fundamental consequence of this view is that meaning is grounded neither strictly in experience, nor strictly in conceptual knowledge, but in both, thus capturing an essential aspect of embodied cognition.

This work is closely related to work on architectures, reported in DR.1.1. In this early deliverable on the first phase of work-package WP.2, we report on some of the hard unsolved problems we have identified on the basis of detailed analysis of some of the processes that will have to occur when the PlayMate and Explorer robots perform their tasks. The analysis used our scenario-driven research methodology. We introduce some preliminary characterisations of the key problems and some preliminary ideas for dealing with them, inspired in part by studies of cognition in humans and other animals. We confirm the conjecture in the CoSy proposal that various kinds of representations are required for different sorts of sub-mechanisms (including for instance representations concerned with planning complex sequences of actions and representations used in producing and controlling fast and fluent movements). The different representations are in part related to different ontologies, since different mechanisms acquire, manipulate and use information about different subject-matter. A substantial part of this report is therefore concerned with first draft, incomplete, ontologies that we expect our robots will need, some parts of which the robots will have to develop for themselves, especially ontologies concerned with objects and processes that have quite complex structures involving multi-strand relationships. We consider alternatives to common forms of representation and suggest that it is worth exploring forms of simulation at different levels of abstraction, that can capture some of what is often referred to as spatial or visual reasoning. We consider in particular requirements for a pre-linguistic robot that is capable of perceiving, acting in and to some extent reasoning about the world before being able to talk about it, and raise questions about how that might relate to learning

[1] https://www.cognitivesystems.org/intranet/offDoc/workplan-v5.pdf

that adds linguistic competence. We note that in animals there is wide variation between species that start with most of the ontology and representational competence they will ever need and those that somehow learn or develop what they need and suggest that further study of those cases may yield clues regarding options for robots of different kinds. Most of this work has not yet been published. This is work-in-progress and much of it remains to be expanded, clarified and polished.

Role of (topic of deliverable) in CoSy

Information processing of several kinds is central to the design of any cognitive system. Information needs to be represented somehow and there are various conflicting views on how that should be done. Addressing the problem of representation in a principled manner is therefore central to the success of CoSy.

Relation to the Demonstrators

During the process of designing and implementing the demonstrators, many decisions will have to be taken about how to represent various kinds of information. For simple demonstrations anything that works will suffice. As more and more ambitious demonstrators are considered, combining wider ranges of competence, in increasingly demanding task situations requiring ever more flexibility and autonomy in the robot, the constraints on choices of representations will become more and more demanding. This work package is therefore required in order to provide a long term view of the requirements, so that decisions taken at an early stage in order to produce the first demonstrations do not constrain the long term success of the project.

It may not be possible within the scope of the four-year CoSy project to demonstrate all the kinds of competence that we are analysing, including learning and development of new representations. But the scientific aims of CoSy involve looking beyond what can be implemented within the scope of the project, towards the longer term. So this work package is viewed as relevant to future demonstrators also, such as the ‘motivating’ hypothetical scenario presented in section 4 of the project Workplan and used in Deliverable DR.1.1 to motivate the requirements analysis for architectures (namely the Robot Fido helping the Hartmut family).

1 Requirements study for representations

1.1 Background: representational issues in natural and artificial systems

Although in the last decade and a half it has become fashionable in some circles (following [Brooks, 1991]) to claim that representations are not needed in intelligent systems we regard this sort of claim as merely part of a strategy to focus attention on a particular narrow class of types of representation. For insofar as animals and machines process information that information somehow needs to be available to the mechanisms that perform the processing. Whatever encodes or embodies the information in a usable way can be regarded as a representation in the general sense of being "something that presents information". In this sense, talk of representations and information-processing is now commonplace among many kinds of scientists including not only Computer Scientists and AI researchers, but also psychologists, neuroscientists, biologists and physicists.

In a full survey we should need to make many distinctions relevant to the role of representations in intelligent systems.

- Different sources of information – from genes, from external sensors, from internal sensors, from other intelligent systems, derived by inference from other available information.
- Different uses of the information, e.g. transient uses in tight feedback control loops, long term storage for multiple uses, recording specific facts, recording generalisations, expressing questions (gaps in information), expressing goals, expressing instructions, or collections of instructions in plans, including conditional instructions.
- Different forms in which the information can be stored, e.g. implicitly in current activation states of sensors or motors, explicitly in propositional records or other data-structures, in patterns of activation in neural nets, in weights of connections in neural nets, in chemical structures or chemical concentrations, in states of the physical body, in external objects, in external symbolic forms, in genetic mechanisms, in social and legal conventions. (Compare Popper's 'Third world')
- We could also distinguish simple or 'atomic' information structures from compound forms of representation where the information represented by the whole depends systematically on both the semantics of the components and on the mode of composition (compositional semantics). There are different ways in which this can happen, some of them studied in connection with logical and algebraic formalisms, others found in diagrammatic and pictorial or other spatial representations (often mistakenly thought to be isomorphic with what they represent, whereas the relationships are usually more subtle [Sloman, 1971])
- The use of mainly numerical forms of representation, as opposed to qualitative and structural forms.
- Associated with the last point is a distinction between representations of probabilities or probability distributions which express uncertainty or indeterminacy, as opposed to determinate representations (which may be abstract or explicitly disjunctive instead of being numerically indeterminate).
- For agents in a changing world there is also an important distinction between representing static facts and representing events and processes, both when they occur and also at other times.
- Besides representing objects which do not themselves represent anything, intelligent systems sometimes need *meta-semantic competence*, namely the ability to represent the contents of structures that are themselves representations of other things, for instance in reasoning about sentences or pictures, or the desires and beliefs of other agents. In addition to that second-order semantic competence agents may need third-order or higher order competence, namely the ability to refer to things that refer to things that refer, for

instance in reasoning about what another individual thinks about one's plans or about a third individual's beliefs.

- Besides representations of actual or believed *facts* i.e. representations which the agent takes as containing information that can be relied on in actions, it is also important to represent *hypothetical possibilities*, for instance, possible explanations of observed facts, possible histories leading up to some situation, possible contents of obscured or distant regions of space, possible future events, and possible contents of the minds of other agents. This ability to represent and reason about what is not taken to be the case can be used in a variety of different ways, for instance in searching among possible explanations for the one to be believed, or searching among possible predictions of future actions, or among possible plans for the one to be adopted.
- A particularly important special case discussed in Section 3 is the ability to identify and represent gaps in one's own information, so that the gap can be expressed as a question which then generates processes intended to fill the gap. These questions may be used entirely within an agent, or may be used in communication with other individuals. Only the latter depend on the use of what we normally call natural language. Evolutionarily older forms of representation may be used by pre-linguistic children and animals that do not talk but may nevertheless be able to ask themselves questions in some sense. Similar things may be needed by intelligent robots even if they do not engage in communication.
- Searching was mentioned above, namely exploring a space of possibilities in order to answer a question, formulate a plan, explain something, prove a conjecture. The process of searching may use representations of items in the space being searched (e.g. actions and their consequences) but also requires a representation of the searching process so that it can be controlled. Different sorts of search may use different kinds of representations, depending on the requirements of the search process and the complexity of the subject matter. Although this is one of the oldest areas of AI research we still have much to learn about ways of controlling and representing search processes, though nowadays there is an unfortunate tendency for research groups favouring one or other type of process to behave as competing factions.
- One of the current topics of research is the extent to which an individual's body, including the available sensors and motors, either reduces or simplifies the need for internal processing, or constrain the kinds of information that can be represented or used by that agent. This is often discussed under the heading of 'embodied cognition', and is sometimes linked to the aforementioned claim that representations are unnecessary for intelligence. However, it is clear that people deprived from birth of arms and legs (e.g. following the thalidomide tragedy of the 1960s) share a great deal of cognitive competence with humans with normal bodies, and blind people share a great deal with sighted people. A full theory needs to explain how this is possible. Some of the confusions found in discussions of embodiment are discussed in [Ziemke, 2002]. Some aspects of the role of embodiment have been known for many years, e.g. stigmergy in many insects, and wall-following as a way of getting out of a maze. But it is likely that there are more to be discovered.

There is a huge amount of literature on all these topics. This report merely addresses a subset directly relevant to both the short term CoSy scenarios (PlayMate and Explorer) and possible longer term scenarios in which the ideas may be applied in domestic robots, such as the one sketched in the CoSy proposal (WorkPlan, section 4):

> The Hartmut family decides to acquire a CoSy system to assist them in their everyday life. A system is acquired from the local Robs-R-Us chain of stores....

The robot (Fido) learns the layout of the house, finds out about and learns to use the machinery in the house (refrigerator, dishwasher, etc.), copes with changing domestic equipment (e.g. newly installed sliding doors), learns to play a child's game by deliberately asking the child to teach it instead of downloading the formal specification, and adjusts its level of play to suit the child. This gives the flavour of the inherently open-ended variety of competences under consideration in this project. Of course the detailed requirements would be different for a robot on a factory floor, a robot fire-fighter, a robot helping a cattle farmer, a robot assistant for interplanetary explorers, or an intelligent robot with a microscopic body helping brain surgeons.

1.2 Constraining the problem space to human-like robots

Our conjecture is that despite all the differences in different kinds of knowledge and skill, along with different kinds of physical body, that might be explored in robotic projects, there is a collection of competences produced by biological evolution found in typical human children in their first few years of life, which provides a generic platform for developing and learning in vastly different directions. This follows from the fact that a typical child in that age range can be transported to any human culture in any geographical region on earth and will develop and learn what it needs to cope with that environment, just as successive generations of children in our culture have acquired knowledge and skills that none of their ancestors had, including the ability to use mouse-driven computer interfaces and other new machinery, and have learnt scientific concepts and theories that were previously unknown. So there is something powerful and general in such child-like competence, whereas the competence employed by an adult human sheep shearer, brain surgeon, plumber, psychiatrist, historian, concert pianist or politician will typically depend on a vast amount of specialised learning of factual information and task-specific skills arising from the individual human's history, which lacks the kind of general applicability of the child's competence produced by the history of the species, but builds on it.

1.3 Some limits of human competence

Note that we are not claiming that a human child's competence is *completely* general. There are undoubtedly limits to what humans can learn and do, and cognitive engineers of the future may wish to develop robots that exceed some of those limits (referred to as 'the human window' by Donald Michie). Indeed humans do not have the power of Turing machines insofar as short term memory limits and tendencies to error of various kinds will get in the way of performing some of the tasks a Turing machine can perform. (That is one of the reasons for giving certain tasks to computers rather than humans.) But this project is guided by the less ambitious goal of understanding and partly replicating generic human capabilities. This constrains our consideration of forms of representation and architectures. For example, we do not consider a robots with a large number of visual sensors distributed around a building, or a town, or with a disconnected body whose multiple arms and legs might be performing different tasks simultaneously in different rooms. We leave that to successors of CoSy, and suspect that their progress will depend on success in understanding competences within the human window.

That raises the question of what the biologically general competences in young children are. That is by no means obvious and although a traditional way to find out is to do experiments to find out what children typically can and cannot do at various ages, we feel that such research, valuable as it is, has to be supplemented by design-based analysis: attempting as engineers to investigate the information-processing requirements for the various task, for instance picking up a teacup, or realising that you don't know where the sugar is when you need sugar. Investigating these requirements can, as far as we know, only be done by careful analysis of the *minutiae* of very detailed scenarios, such as the scenarios we have begun to analyse in this and other deliverables, partly aided by using the scenario-template we have developed here, http://www.cs.bham.ac.uk/research/projects/cosy/scenarios/, and which we expect to continue developing.

As will become clear, we believe that one of the most important features of this generic human competence which is already very sophisticated but still undergoing development in young children, is an understanding of spatial structure, motion and causation in interacting objects of varying kinds and degrees of complexity. This sort of competence is also shared, to some extent with other animals, especially animals that pick fruit, build nests and shelters, or hunt and eat other animals.

Our attempt to characterise that understanding of spatial structure and motion and its relation to perception of affordances, planning capabilities, reasoning about change in the environment, plan execution, and detail control of action drives many of the details of our research.

1.4 Criteria of adequacy of representations

The influential logicist manifesto [McCarthy and Hayes, 1969] defined three types of criteria for adequacy of representations, echoing related criteria of adequacy for linguistic theories previously formulated by Chomsky in [Chomsky, 1965]. In section 2.3 they write:

> "From the artificial intelligence point of view we can then define three kinds of adequacy for representations of the world.
>
> A representation is called metaphysically adequate if the world could have that form without contradicting the facts of the aspect of reality that interests us.
>
> . . .
>
> A representation is called epistemologically adequate for a person or machine if it can be used practically to express the facts that one actually has about the aspect of the world.
>
> . . .
>
> A representation is called heuristically adequate if the reasoning processes actually gone through in solving a problem are expressible in the language."

Those three criteria, with various modifications, have been very influential in the last few decades of AI research, but they are not the only important criteria of adequacy. For instance they say nothing about adequacy in relation to the ability to perceive and act in a physical environment, or the more stringent requirement to act fast enough to achieve specific goals (such as catching prey). They also say nothing about one of Chomsky's criteria, which he labelled *explanatory adequacy*, which required the ability to account for how a child learns language. We can generalise this to a requirement for theories of representation and architecture in an intelligent system to account for how a child-like robot can perceive and act in a physical environment and learn many things, including language. In later sections we indicate how detailed analysis of very detailed scenarios can draw attention to otherwise unnoticed requirements from which we can derive criteria of adequacy. For example the existence of changing multi-strand relationships, discussed in Section 2.13, demonstrates the need to represent multi-strand processes at different levels of abstraction, leading up to a generalisation of the old idea that perception produces representations at different levels of abstraction, to the proposal that perception produces multiple simulations running concurrently at different levels of abstraction.

We do not yet have a complete theory meeting these requirements. So this report is only partial. In a later section 5.3.3 we illustrate more detailed criteria related to specific capabilities of an intelligent robot.

1.5 Varieties of representations in CoSy

Most of our work in this area has been concerned with the very demanding requirements for the PlayMate scenario, including requirements for the ontology to be used by the robot in thinking about its environment, its actions, other agents, and its own mental processes (see Section 2). Similar work will be done later in relation to the Explorer. That will require combining

- representational capabilities related to relatively small and structurally rich 3-D spaces in which many spatial transformations can occur
- representational abilities related to relatively large and to a first approximation 2-D spatial layouts in which moving objects are treated as points or small circles.

So far, we have mostly analysed requirements relating to

- the robot's ontology for spatial objects, their parts, relationships and processes involving them, including actions as well as consequences of actions
- representation in visual mechanisms, including
  - representation of spatial structure (perceived, remembered, planned, or hypothesised),
  - representation of affordances related to detailed structure,
  - representation of perceived processes, including the robot's own actions and the actions of others
  - dealing with missing information about partially occluded objects or other unperceived parts of the environment,
  - dealing with noise and other kinds of uncertainty in images, e.g. due to low resolution,
- planning 3-D actions involving grasping, translating, rotating, inserting, placing, releasing etc.
- representation of general laws about objects, events, processes, such as causal relations, including
  - purely empirical generalisations based on observed frequencies (e.g. lights can be turned on and off by using switches on walls)
  - generalisations based on understanding of causal relations (e.g. pushing one end of a rigid object away from you makes the other end move further away too – and many cases where understanding causation arises from perception of mechanisms, as in old clocks).
- the meta-semantic ontology for perceiving, reasoning and thinking about agents (including oneself), their percepts, beliefs, desires, plans, intentions, actions, and other mental states and processes
- learning all of the above, on the basis of some initial (innate) set of competences.

We have investigated those requirements and found that the current state of the art (based both on our own assessment and also informal consultation of several internationally known vision researchers) is far behind what is needed for achieving the goals of CoSy as originally conceived.

1.6 Requirements for pre-linguistic spatial competence

In particular, achieving the research goals of CoSy will require pushing frontiers in vision, in representing spatial structure and affordances, and in manipulation of 3-D structures. In young children, and presumably in some future robots there are many competences that are manifested well in advance of the ability to understand or produce spoken language, and that raises deep questions about the forms of representations used in that early stage and how they later relate to the representations underlying linguistic competence. E.g. are there generally applicable syntactic forms in the mental processes of prelinguistic children (and robots) that somehow provide a basis for learning the more specialised syntactic forms of particular human languages?[2]

There are also requirements for mechanical construction of arms and hands and for control mechanisms, which are not met by arms currently available off the shelf and affordable, for instance number of degrees of freedom and availability of force feedback, in addition to the requirement of being small and safe for activation close to a person sitting at the same table. These constraints will have to be addressed by restricting the implementations to simplified versions of the scenarios, which will also reduce the scope of the representational issues to be addressed in the implementations, though not necessarily in our theoretical work.

As our research has already identified requirements and tasks that go beyond the resources of the CoSy project we are attempting to formulate those requirements in a manner that can inspire follow-on or complementary projects in addition to guiding research within CoSy. Results of such research done in other projects will be integrated with the CoSy results as they become available.

We turn now to more detailed requirements for future visual systems.

1.7 Beyond current preoccupations in machine vision

Most of the research work being done nowadays on vision in AI is concerned with:

- Recognition/classification/tracking of objects (including face recognition), including prediction of motion
- Optical character recognition (special case of the previous point)
- Self-localisation, route learning and route following.
- Pushing things, avoiding things, blocking things (as in robot football).
- Various special-purpose applications of the above, e.g. floor cleaning and lawn mowing.

Though extremely challenging and also useful for many practical applications such research has important gaps in relation to requirements for an embodied intelligent robot of the sort we are considering. In particular it leaves much work to be done on issues such as the following:

- Perception of structure (at different levels), e.g.
  - perception of 3-D parts and their relationships, especially in novel complex objects and configurations of objects, where components of a configuration may have complex multi-strand relationships (e.g. various portions of one object have various relations to parts of the other objects, such as the relations between corners, edges and faces of one cube and corners, edges and faces of another),[3]

[2]Compare the recent paper by Michael Arbib, ‘From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics’ in *Behavioral and Brain Sciences*, April 2005, http://www.isrl.uiuc.edu/~amag/langev/paper/arbib04BBSmonkeylikeAction.html

[3]For more on this see the discussion of multi-strand relations in the section on ontology Section 2.13.

  - perception of motion in which relationships between parts of one object and parts of other objects change, including things like sliding along, fitting together, pushing, twisting, bending, straightening, inserting, removing, and rearranging – where some processes involve multi-strand relationships with different strands changing in parallel, e.g. one corner Ca of cube A moves towards a corner Cb of cube B, while the faces meeting at Ca rotate relative to the faces meeting at Cb.

- Perception of positive and negative affordances and causal relations, e.g.
  - Possibilities for action, for achieving specific effects
  - Obstructions to action, and limitations of actions

  especially as regards parts of complex objects, which can be grasped, pulled, pushed, twisted, rotated, squeezed, stroked, prodded, thrown, caught, chewed, sucked, put on (as clothing or covering), removed, assembled, disassembled, and many more...; as well as many variations of each of the above.

The particular forms of representation used in many current vision programs are often merely 2-D segmentations of an image, with a vector of numerical and other parameters computed for each region in the segmentation along with perhaps a label which results from recognising the object depicted by that region or some larger set of regions containing it. Such representations provide very little of the information available to human perceptual systems and required for many manipulative tasks.

1.8 The need for analysis

In order to identify requirements for forms of representation and the representational contents of visual and other mechanisms involved in spatial cognition it is necessary to examine many examples of spatial cognition in action, break them down into separate fragments and *analyse the fragments in great detail*. Some researchers would prefer to implement some kind of powerful learning mechanism and simply allow the robot to learn what it needs, perhaps with a great deal of trial and error. That may suffice for solving some practical problems (though there could be risks if we don't know the scope and limits of applicability of what has been learnt). However for the purposes of *science* merely having a robot learn how to solve a task no more provides satisfactory explanations than having infants or animals that learn. Nevertheless, when we have devised suitable representations and ontologies for a subset of the cases we may be able to devise learning mechanisms that can learn a superclass in which similar principles apply. In that case, by letting the machine learn instead of being programmed, we don't give up the goal of understanding how it works.

1.9 Examples from a possible tea-party scenario

Figure 1: Most people see this clearly

Things for us to think about if we really wish to achieve the long term goals of projects like CoSy include kinds of visual input machines may need to cope with if they are to be able to manipulate household objects, as the robot Fido will need to do. The image shown in Figure 1 is monocular, low resolution and not very well focused, yet humans (at least adults in our culture) seem rapidly to perceive the spatial structure displayed. They are not merely able to identify image regions depicting recognised objects (e.g. cup, saucer, spoon). People can also attend to and think about parts of the surface of each of the objects and the spatial relations between the parts, such as where the spoon is in contact with the saucer, and roughly how far along the length of the spoon the contact area is.

More abstract aspects of the scene that are evident include the approximate angle between the long axis of the spoon and the vertical plane through the cup's handle. You can probably also see which bit of the spoon is below the handle of the cup. None of this implies that what is seen is metrically precise. People can say that one angle is slightly or a lot bigger than another, or one object is slightly or a lot longer, or that one thing is between two others, without necessarily having precise information about exact distances, sizes, or angles. In some cases they probably do not even see enough detail to provide exact information needed for reaching and grasping, but that does not matter if the action is performed with visual feedback: discrepancies can then be corrected online. The degree of precision can be tested by asking people to reach out and touch something after lights have been switched off. The success rate will vary depending on the task. In some cases very precise information is acquired and used, for instance when actions are ballistic so that online control is not possible, such as throwing a ball into a bucket.

The fact that humans can do all these things demonstrates that they are doable by some sort of machine, but leaves open the question what sort of machine since we don't know how humans do them. Research into these capabilities will require the development of appropriate forms of representation for dealing with the kinds of generality and abstraction we have described. A common way of doing that is to use probability distributions over metrically precise sets of representations. But that is not the only way, as the history of symbolic AI shows, as also does the usefulness of maps like typical underground railway maps that retain all topological information and only a little metrical information. There may be new forms of representation waiting to be discovered.

1.10 Variations in tasks

Further analysis of representational requirements could be based on investigating how perceptual, planning, and control requirements relate to various ways in which a cup might need to be picked up, depending on whether it is full or empty, whether it is large or small, whether the contents are liquid or not, hot or cold, whether only one or both hands are available, and depending on what is to be done with the cup, e.g. drinking, pouring the contents, moving to a new location, etc. This involves being able to relate the spatial structure to causal relations. For example the cup can be lifted simply by pushing a finger horizontally through the hole in the handle and then moving the finger vertically, but in that case the cup will tilt (why?) and any liquid therein may come out if the cup was nearly full. However, at the same time as one finger applies lift via part of the 'inner' surface of the handle, another finger could apply a suitable sideways force below the handle. In that case the cup will then not rotate when lifted. At a certain age a child might discover that without understanding it (Humean causation), and later on understand the principles (e.g. perhaps grasping a weak and qualitative version of Newton's laws – Kantian causation). The second form of understanding involves lower level principles that can be generalised to more cases, e.g. the principle that an object can be rigid, meaning that normal forces apply to it cannot change the relations of distance and angle between the parts (though they can change relations like 'above').

The above two-finger strategy is not the only means by which the cup can be lifted without tipping over. Being able to represent and use different lifting configurations, choosing them according to the nature of the task, could form a significant subset of the competences of our domestic robot Fido.

However, it may go through a succession of stages, perhaps as a child does, first discovering empirically what does and does not work, and using only representations that are adequate to express the appropriate generalisations; then later gaining insight into *why* certain ways of grasping and lifting produce or prevent certain consequences. It may be that that shift of understanding requires a shift in the forms of representation, from something that merely suffices to produce action to something that can be manipulated in a reasoning process. The obvious thing to try in our current state of knowledge is to use something analogous to neural nets or statistical rules for the earlier stages, then add some symbolic form of representation. But we shall have to provide suitable bootstrapping mechanisms to facilitate the transitions.

In short, whereas many researchers have focused on requirements that arise out of the noise and lack of detail in the image which can affect relations between image regions and high level classifications, we are arguing that a wider range of competences must be analysed in order to identify suitable forms of representation for a robot that is not merely concerned with classification but can also perform many kinds of actions.

How all of that is done in humans and other animals remains to be seen, though part of the answer will depend on how the information in the image, the information in the scene, and various kinds information linking the two are all represented. In particular, we need to be able to see things we cannot easily express in language, yet are perceivable by and form part of the ontologies of active pre-linguistic agents.

1.11 Vision in action

Some tasks for a crow-challenging robot?[4]

Our approach to the analysis of requirements involves investigating in great detail a variety of examples that challenge the state of the art in order to find out what sorts of representational capabilities, what sorts of ontologies, algorithms, learning capabilities, and architectures need to be developed to enable those tasks to be performed. Consider the following task that is loosely analogous to some of the tasks explored 30 years ago in the so-called 'Blocks micro-world'. Using a robot arm with a two-finger gripper, what actions can transform a situation as perceived in one of these images to the other situation?

[4]Mention of a crow-challenging robot is a reference to Betty, the hook-making New Caledonian crow: see http://users.ox.ac.uk/ kgroup/tools/tools_main.shtml and http://news.nationalgeographic.com/news/2002/08/0808_020808_crow.html

Figure 2: Possible configurations in the cups-world

Of course it would be possible, using techniques from industrial robotics (e.g. perhaps techniques used in automated spot-welding), to program a robot to perform the task for very specific, precisely defined, initial and final configurations where the initial and final relationship between the robot and the objects were severely constrained. However if we consider the images as samples from a variety of configurations where the precise size, weight, shape, colour, and surface textures of the objects can vary, and where the initial and final configurations can vary e.g. with the saucer upside down in either or both configurations, with the spoon in the cup or partly through the handle in one or both configurations, and with lighting varying, and with a human tester slightly moving one or more of the items while the task is being performed or placing an obstacle in the way that requires the motion of the gripper or a held object to be changed in mid-action, we can produce a range of example tasks which for a child (beyond a certain age) or possibly even a chimpanzee would all be approximately equivalent tasks, and would equally easy, but would defeat modern vision systems and modern robot manipulation systems.

In part that is because the requirements are inherently demanding, but in part because most of the research effort in vision and in robotics has addressed completely different problems, ignoring the problems involved in performing actions like this. It is interesting that about 30 years ago a one-armed robot developed in Edinburgh Freddy[5], was able to assemble a few toy objects, e.g. making a toy wooden car from a body with holes, two axles and four wheels, starting with a variety of configurations, but always ending up in the same configuration because the toy had to be assembled with the help of a clamp that held components in a predetermined way. However the speed and flexibility of vision mechanisms available in those days (using an Elliot 4130 computer with about 128 Kbytes core memory) was so limited that the robot could not use vision while moving its arm, so there was no hope of online adjustments, and the parts of the objects assembled could not be varied in size or shape between performances of the task, though there initial positions could be.

1.12 Requirements for the cups-world task

A full analysis of requirements of forms of representation for this task is a demanding research project, but there are some aspects of the task that are already fairly obvious. Prior to the action, the robot has to:

- see which objects are involved in the initial configuration, and where they and there various parts are in relation to one another (as discussed in the section on multi-strand relationships 2.13).

[5]See URL http://www.aiai.ed.ac.uk/project/freddy/

- see and remember the final configuration (which might be shown by demonstrating the transformation, by showing a picture of the required state, or by having another configuration nearby which is to be copied, or described using natural language)
- identify parts of objects, or parts of parts, that are relevant to possible actions the robot may need to perform, e.g. the edge of the handle, or the far edge of the handle or a certain portion of the edge of the saucer
- identify possible processes involving changes in multi-strand relationships as the parts of the scene move (i.e. think about possible multi-strand processes)
- identify possible actions making use of all the previous abilities: e.g. grasping *this thing here* from *this direction* using *these* parts of these fingers, etc.
- think about various effects of actions, including changing effects of both continuous processes and discrete multi-step processes.

Many questions arise about the kinds of representations to be used. For example, could the deliberative premeditation use an abstract action schema (operator) with approximate, qualitative, parameters instead of the more definite actual parameters that would be used if the action were performed?

There are problems here partly analogous to problems of reference and identification in language, except that the mode of reference is not linguistic and what is referred to typically cannot be expressed in language because it is anchored in non-shared structures and processes, , as others have pointed out (e.g. Pylyshyn, Kuipers), internal 'attention' processes (perhaps using internal 'virtual fingers') are partly like external pointing processes that can accompany use of indexicals like 'this', 'here', and 'now'.

We conjecture that solving these problems will, among other things, require revival and extension of the ideas put forward in the generally forgotten paper [Barrow and Tenenbaum, 1978] on relationships between image sub-structures and 'intrinsic scene characteristics' i.e. fragments of 3-D structures of scenes. However we shall have to explore a far richer collection of types of image fragments and scene fragments than Barrow and Tenenbaum did, and add consideration of the relationships of those structure-fragments to affordances. (Gibson's main work on affordances had not been published when that paper was being written.)

Intrinsic scene characteristics are related to features of surfaces and their relationships in 3-D. A robot meeting the requirements for CoSy will have to have an ontology that includes many such things. It will also have to be able to represent: instances of the ontology, generalisations about such instances, questions about such instances, and goals and plans relating to such instances. In Section 2 we investigate in more detail some of the requirements for such an ontology. After that we analyse requirements for formulating questions expressing information gaps. This work has produced a partial ontology of question types introduced in Section 3. After that we return to a consideration of how different kinds of information, and different forms of representation play a role within a complex architecture such as will be required for advanced versions of the CoSy robots.

1.13 Example sub-tasks

Many of the things humans can see and normally simply take for granted, will be relevant to tasks the hoped for domestic robot, Fido, will need to be able to do. Examples include things that we can express approximately as follows:

- This bit is concave. This bit is shiny. This bit is a reflection of the cup. *How is 'this bit' identified?* (Compare 'virtual finger theories' – Z.Pylyshyn's FINST, and recent work by Kuipers on a visual 'tracker' (AAAI-05).)
- As I move left and right those reflections move on that edge of the cup. *How are 'those' and 'that' identified? How exactly is 'reflections move on that edge' understood?*
- This is how I should move my hand to grasp the spoon.
  A different kind of 'this' (manner, route, method) – how identified? Perhaps using partially instantiated parameters in some operator?
- layered interpretations (image, low level image structures, silhouettes (2-d), parsed silhouettes, 3-d structures, affordances, many context-dependent features and relations.
- Attention is different in different parts of the system: different things are selected – *image features, objects, object features, locations in the image, locations in the world (relative to: room, table, object, object part...), relations, actions, ways of doing things, routes, other agents, social interactions, ...*

Figure 3: Different ways of processing an image

- What details are 'visible' in a scene depends on how the visual data are processed (Figure 1, processed in different ways gives left or right set of edge features). Can the system use intelligent decision-making about how to process details in different ways in different places? What features will play a role in such control and how will they be represented.
  (This is one of many kinds of focusing of attention.)

2 Ontology for an active robot

We have begun to develop an ontology for use by a robot in the PlayMate and related domains, for expressing factual information about the environment, with special emphasis on perceivable features, relationships, and affordances, along with kinds of actions that relate to spatial structure. As far as we know there is no existing ontology that can be used. At this stage it is not clear how much will have to be pre-designed and how much can be learnt. In parallel with this an ontology for questions expressing information gaps is being developed, which is reported in Section 3. The word "ontology" has a number of different interpretations. These are explained in connection with the history of investigations of ontology in Section 7, below.

2.1 Introduction: the need for ontologies

One of the long term objectives of the CoSy project is understanding how to achieve, within a human-like robot, integration of multiple capabilities that exist in humans, and to some extent in other animals. This includes perceiving, acting, learning, reasoning, and communicating in a complex environment, such as a family home. This requires the robot to understand both general facts about its environment, such as what sorts of things can exist or occur in it and how various sorts of things normally behave, and also particular facts about what actually exists and what actually happens.

Various kinds and levels of *self-understanding* or more generally, *understanding of information-processing systems, including other agents*, will also be important, in order both to explain and debug one's own actions, decisions and reasoning and also to think about corresponding mental states in others. In particular, being able to think about mental states and processes in others (i.e. states and processes with semantic content in others) is required in order to predict or understand their actions, or to help them. Just as it can be useful to know about size, shape, location, contents, strength, functions, etc. of physical objects, so can it also be useful to know how another individual sees things, thinks about things, understands things, or what the individual wants, likes, dislikes, is afraid of, etc., in teaching situations [Sloman, 1992], and in situations where advice or help is given to someone at an earlier stage of development or who lacks some information because it is out of sight for him.

In all these situations, the robot (Fido, introduced in Section 1.1) will need to deploy an ontology, or more precisely a collection of ontologies. In this paper we begin to analyse some of the requirements for its ontologies, restricting our attention for now to physical environments excluding other intelligent agents. Later work will need to address extensions to the ontology to include animate individuals, which for our purposes can be characterised as things that acquire, transform, use, or communicate information.

Note

The word "ontology" is now often used in AI and software engineering circles, though the concept is much older within philosophy, as explained below. As a result of these different contexts of use, the word has acquired a number of different though related meanings. Section 7 provides a brief history and compares different uses of the word. In an engineering context we can distinguish a *design-ontology*, used by designers to produce some system, which may or may not be an intelligent system, and an *application-ontology* which is used by an intelligent system for whatever its tasks are, which need not include designing anything. This paper is primarily concerned with the "application ontology" for a type of robot that manipulates 3-D objects on a table top. As such it is also a subset of the design ontology for people designing such robots. For now we ignore questions about what is explicitly programmed into the robot and what is learnt through interactions with the environment, with or without a teacher.

2.2 Background, and relation to section on question-ontology

This section is one of two papers on ontology developed within CoSy work-package 2 (Representation of space, objects and context), both submitted as part of deliverable DR.2.1 (Requirement study for representations). The companion paper *'Towards an ontology for factual questions'* presented here as Section 3, is concerned with a largely unnoticed requirement for an intelligent agent, namely the ability to discover gaps in its information, so as to be able to take steps to fill the gaps. Doing that requires a means of representing those information gaps, which in ordinary parlance amounts to a means of expressing questions. So the parallel paper presents a first draft ontology for questions. Identifying information gaps and taking steps to fill them need not involve communication with other individuals (e.g. asking questions) but is part of the information-processing *within* an agent. However it may be a pre-requisite for linguistic competence that involves dialogues in which questions are asked and answered.

The ontology for questions presupposes that the questioner already has an ontology for the *contents* of questions and answers to questions, namely an ontology for factual information. Here we present a first draft incomplete ontology for factual information for a CoSy robot (or perhaps a child, or some other animal interacting with objects in the environment). We ask what sorts of information our robots might need, and summarise in a very general way some of the representational requirements for expressing such information. The topic is potentially vast, but has been constrained here by the needs of the CoSy robot scenarios, initially focusing mainly on a subset of the PlayMate Scenario.[6] This scenario is based on a robot able to perceive, manipulate and talk about 3-D objects on a table top, including eventually assembling complex objects from components (hopefully going beyond Freddy the Edinburgh robot).

The ontological requirements of the PlayMate scenario are related to but different from the requirements for the Explorer. There is a lot of ongoing work in AI on mobile robotics on which the Explorer scenario can build, but relatively little on manipulation of 3-D objects, a kind of task which was accomplished relatively late in biological evolution and which seems to have had a crucial influence on human cognitive evolution. This document focuses mainly on requirements for the PlayMate, but will later be extended to include the Explorer. Insofar as the two requirements overlap, it already addresses the latter.

Another restriction is that the ontology presented here does not include other intelligent agents in the environment. So it is concerned only with objects, events and processes involving inanimate physical objects, except insofar as the agent thinking about such objects may act so as to cause, modify or prevent physical processes in the environment.

2.3 This is about pre-linguistic competence

Although the PlayMate scenario ultimately has to include linguistic competence, it is likely that such competence will build on more fundamental and general forms of competence shared by pre-verbal human children and other animals that manipulate objects such as New Caledonian crows.[7] So this document is a first draft incomplete investigation into some of the kinds of information that a pre-linguistic robot with perceptual and manipulative skills may need to cope with, especially one which, unlike Freddy, knows what it is doing and why.

2.4 Propositional components for a physically embedded information user

There may not be any point trying to cover all possible propositional forms and question forms in the CoSy project. E.g. a subset of the ones a two or three year old child can understand and generate may be plenty, at least for the next few years. It is probably wise to start with mechanisms and forms of representation that do not presuppose the ability to use an external language, since those evolved first and are present first in young children, so it is very likely that they are used by the language-

[6] PlayMate is concerned with a tiny subset of the requirements for Fido!
See URL http://www.cs.bham.ac.uk/research/projects/cosy/PlayMate-start.html

[7] Betty, the New Caledonian crow in the Oxford Zoology lab made headlines in 2002 when she demonstrated that she could make hooks out of wire in order to retrieve food in a basket from the bottom of a tube. See URL http://users.ox.ac.uk/.kgroup/tools/tools.main.shtml

using mechanisms when they develop. In any case it seems that the variety of pre-linguistically comprehensible propositions is very large, and worth understanding in order to investigate how robots might work, independently of what is true of humans and other animals.

An animal, child, or robot may have a complex architecture with many different components using information of different sorts in performing different tasks, often concurrently, e.g. controlling eye vergence, controlling posture, controlling breathing, controlling direction of gaze, causing linguistic input to be parsed and interpreted, controlling the digestion of food, the insertion of hormones into the blood, the pumping of blood, etc. Different subsystems will use various specialised representations tailored to different sub-ontologies and different subtasks. Even information from the same sensory source, e.g. visual information, may be transmitted to different subsystems that use it in different ways to derive different information, represented in different formats, for instance in controlling current actions and in planning future ones, or predicting what will happen next in the environment. For now we ignore all those differences and focus only on subsystems that can use factual information that might usefully be expressed, at least partly, in a propositional form (including non-linguistic internal indexical referring devices, as mentioned previously). Later we return to non-propositional, e.g. spatial, forms of representation.

We consider, as an example, kinds of propositional structures that might be relevant to a robot perceiving and manipulating objects on a table top. (Simpler versions of the robot will cope with only a small subset of ontology presented here.)[8]

2.5 Types of entities that can be referred to

2.5.1 Physical object types

For now 'object' will not be defined, though most of the things called objects will be enduring, spatially-bounded, possibly moving, entities, which may have fixed or changing attributes. A more general notion of object is definable as 'anything that can be referred to', which amounts to the same as Quine's notion of whatever can be a value of a variable. This would include such things as times, numbers, colours, shapes, strategies, styles, theories, proofs, explanations, problems, functions of objects, etc. For now we focus on an ill-defined subset of spatio-temporally located objects (what many philosophers would call 'particulars', as opposed to universals, like shapes, colours, numbers and proofs).

Any object will be an instance of one or more (usually a whole hierarchy) of object types. Some of the object types considered here are fairly abstract, but they all have instances with a spatial location and the possibility of relationships to other objects. Examples would be types of physical object for which we have names:

- Object types categorised on the basis of function and shape: box, tray, cup, mug, saucer, lid, bowl, ball (?)
- Object types categorised only on the basis of shape: cube, sphere, cylinder, ball (?)
- Biological object types: lemon, apple, dog, ...

[8] Some aspects of this investigation overlap with the 'Naive physics' project of Pat Hayes and colleagues. See [Hayes, 1985], though that does not focus on requirements for action. For a commentary with some history see Barry Smith and Roberto Casati (1994), 'Naive Physics: An Essay in Ontology' *Philosophical Psychology*, 7/2 225-244. http://ontology.buffalo.edu/smith//articles/naivephysics.html

- Body parts: arm, hand, finger, fingernail, fingertip
- Parts of objects defined by shape, function, ...: Handle, keyhole, rim, base, ...
- Generic object parts: surface, face, edge, corner (2-D or 3-D), hole, crack,

  NOTE: We could also have typical parts for other things, e.g. plants, pieces of fruit, various animals, various utensils. various kinds of rooms, various kinds of buildings, various kinds of towns, etc.
- Abstract entities that exist by virtue of relations between other entities, gap, opening, mouth, passage, enclosure, corridor, ...
- various entities that can occur on or in a surface:
  - 2-D surface features: marks, texture boundaries, colour boundaries, shadows, edges of those, etc...
  - 3-D (shape-determining) surface features: furrows, indentations, protrusions, curvature extremes, saddle points, edges (where two surfaces meet in a well defined curved or straight line)
- Entities that exist relative to a viewpoint, e.g. visible portion of a surface or object, occluding edge of a surface (e.g. of a sphere viewed from a certain direction – the spherical surface itself contains no surface edges)
- thing (catch all ? Or meta-level concept, like a variable)

[The above is to be extended. Some additional relevant ideas about types of spatial entities can be found in CYC[9]. Some first draft thoughts about requirements are here.[10]]

In a different environment there could be different object types including various plant and animal types, types of furniture, parts of buildings (walls, doors, windows), and various types of out-door objects, such as rocks, trees, clouds, lawns, roads, etc. We ignore those complications for now.

There is no requirement that an animal or robot should be able to see *only* things for which it has names, or which it can recognize. It is clear that we are perfectly capable of seeing, and thinking about a complex structure that we have never seen before, but on parts of which we are capable of performing many actions (grasping, prodding, pressing, pushing, pulling, twisting, bending etc.) The ability to see named, or nameable things almost certainly rests on this more basic ability shared with many other animals that act in an environment about which they cannot talk.

A pre-linguistic animal or robot may, however, use some object types that correspond to recognition states of some internal pattern recogniser without being able to use any external labels for those types even if there are internal labels of some kind. (A draft document discussing, among other things, the evolution of internal labels in biological organisms is available at
http://www.cs.bham.ac.uk/research/cogaff/vis-affordances.pdf[11])

[9] See URL http://www.cyc.com/cycdoc/vocab/spatial-vocab.html
[10] See URL http://www.cs.bham.ac.uk/˜axs/misc/ontology-for-a-manipulator.txt
[11] See URL http://www.cs.bham.ac.uk/research/cogaff/vis-affordances.pdf

2.5.2 'Stuff' types

The object types have well defined instances with boundaries between individuals so that they can be counted. The table-top environment might also include kinds of stuff (labelled in natural languages by 'mass nouns'). If X is a kind of stuff (kind of material), such as water, sand, mud, bread, dough, cotton, etc., then one cannot have two Xs though one can have two pieces, lumps, pools, spoonfuls, (in some cases) squirts, or stretches of X, where the additional (chunker) noun, 'piece', 'lump', 'stretch'. etc, is used to refer to a bounded portion of space filled with X.

For a stuff type X it is not possible to ask how many Xs there are without adding a chunker noun, but it is possible to ask how much X there is, since adding more X to a portion of X, or removing some X from it does not stop what is there being X, it merely alters the amount of X (which is the size, volume, weight, etc. of the portion of X).

In a table-top scenario some subset of the following stuff types might occur:

- Rigid Types of stuff: wood, metal, plastic, glass, (stiff card?)
- Non-rigid Types of stuff:
  - Dry cohesive: paper, string, cotton wool, wire, foil, plastic (e.g. film), plasticine, various plant materials (non-dry twigs, leaves),
  - Dry particulate (pourable, stirrable, spillable): sand, sugar, salt, pepper, ... (Compare piles of smaller and smaller marbles, or cubes: type boundaries may not be sharp, or may be task dependent.)
  - Liquid (e.g. pourable, stirrable, spillable): water, paint (could be more or less viscous, sticky, etc. type boundaries may not be sharp, or may be task dependent.)
  - Gaseous: Steam, smoke, wind, clouds.... maybe in another project

Note that many attributes that an object has, as described below (e.g. rigidity, hardness), will be inherited from properties of the stuff of which it is composed. Understanding this relationship may or may not be required for a particular level of child or animal competence. For particular purposes more sub-divisions may be required, e.g. between breakable and non-breakable rigid objects. But such things could come later.

Insofar as surfaces are important for the PlayMate robot, it should be noted that the perceptual and manipulable qualities of surfaces will be heavily dependent on the kind of stuff involved. Picking up a portion of liquid, a piece of thread, a lump of plasticine, a wooden cube all involve different percepts and skills.

2.5.3 Location types

Physical objects (including chunks of stuff) have locations: they occupy space. Locations of objects are important for many purposes, including grasping them, putting things in them, avoiding them, throwing things at them, identifying them when communicating with others, working out where to go in order to see them, etc.

In addition to objects, events and processes can also occupy or occur at a location. Some locations are not filled by the object (or event) but moved through. For instance a pea in a constantly shaken small box moves around in the volume that is enclosed by the box, and that volume moves around in a larger space as the box is shaken. If a ball is thrown the location through which it moves can be variously identified as the room through which it moves the volume above the table over which it is thrown, the 'tube' which forms the space enclosing all the volumes it occupies during its motion, and no doubt many more.

Besides having spatial locations events and processes can also have temporal locations. As with spatial locations there are different sorts of temporal locations.

Not all types of locations are relevant to all contexts. For example, spatial locations on a large farm may include fields, boundaries, paths, ponds, hills, hillsides, hilltops, valleys, passes, etc. none of which occur on a table-top (although sometimes toy versions do).

A significant part of the history of mathematics has been concerned with space and its properties. Euclidean geometry and more recently topology can be seen as attempts to generalise many of the sorts of notions referred to above, while both abstracting away from features of particular physical objects and processes, and also exploring limiting cases, e.g. as locations get smaller indefinitely, or routes get thinner indefinitely, or lines get longer indefinitely, or structures get more and more symmetrical indefinitely. (Straightness, circularity, perpendicularity, are all examples of symmetry in the limit.) It may be that the human ability to understand such mathematics depends on our evolutionarily older ability to understand, perceive, act on or in, and reason about the less abstract, usually physically instantiated, spatial structures, even though it is possible in principle to produce axiomatic specifications of the abstract limiting cases and reason about them using only general logical capabilities.

There has been an enormous amount of research on ways of referring to or describing spatial locations and regions of various kinds (e.g. by Tony Cohn and collaborators at Leeds University), although much of this research has been done without regard to what the information about locations is to be used for. This paper will not attempt to summarise or reinvent that work, but merely gives some simple examples of what might be useful for a robot that manipulates and converses about objects on a table top, or perhaps a young child playing with toys, or even a nest-building bird.

Spatial locations that an object can occupy or move through, or at which an event or process can occur, that might be relevant to our 'play-and-manipulate' context include at least the following, where the word 'generalised' is used to indicate that we are not restricted to the corresponding mathematical notions.

- generalised points
  (the location at which an object is, or to which it is being moved, or which it has come from in a movement: where the sort of location that is relevant will depend on the sort of object)

- generalised regions of surfaces
  (surfaces can be carved up into regions in many different ways for many different purposes, including possible areas where something could be placed, where something could move, where an individual can reach, or see, or in some contexts regions owned or controlled by someone or something).

- generalised volumes of space
  (Like generalised regions, only 3-dimensional).

- routes of various kinds
  (including 2-D and 3-D routes, both relative to fixed objects or relative to parts of moving objects – e.g. the route of a teardrop down one's face).

- locations on or in other objects, or locations.
  These include regions of surfaces, e.g. 'on his left cheek', or portions of volumes, e.g. 'in his mouth', 'around his left kidney', and also parts of locations and routes, e.g. near the centre or boundary of R, where R is a region, at the beginning of R, where R is a route.

A more complete discussion would also specify different kinds of temporal locations that can be referred to in propositions, including absolute locations and relative locations,

It is worth noting that the forms of representation for locations, surfaces, routes, etc. that a mathematician might first think of are not necessarily the most appropriate for a robot or animal. In particular mathematical specifications will normally be very precise so that when what is scene is unclear or ambiguous a large variety of mathematical expressions may be needed to express all the possibilities, and subsequent reasoning or planning may have to handle a large disjunction, which can lead to combinatorial search. It may be possible to simplify thinking, reasoning, planning and control of actions, if, instead of a precise specification a a less precise form of description can be used, which inherently covers a range of different possibilities (like specifying a time interval rather than a time point, or a spatial region rather than a precise location, or referring to a polyhedron rather than a cube, or tetrahedron, etc.)

CONJECTURE: the move towards less precise concepts makes it possible to avoid complex mechanisms for dealing with uncertainty.[12]

2.5.4 Other types [to be completed]

E.g. agents, animate objects, their mental sates and processes, information-processing systems, mechanisms, etc.

2.6 Attributes (of objects, locations, events, etc.)

All sorts of things can have attributes (and also relationships, discussed later). E.g. physical objects can have size, shape, colour, location (and many other attributes); locations can have extent (length, area, volume, width), shape, or location in a larger location; events can have locations, durations, start or end times, participants, causes, effects, etc.

In the sort of domain we are considering, each sort of entity has a collection of attributes which belong to different types. Each attribute type has a set of possible values, e.g. the type colour can have values red, blue, green, and possibly many others, the attribute height may have either chunked qualitative values, such as small, large, huge, or continuous measure values, e.g. 5cm, 5.1cm. 5.11 cm, etc. In some cases the values are linearly ordered (e.g. height) in others not (e.g. shape, discussed further below). When the values are ordered there may be some direction in the ordering that is naturally construed as *increasing* the value whereas the opposite direction is a *decreasing* direction.

Note: insofar as objects grow, develop, learn, become ill, get damaged, etc. their attributes are not fixed. It is not just that the values can change (e.g. size getting larger, colour getting grayer), it is also possible to acquire attributes in a space that was previously totally absent in the individual. E.g. when a child has learnt to do mental arithmetic it makes sense to ask about the speed at which he multiplies two numbers, whereas at an earlier stage there is no speed because he cannot multiply. Likewise if

[12] As discussed in http://www.cs.bham.ac.uk/~axs/misc/ontology-for-a-manipulator.txt

someone has no arms you cannot ask about the extent of his reach or the strength of his grip. In our initial domain we can ignore growth, learning, development, etc., though we may need to consider objects that get damaged.

Some attributes will have possible values that form ordered sets (e.g. height) whereas others will not (e.g. shape, material, uses, or species in the case of living things). Some may be ordered in several dimensions, especially functional attributes such as usefulness for a task, which may have dimensions like ease of use, quality of result of use, difficulty in learning to use, cost of use, speed of operation, etc. (The multi-dimensionality of 'better' is discussed further here.[13])

Here are examples of types of attributes that objects in a simple robot-manipulation domain might have

- colour
- height
- width
- overall size
- weight
- shape
- volume (e.g. of a container)
- hardness
- rigidity/flexibility
- smoothness
- graspability
- stability
- wetness
- stickiness
- warmth
- material
- function

That is not supposed to be an exhaustive list. In fact for an agent that can act on objects it is a seriously incomplete list, for there will also be a range of important object attribute types that are not definable in terms of intrinsic features of the object (e.g. as size, weight, shape and material are) but depend on what the agent can and cannot do to the object and the consequences of doing, or trying to do those things.

2.7 Affordance-based object properties

In fact some of the attributes listed can be interpreted that way. For instance, instead of hardness being an objective (or intrinsic) attribute of objects (or surfaces of objects), it can be interpreted as relative to a particular agent, depending on how much resistance the surface offers to pressure that the agent can apply. Thus a surface that is hard for a weak, small agent might be soft for a much larger stronger

[13] See URL http://www.cs.bham.ac.uk/research/cogaff/sloman.better.html

agent. Similarly, smoothness instead of being objective may be relative to the amount and kind of resistance to motion when an agent attempts to slide one of its body parts along the surface. What smoothness amounts to for such an agent will depend on what sorts of sensors are activated by such sliding actions.

Even properties like weight and volume instead of being represented in absolute terms such as cubic centimeters and grams might, for some animals and some robots be represented in terms of kind of effort required to lift or throw them, in the case of weight, and in the case of volume in terms of what one can put in the object, e.g. a finger, a whole fist, one's whole body, or in terms of the kind of grasp or posture required to hold the object. That sort of affordance-based notion of volume will of course be closely related to shape, and will not have the kinds of invariance properties that the physicist's or mathematician's notion of volume has. (The use of an affordance-based concepts of volume or 'amount' may account for Piaget's discovery that young children do not understand conservation of volume across transformations that do not preserve shape.)

2.8 Use of predicates vs attribute-value pairs

In many cases each of the possible values of an attribute can be applied as a predicate, producing a proposition that is in some sense an abbreviation of a proposition specifying both the attribute and the value. E.g. 'The block is red and square' could be regarded as an abbreviation for 'The block has colour red and shape square'. The abbreviations lose no information because the name of the attribute value unambiguously (at least in some contexts) identifies the relevant attribute.

However when forming questions if we simply start from the unexpanded abbreviation and create a gap by removing the predicate, the result could be too uninformative as regards what sort of gap filler is required. Thus going from the proposition 'the block is red' to the question 'what is the block?' (or what gap-filler makes the proposition 'the block is ...' true?) does not specify precisely enough the intended question better expressed as 'what colour is the block?', which imposes much stronger limits on what counts as an appropriate answer, insofar as it is equivalent to asking what gap-filler makes true 'The colour of the block is ...'.

Note that the ability to refer explicitly to the attribute by some sort of label (internal or external) need not occur first in development or in evolution. There could be simpler stages where the values of particular attributes are implicitly identified as being of a certain type because of how they are produced by sensory mechanisms and how they are used. Thus the output of a temperature detector fed into some sort of neural control system need not be labelled as being a value of temperature, because the output is connected only to parts of the system that make use of temperature information. So in that context, simple predication without any possibility of expansion to the attribute value form is all that can occur and it is all that is needed. Only when more sophisticated architectures evolve, or develop, performing more complex functions, is it necessary or possible to use explicit attribute labels.
(**NOTE:** Explicitness here has nothing to do with consciousness: it is merely a matter of whether some structure or process exists which has a certain sort of function in the system.)

2.9 Object relations

Physical objects, locations, routes and other things can have relations to all sorts of things, of the same type or different type. The relations differ in 'arity', depending on how many things are involved. E.g. they may be binary, ternary, or n-ary for any integer n.

A major distinction can be made between relations based on attribute values (e.g. X is taller than Y, means the value of X's height attribute is larger than the value of Y's height attribute), and likewise for other relations of ordering, e.g. X is between Y and Z in height, relations involving measures, e.g. X is 14 cm taller than Y, etc. The types of attribute-based relations involving attribute A that are possible will depend on the structure of the set of possible values for attribute A, e.g. whether it is totally or partially ordered, whether it is continuous, dense, discrete, etc., whether the values are themselves structured (e.g. vectors or trees or nets), and so on. For example, one object can be the mirror image of another if the shape of the first and the shape of the second are essentially the same except for a single reflection, so that together the objects can form a symmetric structure. Another kind of relationship that can hold between X and Y depending on shape is X being isomorphic with a part of Y, e.g. if X is a sphere and Y consists of two spheres joined by a cylindrical rod.

There are other relations that are more complex and are not derivable from or definable in terms of the intrinsic properties of the objects but depend on where they are located in space and or time. We'll call those extrinsic relationships, and return to them later.

2.10 Intrinsic relations: Relations based on attribute values

Some of the relations involve comparisons of attribute values. There are two main sub-cases, namely equality and inequality of values, and the inequality sub-cases can be further sub-divided, depending on the structure of the set of possible values, e.g. whether it is totally ordered, whether the total ordering has a kind of asymmetry supporting a notion of more or less, whether it is a continuum of values or not, whether the values vary in different dimensions so that they form some sort of vector space, etc.

For example, if attribute A has possible values Av1, Av2, ... then propositions can be expressed regarding two objects X and Y saying

- 'X is the same A as Y' (e.g. X is the same colour/height/size as Y),
  meaning something like: the value that fills the gap in 'The A of X is ...' and the value that fills the gap in 'the A of Y is ...' are the same thing.

  Variants of this use a relation weaker than having identical values, and instead specify values that are close in the ordering, which could be expressed as 'X is like Y in respect of A', or using phrases like 'similar to' 'close to', etc.

- 'X is more A than Y', or 'X's A is exceeds Y's A' (e.g. X's size, height, hardness, smoothness, exceeds Y's)
  meaning something like: the value that fills the gap in 'The A of X is ...' is 'bigger' (or 'higher') than the value that fills the gap in 'the A of Y is ...'.

In many natural languages there are alternative constructs for expressing this sort of idea, some of which use the attribute name with a modifier, e.g. 'more height', 'less weight', some of which use a comparative form of an attribute value, e.g. bigger, weightier, smaller. As usual we are not concerned with the precise syntax but with what might be expressed.

Although most of the examples given involve relations between physical objects there are also relations between events and processes that are intrinsic. For instance one of the attributes of a process is its duration. So one process can have a longer or shorter duration than another, or a duration that differs by 2 seconds or 2 days, etc. One process may involve something happening faster than another, e.g. rotation, or colour changes, or speed of motion.

The examples given so far of relations based on attribute values are all binary. However insofar as as attribute values are ordered, and many different relations can exist between items in an ordered set, there will be derived non-binary relations between objects with those attributes. For instance

- Ternary relations
  For example: X is between Y and Z in A (e.g. in height, in colour, in size) means something like the attribute of type A of X is between the attributes of type A of Y and Z (in the ordering of attribute values of type A). This could be true if A is height, and the attribute values of X, Y and Z are respectively 6cm, 3cm and 22cm, or, 6cm, 22cm and 3cm (since the proposition says nothing about the relative position of Y and Z). There are many other Ternary relations, such as that X and Y differ from Z in attribute A by the same amount.

- Quaternary relations
  As with ternary relations many quaternary relations can be formulated on the basis of attribute values, for example the proposition that W differs from X in A more than/the same amount as/less than Y differs from Z in A.

Exactly which sorts of n-ary relations can exist between objects with attributes of type A will depend on the structure of the space of values of A. If there is a small discrete set of values, e.g. small, medium and large, then far fewer relationships will be possible than if there are many values, or if the values form a continuum (or dense set).

If the set of attribute values has some sort of distance metric, then comparisons of distances can be used to construct yet more attribute-based relationships, which are not necessarily elegantly expressible in English. Are some languages more suited to saying this sort of thing than others? Examples are:

- X is more tall than Y by a larger amount than Z is more tall than Y.

- W is more red than X by the same amount as Y is more red than Z.

In the case of some qualitative attributes, such as colour, taste and smell, we find it useful to describe things in terms of closeness to particular important or interesting attributes, for instance describing X as more red than Y (redder) or more sweet than Y (sweeter). In contrast we don't say of X's height that it is more 10cm than Y, as a way of saying that X's height is closer to 10cm than Y's height. However in principle there is no reason why a particular language should not have such syntax added to it, if that would prove useful, e.g. because some heights have particular social, religious or economic importance, for certain objects.

All the types of propositions expressing relations derived from ordering or other relations between attribute values allow the removal of components leaving gaps on which question-forming transformations can operate. For example, introducing different gaps in the proposition *'X is between Y and Z'* in A, could produce such questions as

- What is between Y and Z in colour? (one gap)
- In what respect is X between Y and Z? (one gap)
- What is X between in colour? (two gaps)

2.11 Relations based on shape

The examples so far have used attributes whose values are essentially points in a space of possible values, where the points themselves (e.g. heights, widths, colours(?)) have no structure, though the space may be wholly or partially ordered, discrete or not, finite or not, and so on. One particular set of attributes has elements which instead of being mere points in a space (or even vectors in a fixed dimensional space) are structures of varying complexity, namely shapes. Two objects can have shapes that in turn can have attributes of varying complexity, for instance, number of edges, number of holes, relative lengths of holes and edges, being convex or concave all over or in parts of the surface, having grooves, having dents and many more.

The precise set of shapes that any individual can think about and ask questions about will vary from species to species as well as from individual to individual within a species. Of particular interest in robotics and (human or animal) psychology are the relationships between shapes of objects and the actions that individuals can perform. Whatever set of names we have for shapes will never be enough because shapes can be made more and more complex indefinitely by adding components, replacing components, joining two or more shapes in various ways to make a more complex shape, adding grooves, hollows, bumps, holes, and so on. The fact that such transformations produce shapes that cannot be recognised does not imply that they cannot be seen. On the contrary, seeing shapes needs to come before recognising them when they are re-encountered.

The point of all that is that it is a mistake to think of perceptual systems as producing only propositions about features and relationships of recognised (or named) whole objects. Rather there will be many object fragments and surface fragments that are seen, and relations between them that are seen, including not only relations within images, which don't concern us for the moment, but also relations within the 3-D environment.

Thus if there are propositions expressing what is seen they may be in part like parse trees summarising a network of relationships between simple or complex fragments, which may or may not be recognised fragments.

The process of question formation through gap manipulation then will have to operate on these structures, and the results may not be easily expressible in familiar human language. For example it may be possible to wonder whether a partly visible object or fragment has a shape for which there is no label but is (internally) identified by the perceiver as 'that shape', referring to the shape of a wholly visible component of the scene.

Similar comments apply to perceived events and processes, such as the motion of an object. For instance a perceiver may wonder what kind of action that it can produce could cause X to move as Y is seen to move, where there is no name or prior label for the latter type of motion. (E.g. think of a dancing student wondering how to replicate a teacher's actions.)

2.12 Extrinsic relations: Spatio-temporal relations

So far we have considered intrinsic relations based on attribute values, including relations like being more or less A, where A is an attribute, or having an attribute value between other attribute values. Among these are relations like being the same shape or having a shape with fewer concavities, that depend on the shapes of the objects standing in the relations.

There are also *extrinsic* relations that depend not only on the attributes of the objects but also where they are, and how they are oriented in space, and in the case of events and processes some some of the relations depend on temporal location, i.e. when they occur, when they start, when they end, etc.

There are also very many relations that depend on spatial location, orientation, and shapes of objects. Some of these involve metrical relationships, such as being similar in shape, or being a certain distance apart, or being adjacent, whereas others can be thought of as purely topological, such as being inside, or being linked (as two rings can be). Others are a mixture.

For instance when a hook and a ring are joined up as in a toy train where one truck pulls another, it is possible to think of the relation between hook and ring that makes pulling happen as metrical rather than topological because bending the hook so that it is lifted out of the ring and then makes pulling impossible will change only metrical properties, not topological properties. However if we think of the hook as part of a ring with a missing part then we can think of the hook and its missing part as linked to the ring, a topological relationship.[14]

2.13 Multi-strand relations and multi-relation facts

Although two point-like objects may have a small set of relations of distance and direction (which can be combined into one relation vector) when two objects have complex structures they can have multi-faceted relations because the relations between the two objects are composed of many relations between their parts, as noted in Section 1.7. For example if a small cube (SC) rests on a large cube (LC) we can have relations like the following between the objects and their parts:

- SC is immediately above LC
- bottom face of SC and upper face of LC face in opposite directions
- bottom face of SC is in contact with upper face of LC
- Edge E3 on bottom face of SC coincides with part of Edge E17 of top face of LC
- Vertex V5 of SC coincides with vertex V6 of LC
- Face F4 of SC meets face F20 of LC perpendicularly

and many more. Thus a vision system that can merely report that SC is above LC is inadequate for tasks where multi-strand relationships are important, as in many assembly and manipulation tasks.

Many of the objects in the environment are complex objects composed of parts. It is often thought that complex objects are best described in terms of hierarchical decomposition yielding a tree-structured description where nodes are mostly linked by a ‘part-of’ relation. But what we actually see and what we often need to know is far more than the part-whole decomposition. A partial acknowledgement of this point is in this online ontology for bicycle parts, which besides the ‘part-of’ relation includes *constraints* between parts:

http://www.ksl.stanford.edu/htw/dme/thermal-kb-tour/bike-domain.lisp.html

Moreover, as also noted in 1.7 a *process* involving two structured objects moving in space will involve a collection of sub-processes where several relationships all change in a systematically constrained way. A robot observing and producing such processes needs to be able to represent and reason about *multi-strand changes in relations*. We could call these ‘multi-strand processes’. Multi-strand processes are, of course, familiar in physics and engineering, where the most highly developed form

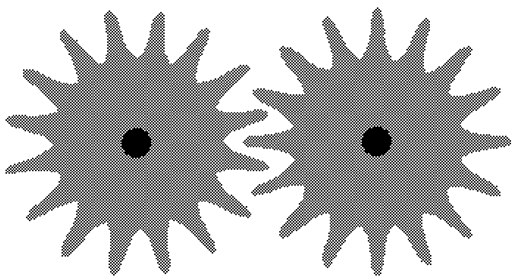
[14]The need to move between more or less metrical and topological relationships in solving problems is discussed in a little more detail in [Sloman, 1998] and in A.G. Cohn, S.M. Hazarika, ‘Qualitative Spatial Representation and Reasoning: An Overview’, (2001). http://www.comp.leeds.ac.uk/qsr/pub/funinfreview.ps.gz

of representation is a collection of differential equations relating such things as positions, distances, volumes, currents, voltages, pressures, tensions, torsions, and their rates of change and possibly also higher order derivatives. But that form of representation does not capture the collection of changing relationships that are typically of interest in an environment in which an animal or robot perceives and acts. Verbal descriptions are also often clumsy and inadequate, which is one reason why we often prefer diagrams, working models, or complex gestures or demonstrations, for instance in teaching someone how to tie shoe-laces, or how to dress a child. Logical descriptions can handle arbitrary collections of relationships and changes of relationships but it is not clear what would be best for a robot involved in domestic manipulative tasks.

2.14 Multi-strand relationships and causality

This is deeply connected with our everyday notion of causation, especially the (Kantian) deterministic notion of causation, rather than the (Humean) correlational, probabilistic notion of causation that has recently received a great deal of attention in AI and psychology (e.g. in connection with Bayes nets).

For example, part of the notion of an object being *rigid* is that its parts always move in such a way as to maintain a collection of metrical relationships including distance, direction, parallelism and similarity. Understanding what this involves, combined with the notion of impenetrability gives a way of thinking about what happens if two rigid bodies are close together, constrained in some way, and one of them starts to move, such as two meshing gear wheels each constrained to rotate around an axle at its centre, as in the figure, where most people will easily see what must happen to the wheel on the right if the wheel on the left rotates clockwise, even though this is a complex collection of multi-strand processes in which new relationships come into being and then later cease to exist (e.g. contact between teeth of the gears).



Depending on the shapes of the wheels and their teeth there may be situations where rotation is seem to be impossible unless the wheels move apart. More subtle features of the changing relationship can give clues as to whether there will be friction and wear during the rotations, an example of longer-term causation. It is not clear whether existing forms of representation used in computers can easily be applied to the task supporting such causal reasoning.

There is a long history of attempts to develop symbolic forms of representation suited to multi-strand relationships, including semantic nets in the 1960s, Minsky's Frame Systems [Minsky, 1978], formalisms such as KL-ONE, object-oriented extensions to AI programming languages (such as FLAVORS, LOOPS and CLOS), and more recently systems such as CLASSIC using description logics[15]. These forms of representation are essentially all discrete and suited to reasoning about discrete rela-

[15]Online tutorials on description logics are listed here http://dl.kr.org/courses.html

tionships, differences, and changes, although in principle some of the numerical attributes represented may have real values. In parallel with these developments many people have attempted to understand the significance of visual forms of representation (see the references in DR.02.01).

A complete survey of spatial relations, spatial structure, spatial changes, causation in spatial changes, and ways in which they can be represented in intelligent agents would require a book at least. However a few points are worth making about some of the kinds of variation a perceiver may need to take account of.

One of the requirements for a visual system in a robot observing complex motions of complex objects is that it should be able to represent combined translations and rotations of rigid, jointed and flexible objects whose parts move relative to one another while the whole object moves, including for instance a lion seeing a deer during a chase. Birds building nests by inserting one twig at a time in such a way as to form a new complex rigid object have related requirements. The processes a human-like domestic robot will have to be able to cope with, like those already coped with by many animals, involve a mixture of *continuous* and *discrete* changes (e.g. a cup approaches the rack in a dishwasher through empty space and then is in contact with the rack and constrained from moving downwards or horizontally, or rotating). Although there has been much psychological research on human abilities to visualise continuous changes, as far as I know there are no forms of representation used in AI that meet all these requirements, though it is possible that special cases can be handled by a mixture of discrete data-structures and techniques based on differential equations or spreading activation mechanisms in neural nets. It is possible that a survey of techniques used in systems for generating dynamic displays in CAD systems and computer games will reveal something relevant.[16]

One form of representation that looks particularly promising involves the use of *simulations*, where a collection of data-structures and programs provides a virtual machine running in the computer with structural relations that correspond in a useful way to those in the objects represented, and where the programmed constraints on the changes in the data-structures correspond to constraints in the physical environment. However, for many purposes a very detailed and precise simulation would be both impractical and overkill, e.g. for high level planning of actions and their consequences. It seems that investigating the use of varieties of simulations at different levels of abstraction could yield promising insights into ways of giving a robot something like human abilities to perceive, act in and think about the sort of domestic environment the robot Fido will have to deal with. We could start doing this for very simple table-top manipulations. I conjecture that the use of this sort of simulation process will turn out to be the basis of many observations based on introspection about the role of visual or spatial representations in reasoning and problem-solving. The actual processes are not spatial but self-monitoring mechanisms interpret them as spatial because that is part of the mechanism whereby they are usefully applied to reasoning about spatial processes.

2.15 Local vs global spaces .... to be extended

Some relations are relative to an object's intrinsic frame of reference, e.g. in front of, on the left side of, behind, when applied to something like a car, a horse, or a person. For objects that don't have an intrinsic front and back, such as a ball, a box or a banana, the very same words can be used to specify a relationship based on the perceiver's viewpoint, so that 'in front of' is interpreted as equivalent to

[16] A useful collection of ideas is summarised in Cognitive Models of Dynamic Geographic Phenomena and Their Representations: Report of a Specialist Meeting held under the auspices of the Varenius Project (October 28-31, 1998). Eds Stephen Hirtle, and Alan MacEachren. Online at
http://www.ncgia.ucsb.edu/Publications/Varenius_Reports/Cognitive_Models.pdf

something like ‘nearer to me’. It is also possible to refer to front-back relations between things that don’t have a front-back distinction, by referring to another object that does, e.g.

‘The ball was in front of the horse and the box in front of the ball’

For both the PlayMate and the Explorer it is important that the visibility of objects can change without their existence being accepted (Piaget on object conservation.) E.g. things on the table that are temporarily out of view because they are occluded, or because you are looking in a different direction can still be remembered. A video of a three year old child playing on the floor with a toy train set that surrounds him illustrates this very well

http://www.cs.bham.ac.uk/˜axs/fig/josh_tunnel.mpg [5MB]
http://www.cs.bham.ac.uk/˜axs/fig/josh_tunnel_big.mpg [15MB]

The retinoid mechanism, and related mechanisms described in [Trehub, 1991] were designed to explain this, along with other phenomena related to the fact that visual information persists across saccades, though it is not clear that it is adequate to the task.

Moreover, there are big differences in the representation of local space (i.e. immediately surrounding the viewer) and the representation of more global terrain, such as one’s house, village, or country. The local space seems to be at least partly based on a viewer-centered form or representation (which therefore has to change as the agent moves around), and is rich in 3-D information whereas the global space is more persistent and seems to be (at least for humans, though perhaps not for flying animals) largely 2-D, as is the case for most computer models of way-finding and learning about locations.

Some of the local and global spatial representations make use of frames of reference defined by specific objects, including the agent’s own body, another agent’s body, part of a body (e.g. an arm), the room, the table, the toy aeroplane on the table, etc. Some objects may simultaneously be represented as occupying locations within different reference frames. E.g. fingers have locations relative to the hand, the arm, and the whole animal, as well as locations in the room. Which ones are important and need to be made explicit can change dynamically according to task requirements. For instance relations between finger and nose are normally not relevant, yet they become relevant when someone is seen as scratching his nose, or an agent needs to scratch its nose.

This argues against assuming that there is one hierarchical part-whole model for each complex object from which all information is derived as needed (e.g. the Marr/Nishihara model based on connected generalised cylinders, and Hogg’s and Mackworth’s PhD theses.(???))

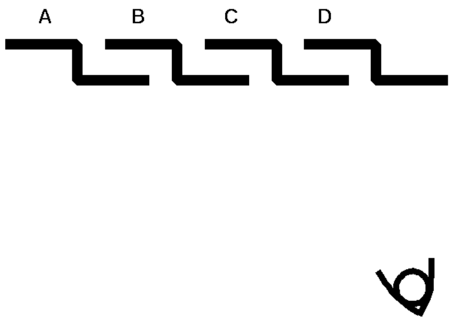
2.16 Integration, zooming, etc.

If a robot uses multiple forms of representation, that raises questions about how they are integrated. In principle there be only a collection of completely separate task-specific representations used by different sub-modules in the architecture which perform different functions. (This is one of the themes of [Minsky, 1987].) However, it is clear that humans (at least adult humans in our culture) have an ontology that somehow combines the local 3-D and large scale 2-D information within larger scale 3-D (or more accurately large scale 4-D space-time) — information structures, that allow zooming in and out according to the needs of the task. Recently such zooming operations, which in the past we could perform only in our heads, have been made available on computers, e.g. in CAD systems, but at present such systems do not share our understanding of what they are doing, and the representations are useful only for generating images or controlling specific actions, such as parts manufacture or checking design constraints. E.g. they do not, as far as I know, support understanding of how a small scale 3-D structure (e.g. a toy house) can provide information about a larger structure (the real house) that might be useful for route planning.[17]

2.17 Indeterminacy of spatial concepts

Many of the relationships for which we have names, like the ones mentioned above also have a kind of indeterminacy as to what regions are involved: e.g. if you look at a box in the middle of the table, and consider under what conditions you would describe a ball placed anywhere on the table as being to the left of the box, there may be considerable regions of indeterminacy. It is sometimes thought that such cases reflect statistical concepts with a probability distribution. An alternative view is that they are actually higher order concepts with an implicit reference to the *purpose* or *task* for which the relationship is being described. If there is no specific purpose or task the question about where the boundaries between the regions (left, right, front, back, etc.) are is essentially pointless, like asking whether a specific rock is or is not big without having any criterion for things being big enough. If forced to answer such pointless questions (e.g. in a psychology experiment asking which locations are to the left of that box) people may then use a default rule that covers many of the normal reasons for mentioning the ‘left of’ relation.

Instead of postulating a probability distribution it may be more accurate to regard this as a case where there is no answer, and therefore the statement ‘X is to the left of Y’ in such a case has no truth value, like the statement ‘My lawnmower is better than my car’ taken out of any context specifying the relevant respects of comparison.



Many spatial relations involve an order which depends on location or a combination of location and other attributes. For instance things may be ordered according to their distance from some reference location (e.g. the viewer) or according to what partially occludes what. ‘X occludes Y’ often implies ‘Y is further away than X’ but need not where complex shapes are involved. For instance, when the figure is viewed from somewhere below D, A occludes B, B occludes C, C occludes D, but A may be furthest from the viewer and D nearest.

2.18 Affordance-based relations – and embodiment

For a perceiver that is capable of manipulating objects, many spatial relations are concerned with what motions are possible or constrained. For example saying that there is a gap between X and Y could in some contexts refer to the possibility of some object currently under consideration being able to move between X and Y. There will also be many relations for which we don’t have verbal names but which

[17] In that respect they are like the two-year old children described in a recent article ’Mindful of Symbols’, Author Judy S DeLoache Scientific American, August 2005, pp 60–65.
http://www.sciam.com/article.cfm?articleID=000ACE3F-007E-12DC-807E83414B7F0000&sc=I100322

an expert manipulator learns to perceive and think about as relations that facilitate or obstruct actions, or which change during actions.

For example, when a child lifts a cut-out picture of a car from its ‘home’ in a flat sheet of wood, and then attempts to replace it, she may be unaware of the requirement to line up the boundary of the car shape with the boundary of the hole it left behind, and merely think of putting it back in the same general location. Such a child can be seen to try to get the piece back into its hole merely by pressing hard on it.

At a later stage the child has learnt about the significance of the additional spatial relationship, and appreciates the need to do some sliding and rotating as well as pressing, in order to get the shape back in place. At that stage the child need not have any *name* for either the relationship of alignment or coincidence between corresponding parts of the outline of the car and of the hole, nor for the changing relationships that occur while it is being moved around in order to get it into position to drop into place. Yet the child may understand the relationships, may be able to think about what sort of misalignment exists (e.g. where there is a boundary mismatch) may be aware of information gaps, may change its viewpoint, or move its hand out of the line of sight, in order to obtain missing information and then use the newly acquired information in order to complete the task. Thus there is an intermediate stage in which a question is considered, for which no linguistic expression is available (at least not to the child), and an answer found, for which there is also no linguistic expression available.

Similar remarks can be made about tool manipulation and manipulation of items of food or nest-building materials in other animals, e.g. squirrels, monkeys, chimpanzees, crows, etc., none of which can use anything like a human language.

2.19 Affordance-based and mathematical relations

In ways that remain unexplained, a learner seems to be able eventually to understand that there are spatial relations that may have originally been learnt about through their relevance to action but which can be described in a manner that is independent of how they are perceived or used by acting animals or robots but can be expressed in an abstract manner in terms of their own intrinsic properties and relations. The invention of Euclidean geometry as axiomatised first informally a few thousand years ago, and then more formally in the 20th century, was perhaps one of the most important such discoveries in the history of human thought, and it is not clear whether any other animals have such an affordance-free conception of space and spatial relations.

One of the tasks for CoSy and related projects in the longer term is to explain how that conceptual change can occur within an individual: what sorts of forms of representation, mechanisms, architectures are needed.

2.20 Self-knowledge

Nothing said here about needing, acquiring and using information of a particular type implies that the child or the animal *knows* what it is doing. That kind of self knowledge requires additional architectural resources and meta-semantic capabilities supporting reflective self-description. It is possible that many animals that are very good at thinking about and solving problems are totally incapable of thinking about or even being aware of their own problem-solving. This is probably true of much of what happens in young children: the meta-management architectural layer required for self-understanding does not seem to develop until some time after many other capabilities have developed. In general, knowing X, and knowing that you know X require quite different mechanisms, including different forms of representation whose semantic contents presuppose very different ontologies. The latter requires a meta-semantic ontology which includes things that manipulate information, whereas the former does not.

2.21 Further details [to be reorganised]

Some comparative relations can be used as ternary relations, e.g. 'X is further from A than from B', 'X is further from A than Y is'

Other things to sort out

- further
- moreleft
- moreright
- moreinfront
- morebehind
- derived (parametrised) comparative relations: closerin (colour, height, size, distance, ...)

Similar kids of ternary relations can be generated from features that allow 'more' or less', e.g more red, more high, more fast

more kinds of relations

- QUATERNARY
  - W is as far from (close to, higher than) X, as Y is (to) Z
- GEOMETRICAL RELATIONS
  These can involve arbitrarily many objects. I.e. the relations can be
  - unary (is a triangle),
  - binary (contains)
  - ternary
    - form an equilateral triangle
    - form an isosceles triangle
    - are collinear
    - form a acute/obtuse/right angle (X Y Z)

  etc.

OTHER PROPOSITIONAL FORMS

- x is P
- x is at X
- x is a K
- morethan (X's Y's)

- Existentially quantified
- Universally quantified
- Most (Xs are Ys)
- More (Xs are P than Ys are)
- Numberof Xs is N
- Numberof Xs has some numerical property.
  (E.g. even, graspable simultaneously, beyond our budget, affordable, ...)

2.22 What about a non-linguistic (pre-linguistic) agent?

A significant subset (what subset?) of the above proposition types and question types may be capable of being considered by animals without human language and pre-linguistic children. The questions will not be posed externally, but internally, and will determine goals for cognitive and other processes.

So that raises the question of what sort of formalism could do these things *within* a cognitive agent, as opposed to *between* speakers?

It may be that answering that question will give us deeper insights into what goes on in speakers, for they will presumably, to some extent, build on the pre-verbal capabilities that evolved earlier.

Later on the internal processes were certainly expanded by the availability of linguistic constructs – e.g. things like

> 'By how much is the distance from A to B bigger than the distance from C to D?'

Could a pre-linguistic animal (or child) wonder about such a question?

Some of these issues are posed in relation to images of a cup, saucer and spoon in this short presentation: http://www.cs.bham.ac.uk/research/cogaff/challenge.pdf [18]

2.23 Representations in reactive systems: speed, and fluency using implicit representations

In addition to the pre-linguistic cognitive mechanisms we need to refer to what are sometimes referred to as *sub-cognitive mechanisms*, a notion that overlaps with, or may be equivalent to some uses of, the notion of *reactive mechanisms*.

The spatial competence of an octopus is amazing. Many animals, including insects and and other invertebrates, and perhaps many vertebrates, such as fish, amphibians, reptiles, many birds and some of the less intelligent mammals all seem to have considerable spatial competence yet probably lack the forms of representation and the ontologies described here. (It's hard to be absolutely sure about exactly what forms of information-processing they do and do not have, but let's ignore that question for now.)

For example bees land on flowers suitably positioned in order to get nectar and pollen, they find their way back to the hive, they can later return to the same location, and they can even communicate direction and distance to other bees by doing their famous dance. Spiders make intricate webs and when insects get caught in the web a spider can take appropriate action. Many animals find and eat food which may be animate and try to escape or inanimate and have to be removed from whatever it

[18] Elaborated in CoSy Deliverable DR.02.01 on Requirements for representation

is attached to, e.g. grass, berries, leaves, nuts, fruit, etc. Many animals make nests and termites make elaborate cathedrals. There are also complex spatial actions involved in mating, and egg-laying.

For many kinds of actions behaviour may be controlled by information that is implicit in the changing patterns of activation of various parts of a control system that includes sensors, effectors, various kinds of feedback and feed-forward mechanisms, and may even be hierarchically layers. In such a system actions and the consequences of actions are represented only while subsystems are active in producing and controlling those actions. Likewise the contents of sensors are represented only while the sensors are being stimulated. Past actions and percepts are not representable, though there may be gradual changes produced by hysteresis and and adaptation mechanisms, as in neural nets that change synaptic weights over time.

These *implicit* means of encoding information in the states of dynamical systems can be very powerful and are probably the only form of representation available to the vast majority of biological species. However[19] such forms of representation have major limitations and evolution seems to have discovered, in at least some species, additional more explicit forms of representation that support such competences as thinking about unobserved objects or spatial regions, predicting sequences of future events, explaining observed phenomena in terms of unobserved phenomena, making structured plans, having motives that refer to the remote future, and even thinking about thinking. These are often thought to be what distinguishes cognitive systems (though some proponents of dynamical systems claim these are just a special case, which is either trivially true because everything that works is a dynamical system, or false if dynamical systems are characterised on the basis of models from 19th century physics based on differential equations).

However, for the purposes of CoSy, if robots are to be capable of performing fast fluent actions such as rapidly and smoothly picking up a cup of tea and passing it to a person, then many of the details of such a process will probably need to be controlled by an insect-like reactive mechanism with many feedback loops and continuous (analog) online control systems, unlike a typical plan execution mechanism.

It appears that in humans, and possibly other animals, there are both sorts of systems and they are closely integrated in complex ways that are not yet understood, but one of them could involve the more symbolic, cognitive, deliberative system making use of explicit representations *training* a reactive, dynamical system that uses only, or mostly, implicit information structures. Repeating actions under the control of a deliberative mechanism seems to be a common means of developing speed, accuracy and fluency in many human activities, including running, jumping, grasping, throwing, many athletic skills, musical skills, linguistic skills and skills at operating complex machines, including driving bicycles and cars. What is not so generally appreciated is that similar extensions of competence can occur in low-level perceptual mechanism when trained in various repetitive tasks under the control of deliberative mechanisms. Examples include learning to read text fast, learning to read music, learning to ‘read’ the behaviour of opponents in various sporting activities, and perhaps learning to read facial expression, posture and gestures in many forms of human interaction.

Another kind of relation between the two sorts of mechanisms can be the development of reactive mechanisms that generate new motives or even alarm signals for the deliberative system so that its focus of activity can be switched or its actions modulated when something important has changed in the environment (or in the body).

If all that is right, then it may also be a long term requirement for CoSy robots that they too should

[19]As discussed in an online draft, incomplete paper, http://www.cs.bham.ac.uk/research/cogaff/sloman-vis-affordances.pdf

have both reactive sub-symbolic and deliberative explicitly symbolic mechanisms if they are to be able to function as domestic helpers for the aged or infirm, for example.

(An alternative possibility is that speeds of computation will increase so fast that extremely fast cognitive mechanisms constantly planning and re-planning at speeds that are impossible in brains will be able to take over all the functions that in humans, apes, etc. are performed by evolutionary very old reactive mechanisms. There are reasons by physical impossibilities, including power-weight and other constraints, may rule that out, but the issue will not be discussed further here. At least we seem to have an existence proof of the effectiveness of a hybrid system, even if we have no proof of uniqueness or optimality.)

An example of an attempt to produce a hybrid symbolic/neural system meeting some of these requirements can be found in Ron Sun's CLARION project http://www.cogsci.rpi.edu/˜rsun/clarion.html Discussion of how well it meets the requirements specified here will be left to another occasion. It is discussed briefly [Sloman and Chrisley, 2005][20].

2.24 Nature Nurture Issues

Suppose it is true that at least some natural language users are capable of understanding the full variety of types of propositions and questions sketched here, in at least some sense of "understanding" involving knowing what sort of thing counts as a correct answer, without necessarily knowing the answer or even known how to find or verify the answer – because that might require technology that has not yet been invented, or in some cases might require use of a theory that has not yet been discovered). This raises the question how such understanding arises?

One possible answer is that it arises out of the use of the sorts of mechanisms sketched in a paper presented at IJCAI05[21] discussing the precocial/altricial spectrum, where altricial species (or altricial skills) instead of being restricted to

(a) learning by the use of positive and negative reinforcement mechanisms that gradually transform weights in some sort of statistical mechanism (e.g. a neural net) can also be learnt by
(b) a combination of more 'symbolic' mechanisms:

- exploring (randomly or otherwise) effects of many kinds of distinct sensory patterns and action types (starting with innate ones)
- noticing, categorising, and inventing (internal) labels for 'interesting' cases
- using some (innate? learnt?) 'syntactic' mechanism for internally recombining the labels in various ways to generate new, larger perceptual and action structures, which are also explored and if found interesting stored, labelled, and made available for re-use

Such mechanisms can support discrete, creative learning and discovery steps producing quite large changes in competence.
(The idea goes back at least to work by Oliver Selfridge, as reported in this toy demonstration http://www.cs.bham.ac.uk/research/poplog/teach/finger[22] which can be run using the poplog system.[23]

[20] Available online here http://www.cs.bham.ac.uk/research/cogaff/04.html#cogsys
[21] Co-authored with biology Jackie Chappell, http://www.cs.bham.ac.uk/research/cogaff/altricial-precocial.pdf 'The altricial precocial spectrum for robots'
[22] See URL http://www.cs.bham.ac.uk/research/poplog/teach/finger
[23] See URL http://www.cs.bham.ac.uk/research/poplog/freepoplog.html

Defining that hypothesis about altricial mechanisms properly will have to wait for another occasion. For now it is important only that these mechanisms have implications both for the innate ontology of the robot or animal and for how the ontology develops over time.

2.25 Some References on ontologies

These references were found with the help of Google. The list is neither complete nor authoritative nor representative. It's just what I found in a fairly short time spent searching, with some help from Push Singh.

Pylyshyn, Z.W., 'The role of location indexes in spatial perception: A sketch of the FINST spatial-index model,' *Cognition* 32, 65-97, 1989.
http://ruccs.rutgers.edu/ftp/pub/papers/cognit89.pdf[24]

See OpenCyc on 'Spatial Relations'
http://www.cyc.com/cycdoc/vocab/spatial-vocab.html[25]

> This document describes collections, predicates and other Cyc constants that are used to represent spatial objects and relations. See also documents for Groups, Quantities, Movement, Paths & Trajectories, Parts Of Objects, and Geography.

The 'Simple bicycle design ontology' referenced in Section 2.13 is part of the project 'Model-Based Support of Distributed Collaborative Design' (Richard Fikes, Edward Feigenbaum, Sheila McIlraith, Robert Engelmore Todd Neller, Liang Zhu. Yumi Iwasaki, James Rice) at Stanford, described here: http://www.ksl.stanford.edu/htw/

A.G. Cohn, S.M. Hazarika, Qualitative Spatial Representation and Reasoning: An Overview (2001), http://citeseer.ist.psu.edu/cohn01qualitative.html

Brandon Bennett, Modal Logics for Qualitative Spatial Reasoning (1996), http://citeseer.ist.psu.edu/bennett96modal.html

Jochen Renz, Bernhard Nebel, Efficient Methods for Qualitative Spatial Reasoning (2001), http://citeseer.ist.psu.edu/renz01efficient.html

[24] See URL http://ruccs.rutgers.edu/ftp/pub/papers/cognit89.pdf
[25] See URL http://www.cyc.com/cycdoc/vocab/spatial-vocab.html

3 Ontology for information gaps: questions to oneself or others

In parallel with the work reported in Section 2 an ontology is being developed for kinds of information gaps that such a robot may identify as indicating information it should seek (an ontology for questions it can ask itself or others).

This work is an investigation into some of the variety of types of questions that can elicit information or at least express a need for information, or specify a gap in information available to some information-user. This is not a question about communication between agents (e.g. asking questions) but about processing *within* an agent. An agent deciding what to do might discover that it needs some information, e.g. about the current environment, about whether preconditions for some action are satisfied, about what the consequences of the action could be, etc. Such information is, for now, restricted to an environment without other intelligent agents.

An agent (e.g. an animal or robot) may sometimes discover gaps in its information, which can lead it to take steps to fill the gaps, such as looking in a new direction, processing its current sensory information differently (e.g. attending to specific features and relations), going somewhere else, making a prediction, recalling something, or deriving new information from existing information. If there are other agents who might know the answers it can also ask questions, but not all questions are concerned with getting information from another individual.

Both the analysis of required types of factual information and the analysis of possible information gaps and what to do with them are huge topics, but the present discussion is constrained by the needs of the two scenarios being developed within CoSy, initially focusing on the PlayMate robot scenario.[26] This scenario is based on a robot able to perceive, manipulate and talk about 3-D objects on a table top, including eventually assembling complex objects from components and talking about the process with a collaborator, a teacher, or a pupil. For present purposes we shall focus only on the physical environment, leaving for another paper the extension of the question ontology to include questions about beliefs, percepts, thoughts, capabilities, and current mental state of another individual (or oneself).

Although our scenario ultimately has to include linguistic competence, it is likely that such competence will build on more fundamental and general forms of competence shared by pre-verbal human children and other animals that manipulate objects – such as New Caledonian crows.[27] So this document is a first draft incomplete investigation into some of the kinds of information gaps that a pre-linguistic robot with perceptual and manipulative skills may need to cope with, especially one which needs to know what it is doing and why.

3.1 Self-knowledge may be about gaps in knowledge

There is growing interest in self-aware computing systems. See for example [McCarthy, 1995], and the position statements for a DARPA workshop on self-aware systems held in 2004[28]

What is not always appreciated is how important it is for an intelligent system to discover gaps in the information available to it so that it can take steps to fill the gaps. Later, we'll see that some kinds of intelligence require the ability to cope not only with gaps in information about *the current*

[26] See URL http://www.cs.bham.ac.uk/research/projects/cosy/PlayMate-start.html

[27] Betty, the New Caledonian crow in the Oxford Zoology lab made headlines in 2002 when she demonstrated that she could make hooks out of wire in order to retrieve food in a basket from the bottom of a tube. See URL http://users.ox.ac.uk/.kgroup/tools/tools.main.shtml

[28] http://www.ihmc.us/users/phayes/DWSAS-statements.html

situation, but also gaps in information about *hypothetical situations* that could arise, if certain actions were performed. Identifying and dealing with those gaps will require more complex architectures than simply noticing that you don't know what is behind you. For now we ignore those complications and discuss only gaps relative to what is the case.

3.2 The need for an abstract syntax

Asking questions externally requires formulating sentences in a public language using syntactic constructs of that language to express both that the utterance is a question and also what the semantic content of the question is, namely what information is required. In contrast, for internal purposes a robot or animal needs only a form of representation that expresses the *content* of the question, since the fact that it is a question could be determined by the information-processing context in which the question is being formulated.

This requires a means of representing those information gaps, which in ordinary parlance amounts to a means of expressing questions. We show how to derive a wide variety of question forms from a language that is capable of expressing factual information. The simplest example would be taking whatever expresses some fact (e.g. the internal representation of 'The blue box is empty') and treating it as having an unknown truth-value instead of treating it as true. More complex types of questions of various types, require more or less complex syntactic transformations of a propositional expression, as illustrated in the remainder of this paper. The presentation here is informal, since we do not presuppose any particular well defined formalism for expressing propositions, though we assume at least the richness of first-order, and possibly second-order propositional calculus. How exactly that kind of syntax is expressed in the mind of an animal or a robot is a separate question. We also leave open the possibility that some animals and robots, and of course young children, will be able to cope with only a subset of the forms and transformations described here. We postpone for now the task of explaining how to derive questions from non-propositional formats for expressing information, for example maps or diagrams, even though in biological evolution and child development it is possible that those came first.

3.3 Driving idea

Concerning questions, the driving idea is that every question content (every specification of a need for factual information) can be expressed by starting with a true/false propositional form then creating zero or more gaps in that form by removing components and then applying various question-forming operators to the resulting gappy structure, the simplest such operator being 'Is it true that...', which requires no gaps. This point is independent of exactly how the logical forms are represented, as long as they are 'Fregean' representations, in the sense of [Sloman 1971][29]). There are many different syntactic forms with the same semantics that can be used in intelligent systems. The comments about question formation apply to all of them. In systems using hybrid representations, e.g. a mixture of logical forms and map-like structures (labelled 'analogical' in [Sloman 1971]), some generalisation of the method for deriving question forms from propositional forms would be required, but that is a topic for another occasion.

In general, several questions can be derived from any proposition, by creating gaps in different places, and by applying different question-forming operators to the gaps.

[29]See URL http://www.cs.bham.ac.uk/research/cogaff/sloman-analogical-1971.pdf, also available in HTML: http://www.cs.bham.ac.uk/research/cogaff/sloman-analogical-1971

Obviously, developing the ontology for questions (information-gaps) presupposes a prior ontology for the kinds of propositions from which the questions could be derived. So a substantial part of the ontology presented here is common to both factual propositions and factual questions.

What follows is not a totally general ontology. The survey is restricted to types of factual information that may be discovered, or needed, or used, by an animal or machine with a physical body, located in space, enduring over time, and perceiving things, preforming actions and thinking about objects, events, and processes (including actions) that might be relevant to an animal or robot manipulating and thinking or talking about 3-D objects on a table top, as in the PlayMate[30] scenario of the CoSy project.

The survey is concerned mainly with factual information and gaps in factual information, though some of the ideas might be applicable to other forms of information (e.g. control information, as expressed in imperatives, discussed briefly below).

3.3.1 Questions for thinkers as well as communicators

Some suggestions are made about possible forms of propositions and forms of derived questions that can arise both in dialogues and in thinking. I believe that the ideas are widely applicable in the design not only of dialogues between agents (as in PlayMate) but also agents that have dialogues with themselves, for instance when planning or reasoning, or wondering what to do, or why something happened. In the latter cases the forms of question and answer need not necessarily be translated, or even translatable, into external linguistic forms. Some of them could be used by pre-verbal children or other animals. Exactly which subsets can be used by which types of individuals is an empirical question which this document does not address, though it ends with some brief comments on how this relates to pre-verbal cognitive processes in a predominantly altricial species. The altricial-precocial spectrum in organisms, and its possible application to robots, is discussed in another paper.[31]

I don't claim that any of this is original: the idea about questions arising out of gaps is based on something dimly remembered from my youth as a philosophy student in Oxford 1957-1962, but I cannot be more specific as to where the ideas came from, except that it builds heavily on the logical investigations of Gottlob Frege[32] about a hundred years ago. The powerful idea that gaps in propositions can be filled in many ways, including the use of quantifiers was, I believe, independently discovered by G.Frege and C.S.Peirce, breaking away from the previously held (Aristotelian?) view that in a sentence like 'All As are Bs' the subject of the sentence is some entity referred to as 'All As'. On Frege's analysis the sentence is to be understood as asserting that any gap-filler that makes '... is an A' true will also make '... is a B' true. (There are people, e.g. the philosopher Fred Sommers, who believe that Aristotle was right and Frege wrong, but I shall ignore that debate.)

Logicians will recognize that much of what is said about gaps here can be expressed using Lambda Calculus, the formal notation Alonzo Church developed inspired by the rather more clumsy formalisms invented by Frege. I have decided to keep this presentation informal, instead of using lambda expressions or a similar notation.

Looking for related work on the internet, I found a recent PhD thesis which develops ideas about question types very close to the ideas about questions presented here:

[30] See URL http://www.cs.bham.ac.uk/research/projects/cosy/PlayMate-start.html
[31] See URL http://www.cs.bham.ac.uk/research/cogaff/altricial-precocial.pdf
[32] There are many other overviews of his work available online and in print, e.g.
http://plato.stanford.edu/entries/frege, http://www-groups.dcs.st-and.ac.uk/.history/Mathematicians/Frege.html

Debra Thomas Burhans (2002)
A Question Answering Interpretation of Resolution Refutation[33]
PhD Thesis, State University of New York at Buffalo Department of Computer Science and Engineering.

Nils Nilsson kindly read a draft of this document and commented that some of the early work in AI used ideas similar to those presented here. For instance work by Cordell Green[34] and others in the 1960s. It was pointed out for example that an existentially quantified proposition could be interpreted as a question whose answer would be a list of possible values for the variable in the quantifier, a relationship which in natural language would be clear between ‘Is it true that someone was at the party’ and ’Jill, John, Jonas and Judy were at the party’. Closely related ideas led to the design of Prolog and other logic programming languages.

3.3.2 Varieties of non-information-seeking types of questions

Before moving on to the task of specifying types of information relevant to our project I’ll make a few remarks about restrictions on the types of questions that are relevant at least to this document.

It is clear that questions can have many different pragmatic functions, apart from eliciting information, for example:

- testing someone’s knowledge
- getting someone to think about a problem (Socratic questions)
- teasing someone
- challenging someone
- being ironic or sarcastic
- drawing attention to something
- being rhetorical (in various ways)
- making a request (i.e. trying to get something done)
- asking for a definition (I constantly have to say to people: ‘what do you mean by “emotion”?’ in response to some comment or question about emotions.)

Several taxonomies of question types have been proposed, including the taxonomy often referred to as “Bloom’s Taxonomy”, which distinguishes different sorts of pedagogical functions of questions.[35] Ways of classifying questions in terms of their communicative or other social functions functions will be ignored here. We are concerned only with the ‘core’ case: use of questions to specify missing but possibly required factual information. I consider *only* the semantic content of questions and answers, ignoring all pragmatic and performative issues, and leaving open the syntax to be used in a working system. Any syntax used here is illustrative only.

[33] See URL http://www.cse.buffalo.edu/.burhans/debra.ps
[34] C. Green, Applications of Theorem Proving to Problem Solving, Proceedings 1969 International Joint Conference on Artificial Intelligence (IJCAI 69), pp. 219–240
[35] See http://honolulu.hawaii.edu/intranet/committees/FacDevCom/guidebk/teachtip/questype.htm and http://www.officeport.com/edu/bloomq.htm, and for a different type of taxonomy compare: http://www.usingenglish.com/glossary/question-types.html

3.4 Definitions:

- *A question* as discussed here, is a specification of a factual information gap in an information user.
- *An answer to a question* is a an information structure that fills, or reduces, the gap (or in some cases, specifies how the gap could be filled).

This leaves completely open what kind of medium is used to express either questions or answers, and what syntactic forms are used within any particular medium. It also leaves open what kinds of use are made of the information, how gaps are detected, how they are filled, how genetic information, learning or development can produce any of the competence required, how brains operate, etc.

3.4.1 Key ideas:

The key ideas are:

1. There is a basic type of factual question, a Yes/No question, which refers to a proposition that is capable of being true or false and requests the information whether the proposition is true or false,
2. There are ways of generating other types of factual question by creating one or more gaps in the proposition and specifying requests for information about ways of filling the gaps so as to make the proposition true (an example would be 'Which individuals satisfy: X is P?')
3. Many, though not necessarily all, factual information gaps in information users involve needs that can be expressed as factual questions of the above kinds. (Note that two kinds of gaps are involved: information gaps in intelligent systems and gaps in propositions produced by removing components of those propositions.)

E.g. if a proposition asserts 'A did X' then the question form 'Who did X?' is a natural language expression of the logical question form requesting information that can fill the gap in '... did X?' so as to produce a true proposition. Some questions are related to a single gap, others to two or more gaps, as illustrated below. Some questions refer to a gap on the assumption that there is a unique gap-filler that would make the proposition true ('Who is the murderer?') whereas others allow the answer to specify any number of fillers that could produce true propositions, or to specify the number without listing them, e.g. 'Who saw the event?' (answer: Tom, Trudy, Tim, ...) or 'How many people saw the event?' (answer: none, or five, or...).

Some questions specify two or more gaps and ask about individuals standing in some relationship (e.g 'Who is taller than whom?').

Since propositions can take many forms, with unbounded complexity, the variety of types of gaps and gap-fillers is essentially unbounded.

Sometimes the specification of information needed requires two gaps to be filled by the same thing. E.g. 'Who shot himself?' involves two linked gaps in the propositional form: ' ... shot ... '. Such linking is normally expressed in formal languages by the use of variables e.g. 'x shot x' or 'Shot(x, x)', In natural languages other devices are used e.g. use of pronouns, words like 'himself', 'itself', 'themselves', etc. (Failure to understand what's going on here leads many people to believe that we all have spooky entities called 'selves' inside us.) The natural language devices are irrelevant to the aims of this paper, which is concerned only with cognitive structures and processes that may be common to users of many different languages and to some intelligent systems that do not use what we would call an external language.

3.4.2 Non-factual questions

A more general definition of 'question' would include *control* information gaps, as expressed in questions like the following, whose answers could be imperatives rather than factual statements:

- What should I do?
- Is it better to do X or Y?

For now these are not considered, though some of what is said here would also be applicable to such cases. E.g. some of the ways in which different question forms can be generated from a single factual proposition by inserting gaps and applying appropriate operators to the proposition with gaps, would also be applicable to ways in which different control questions can be generated by inserting various kinds of gaps in an instruction, or imperative, of the form 'A, do X?' (e.g. talking to myself: 'Me, do X'). Questions like 'What should I do?' could be interpreted (at least sometimes) as requesting the gap to be filled in the self-directed imperative 'Me, do ...'. The difference is that instead of depending on the notion of a true proposition, control questions depend on the notion of an imperative or instruction being accepted or acceptable. There are many complications that arise from the fact that two imperatives that are independently acceptable can be inconsistent in the sense that they cannot both be acted on whereas two true propositions cannot be inconsistent. Another complication is that often a control question is not about what to do but which of two options is better[36], a topic with many complications that we shall ignore.

3.4.3 Varieties of answers

Here we discuss only answers that provide information to fill gaps specified in factual questions. However, as indicated above, questions can have many different functions, and so can answers. In dialogues between information users there are things that we would call answers that perform other functions besides giving answers, such as apologizing for not knowing the answer or being too busy to respond, or refusing to give the information, or challenging some aspect of the question – e.g. its relevance to some shared goal context, its assumptions (e.g. that there is a unique gap filler), its politeness, the right of the asker to hear the answer, the availability of information that provides an answer, etc. None of those 'pragmatic' answers will be discussed here. The scope of this document is very limited.

I believe that every question of the types surveyed below can be expressed in English, though some of the questions have very complex structures and therefore expressing them accurately in English may be difficult, or at least clumsy. E.g. consider questions derived from various ways of creating gaps in this proposition:

A1, B1 and C1 are in the same order on line L1 as A2, B2 and C2 are in on line L2.

(some English speakers would omit the second 'in'.) A possible question derived from that would be:

[36]See URL http://www.cs.bham.ac.uk/research/cogaff/sloman.better.html

Which three things in the room are on the same order on line L2 as A1, B1, C1 are in on which line?

Linguists may be interested in debating whether this really is English or not, and if not why not, whereas for my purposes that is irrelevant since, like many other non-English sentences, the meaning is clear, as in 'Me go home after me eat', and many of the things young children and non-native speakers say. This raises important questions about what is going on in a child who wishes to ask a question but does not yet have a way to express it that will be understood by others. However that is just a special case of the ability of children to have something to communicate without yet having the linguistic competence, a situation that will be familiar to every parent.

The answer to the above question may specify a complex set of alternative ways of producing true propositions by referring to objects in the room to fill the gaps formed by removing A2, B2, C2 and L1 from the original proposition.

Not only generating, but also understanding an accurate expression of a complex question may also be difficult. Often there is a formal mathematical way of expressing the question that mathematicians would find clearer or more succinct, though non-mathematicians may find it incomprehensible. One of the problems of learning to do mathematics is learning ever more sophisticated question-forming and answer-forming techniques, though I have no idea how many mathematics teachers understand this. It is very likely that some of the more complex forms of propositions and questions discussed here could never be understood by young children or other animals. Likewise some of them may be beyond the grasp of our robot. Nevertheless the theory subsumes them.

3.4.4 Questions and propositions in non-linguistic (pre-linguistic) information users

As far as I know every human language is capable of expressing all the forms of propositions and factual information-seeking questions mentioned here, but it is theoretically possible that some natural languages lack the syntactic expressiveness required for some forms of questions, just as young children do. Whether some natural language lack the expressive power to do everything described here is an empirical issue, that is not relevant here. The study of the precise forms of syntax used in a particular language to express one of these question types is part of empirical linguistics, and I have nothing to say about that. However, I conjecture that every information-seeking question that is expressible in any language is an example of the sort of schema for generating questions presented below, and every proposition that is expressible in any language is an example of one of the propositional forms we are discussing. (Though the actual list in this document is incomplete: the paper is unfinished.)

But there are more things to be thought and questions to be asked that we can express linguistically. For example, a person looking at a map to find a suitable route between two towns is engaged in acquiring information that is not necessarily expressible verbally. The answer may take the form of identifying a sequence of roads marked on the map, or if the terrain has no roads and the map shows contours (as in an orienteering map) the answer may be a trajectory that could be drawn on the map, or visualised while looking at the map. So some factual questions have contents and have answers that are expressed by non-linguistic means, e.g. using kinds of attention-focusing mechanisms ('virtual fingers' (Pylyshyn's FINSTs (1989))[37]) to specify entities or other things referred to for oneself (even if one is a robot). The work of Trehub(1991) is also relevant.

[37] See URL http://ruccs.rutgers.edu/ftp/pub/papers/cognit89.pdf

Public, external, languages used for communication between agents constitute only a subset of the forms of representation required by intelligent agents. Many kinds of information structures are used in various parts of a complex agent architecture, to store information of different kinds, referring to external or internal entities, states or processes, at various levels of abstraction, stored for different time scales, used for different purposes. There is no reason to believe that all such semantic contents are expressible in external languages, such as human languages.[38]

So there may be types of questions and types of answers that are not accurately expressible in any language but are important for the cognitive functioning of some non-linguistic animals and pre-linguistic children. It is possible that this kind of pre-linguistic mental function provides part of the infrastructure for natural language as used by humans and therefore needs to be understood in order to develop an accurate theory of language understanding by humans.

Pre-linguistic question formation and answer-seeking or answer-providing processes (within an individual) may also be important for some robots, in particular robots that are able to perceive and act intelligently while lacking the ability to communicate in any human-like language.

3.4.5 Types of question structures and answer structures

We turn now to a first draft informal elaboration of the points made above about ways of deriving information-seeking questions from gaps in propositions. This requires a specification of forms of propositions, on which more will be said in a later section. For now I am going to assume that all the propositions required are of a type that could be expressed in first or higher order predicate calculus, with certain modal operators added (e.g. to express causes, purposes, etc.) Instead of a general formalisation, I present only examples, which should suffice to explain the structure of the ontology of questions being presented here. Remember that arguments of predicates and functions need not be denoted by words or linguistic expression: they may be objects of current attention identified by their relationship to what is currently seen or thought about.

A question specifies a request for information or an information gap. The basic form of answer is "yes" or "no" answering a question that asks what the truth-value of a proposition is. More complex answers correspond to questions that are derived from a proposition in various ways, described below. A question has a structure, which may be more or less complex, as explained below.

Answers will be information structures (or structured information items) capable of filling such information gaps, or specifications for sets of information structures that can fill gaps. In many cases the answer to a question can be given in alternative forms: e.g. if the question requests an example of something, different examples may be given in answers. However, in general the form of a question constrains the forms of appropriate answers. For example, if the question is "Is it raining?", the answer "thirty seven" is inappropriate, and if the question is "How many people are in the room?" the answer "yes" is inappropriate. If the question is "What happened next?" only a complete proposition is an appropriate answer (though the proposition may be identified by a referring device, e.g. "What I just saw happened next").

Despite the use of English to make all these points, it is important to stress that this is not a document about natural language syntax, but about kinds of factual information that may be needed by an information processing system.

[38]For some examples, see http://www.cs.bham.ac.uk/research/cogaff/challenge.pdf

3.5 Question forms

This section describes, rather informally, and incompletely, a principled collection of transformations of propositions in order to specify information gaps. A question, from this standpoint is just an expression of an information gap, and an answer is an expression of the information that fills the gap.

3.5.1 Yes-no questions: Proposition and its negation

Those are the fundamental sorts of questions, merely asking about the truth or falsity of a proposition.

3.5.2 Derived questions: operations on propositions with gaps

There are many question forms that can be generated by producing one or more gaps in a proposition. The gaps can be of many kinds: object, attribute-type, attribute value, relation, action, purpose, manner, instrument, cause, location (different sorts), time (different sorts),

Some of these occur so often that we have special words to identify the type of gap to be filled:

> who, what, which, where, when, how, why, ...

Sometimes there is more than one gap

> 'Who broke what?'
> 'Who murdered whom when?'

Different question forms relate to whether a unique gap filler (or a unique tuple of fillers) is requested, or a the set of all, or the number of the set of all, or some statistical property of the set (are most X's P?), whether referents are identical, or whether an identified entity fills some gap, etc. etc. For example:

> 'Is fred the person who...,?'
> 'Did fred solve it in the same way as Mary?'
> 'How many people solved it in the same way as Mary?'
> 'How many things did Fred do in the same way as Mary?'
> 'Is Fred cleverer, bigger than Joe, ...'

It is worth noting that from a logical point of view many things are clearly interpretable whose most 'logical' expression in natural language would be regarded as ill-formed or bad style, such as

> 'How many people did fred solve it in the same way as?' 'Fred solved it in what as Mary?'
> Possible answers to the latter include: 'in the same way', 'in the same room', 'in the same time', etc.

From the present point of view the best way to express these questions in natural language, or even whether they can be expressed in a particular natural language, is of no consequence.

3.5.3 Some common question forms

- **WH-singular-unary:** Proposition with one gap, requests unique filler:

  Who married Mary?

- **WH-singular-binary:** Proposition with two gaps, requests unique filler-pair

  Who married whom? (also ternary, etc. 'Who gave what to whom?')

- **WH-plural-unary:** allows plural answer:

  Who was at the party?
  Answer: Fred, Mary, the man in the moon, and my uncle.

- **WH-plural-binary:**

  Who talked to whom at the party?
  Answer: Fred talked to Mary, Sue talked to Joe and to Tom, etc.

3.5.4 More complex derived forms

Note that for every type of proposition P more complex propositions can be formed by embedding P in a larger context, possibly including other propositions, Q, R, ..., and then creating one or more gaps in the context from which questions can be derived. Examples include:

- P because Q
  leading to "What caused P?", or various sorts of "Why" questions about P.
- Q because P
  leading to "What effect did P have?", and similar questions.
- P in order to Q
  where P describes an action that has been done, is being done, or will be done
  leading to "Why is P being done?" and similar questions. However this presupposes the existence of some sort of intelligent agent with intentions in the environment.
- Q in order to bring it about that P
  where P describes a state of affairs
  Possibly leading to something like "How was P done?" or related questions.

Other examples specify means by which something is done, route taken, style or manner, or other qualifiers of P.

For each question form and each situation determining an answer there may be varied verbal answer forms possible (e.g. different ways of identifying an object) with no context-independent means of saying which answer or answer form is best. Sometimes selecting between the different modes of expressing a request for the same information depends on facts about the dialogue context other than whatever is referred to in the question (e.g. the answerer may be impatient with the questioner, or may want to draw attention to a mistaken presupposition of the questioner, or ......)

3.5.5 Further ways of deriving questions from propositions after creating gaps

To be continued .... there is a lot of existing work on this, including work in linguistics, NLP, Burhans (2002) and others. Some sample ways of deriving questions by filling gaps

- How many ...?
- Are there more than N ...?
- Is the number of ... the same as the number ...?
- If P will Q ??
- Why P ? — possible answer: P because Q
- Why Q ? — possible answer: P caused Q
- Why P ? — possible answer: P in order to Q
- How P ? — possible answer: P using X
- etc.

The above set of points about forming questions by applying various kinds of operators to gaps in more or less complex propositional forms generates a VERY large variety of types of questions (And answers).

Since for any proposition different gaps and different numbers of gaps can be created in it, and for each gap or each combination of gaps different questions can be formed using different question-generating operators, the variety of forms of questions (or more precisely information requirement specifications) must be greater than the variety of types of propositions.

3.6 Categorical and hypothetical information gaps

A more detailed theory will need to specify features of cognitive architectures that allow an information-user to discover that it needs some information that it does not have. It will require mechanisms that allow the missing information to be identified with sufficient precision to generate actions to remedy the deficiency[39] which may involve

- re-focusing attention
- re-processing current sensory information in some new way
- accessing and manipulating previously acquired information
- performing some external action to get more information (e.g. looking in a different direction)
- communicating with another agent in some way, whether non-verbally or verbally (e.g. a child may say ‘What’s that?’, and turn an adult’s head to look in the right direction.)

One very obvious way in which an AI system can encounter an information gap occurs if it is attempting to form a plan to achieve some goal by combining action operators for which preconditions and consequences are known (as in STRIPS, and many AI planning systems since STRIPS).

[39] Jackie Chappell has demonstrated a video of Betty, the aforementioned New Caledonian crow, performing a task involving using a stick to pull or push food along a horizontal tube with a trap in the middle, into which the food may fall. At one point Betty seems to realise that she is not sure of the exact situation. So she releases the stick, goes to another position to get a side-on view of the tube, stick and food, then comes back and completes the task.

The system may discover that some potentially relevant action has a precondition, and then on attempting to decide whether the precondition is satisfied or not it may discover that it does not have the information, in which case it has to find out, perhaps by looking. (This is unlike a typical use of Prolog in which programs are written to work on the assumption that anything they cannot prove is false.) Another example would be a visual system attempting to identify some object or to perceive the relationship between two objects (e.g. whether they are touching), which then discovers that it cannot decide because some or all of what it is looking at is hidden by a large object. Having identified missing information an intelligent agent with sufficient meta-level knowledge may be able to take steps to fill the information gap. In more primitive animals or robots a purely reactive mechanism might generate information-acquiring actions without any self-knowledge about what is going on.

For a program which constructs a plan made of a sequence of actions, things get more complex. If a STRIPS-like forward-chaining planner has formed a partial plan, consisting of applications of operators, O1, O2, O3, and it then wishes to check whether operator O4 could continue the sequence, it should not check O4 for correctness in the *current* situation, but only in the *hypothetical* situation that would arise after performing the first three actions. The ability to ask a question about a hypothetical situation requires the ability to use expressions whose semantics do not refer to the world as it is. This requires architectures and mechanisms that probably do not exist in the vast majority of animals, and did not exist in machines until the development of AI.

Things get even more complex in a backward chaining planner. Such a planner may discover that the required goal state G would exist if it were to apply an operator Oa, because the effects of Oa include making G true. If Oa has no preconditions or all its preconditions are already true, then things are simple: required plan contains only the action Oa. However if Oa has preconditions, they may be false, and unlike the case of a forward chaining planner, there does not yet exist some specification of a hypothetical situation relative to which the planner can check the preconditions. Rather we have to ask whether there is some way building a sequence of actions starting in the current situation such that at the end of the sequence, the preconditions of Oa would be true.

So whereas previously there were two sorts of questions that arise when considering a possible action Oa, namely

- are the preconditions of Oa true now
- would the preconditions of Oa be true if the currently considered partial plan were executed

in backward chaining planners we have to ask

- is there a possible partial plan starting from the current situation such that if it were executed then the preconditions of Oa would be true

This is a more complex case because instead of simply checking Oa against reality, or against an existing data-structure, filling the information gap requires exploring many possible scenarios.

Thus we have at least three distinct kinds of question in a problem-solving system

- is P true now?
- would P be true in some already specified hypothetical context C?
- is it possible to find a way to specify a hypothetical context C, satisfying some constraints, such that P would be true in C?

Further complications arise if the preconditions of an operator include variables. For example a grasping operator may not specify which object is to be grasped, but it may specify that whatever the object is it should be clear for grasping. In that case where we have previously talked about checking whether a proposition is true (in reality or in some hypothetical situation), we would instead need to talk about a *propositional form* (e.g. 'x is graspable') and ask whether that form has instances that are true, or would be true in the hypothetical situation. Examples of a natural language expressions of such questions are

- Is anything that is graspable clear?
- Which graspable objects are clear?
- Would anything graspable be clear if I were to do ..... ?
- Which graspable objects would be clear if I were to do ..... ?

(The 'Which' questions, whose answers give a list of options tend to be more appropriate to breadth-first searching, the others to depth-first searching.)

As anyone familiar with AI planning and problem-solving programs will know, the need to identify and reason about information gaps in hypothetical situations can add considerable complexity to task of keeping track of all the actual and possible states of affairs that could arise if various actions were performed.

Humans vary in their ability to perform these tasks, and very young children seem to be quite incapable of processing information about a collection of hypothetical situations, so this ability is a product of learning or development in humans. It is also generally very limited, insofar as humans cannot match the combinatorial searching capabilities of even moderately powerful computers, though they may compensate for that in various ways, not discussed here. Moreover, some humans seem to have and use these abilities without being at all aware that that is what they are doing. It may be that some of the representations of information gaps are implicit in the states of activation of procedures and mechanisms for dealing with such gaps. It requires additional architectural sophistication to be able to represent such things explicitly and reason about them explicitly.

SOAR[40] is a well known example of a problem solver that can discover that in order to solve one problem it may have to solve another which includes discovering something. Many such mechanisms are theoretically possible and we shall not attempt a survey here.

We can expect different subsets of the theoretically possible variety of question forms and information-obtaining processes to occur in different species, in different robots, and perhaps in the same individual at different stages of development. Where linguistic forms are used, different languages will encode them in different ways. We can expect young children to be unable to cope with the linguistic forms required to ask questions relative to some hypothetical situation, and then only later to learn the extra syntactic constructs (subjunctive forms) required. Whether they can do it internally in thinking and reasoning about hypotheticals before they can do this using an external language is an interesting question.

Yet another variation on the theme discussed here is the ability to ask and answer questions about the mental states of other individuals, e.g. in predicting or explaining actions or trying to decide how to advise or explain something to them. AI research on this task has explored both explicit reasoning about mental states of others, for instance using doxastic/epistemic modal logics, and something closer to implicit reasoning using 'simulative mechanisms'.[41].

[40]See the SOAR homepage http://sitemaker.umich.edu/soar

[41]Barnden's ATT-META project (http://www.cs.bham.ac.uk/~jab/ATT-Meta/) has developed a tool using simulative reasoning mechanisms for dealing with metaphors, mental states of other agents (possibly nested) and which can also be used

The ideas expressed here are very general. When they are applied to the full variety of forms of propositions described in the companion paper on ontology for PlayMate, it may turn out that there are additional possibilities and problems not allowed for here.

3.7 Some References [to be extended]

These references were found with the help of Google, after the first draft of these notes was written. The list is neither complete nor authoritative nor representative. It's just what I found in a fairly short time spent searching. The Burhans thesis mentioned at the beginning seems to be closest in spirit to what I have been proposing regarding question formation, though I have not yet read it all.

Debra Thomas Burhans (2002)
A Question Answering Interpretation of Resolution Refutation[42]
PhD Thesis, State University of New York at Buffalo Department of Computer Science and Engineering.

List of question types: http://www.usingenglish.com/glossary/question-types.html

Metzler, D. and Croft, W.B., (2003) "Analysis of Statistical Question Classification for fact-based Questions" *Information Retrieval.* PDF[43]

Jean-Pierre Ko (2000)
Any questions left? Review of Ginzburg & Sag's *Interrogative Investigations.* J. Linguistics 40 (2004), 131-148. f 2004 Cambridge University Press DOI : 10.1017/S0022226703002354, UK. PDF[44]

Jonathan Ginzburg and Ivan Sag, (2000) *Interrogative Investigations,* CSLI Lecture Notes 123. Stanford, California: CSLI publications,, Pp. xii + 449.

for general counterfactual or hypothetical reasoning

[42]See URL http://www.cse.buffalo.edu/.burhans/debra.ps

[43]See URL http://ciir.cs.umass.edu/pubfiles/ir-323.pdf

[44]See URL http://lingo.stanford.edu/sag/revs/koenig-rev.pdf

4 Multi-layer perception and action sub-systems

In the report on architectures and in the CoSy proposal we refer to the need for architectures in which there are several (many?) kinds of concurrently active processing (mostly totally unconscious). Different types of processes will use different ontologies, because they process information referring to different things, and they will need different forms of representation because they are not all processing information in the same way. For instance some components may be involved in tracking and controlling quite fast moving physical processes whereas others are concerned with hypothetical reasoning about possible future actions or possible past events that might explain something perceived, and yet others may be reasoning about the perceptual, motivational, and reasoning processes in other intelligent systems (or in itself). So we can expect such an architecture to include

- Reactive processing (including innate and trained reflexes and more complex internal reactive processes, many of them representing information only implicitly in states of activation of subsystems)
- Deliberative processing (capable of detachment from what exists, using structures, relationships, compositional semantics, reference to hypothetical futures (predictions/plans), pasts (explanations), unseen things (what made that noise? what's behind me? etc.)
- Reflective-meta-management:
  - including self-observation such as noticing how what is visible changes as one moves left or right, or up or down, or as one moves nearer or further, or noticing that one can guess where the invisible part of the rim of the saucer is, or noticing that the reflection of the cup in the bowl of the spoon temporarily caused some confusion
  - needs meta-semantic capabilities (reference to things that process information, their states, actions, goals, etc.)
  - both inward and outward directed
  - inward directed meta-semantic processing requires architectural support

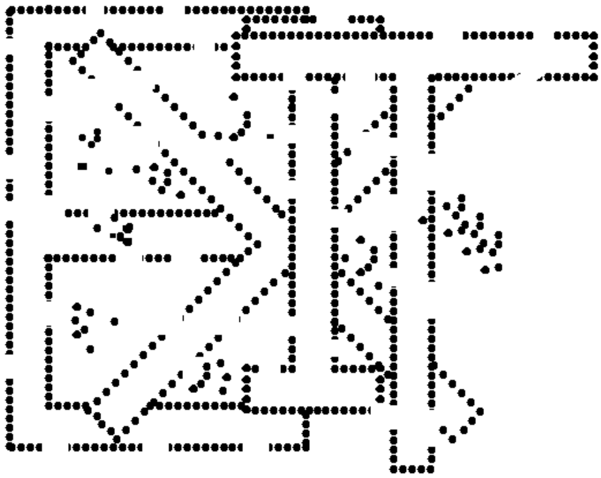
Detailed analysis of these architectural components and their interactions can lead to requirements for the forms of representation.[45] In CoSy deliverable DR.1.1 we discuss the requirement to support intelligent "barge-in", where some processes interrogate, interrupt, redirect or modulate others. This requires the processes that are acted on to use forms of representation that are not as opaque as many common implementations of perception, planning and learning programs are (e.g. values of local variables in complex procedure activation stacks).

Another requirement is that perception and action components in the architecture may have multiple layers processing information at different levels of abstraction.

- Processing intermediate perceptual layers explicitly supports 'hierarchical synthesis' (Neisser and others), and 'low commitment' intermediate inferences (Marr), structure-sharing between hypotheses.
- Intermediate and high level actions can be initiated in different parts of the architecture – e.g. innate and trained reflexes in the reactive subsystems, plan execution in the deliberative system, expressing self assessments in the meta-management system (anger, submissiveness, contriteness, ...)

[45] Many examples of such detailed analysis can be found in Minsky's draft book 'The emotion machine' on his web site.

In some work done around 1977 (summarised in Chapter 9 of [Sloman, 1978]), it was shown how multi-level analysis of pictures such as this, allowed rapid recognition of a word using mixtures of bottom up and top down processing:

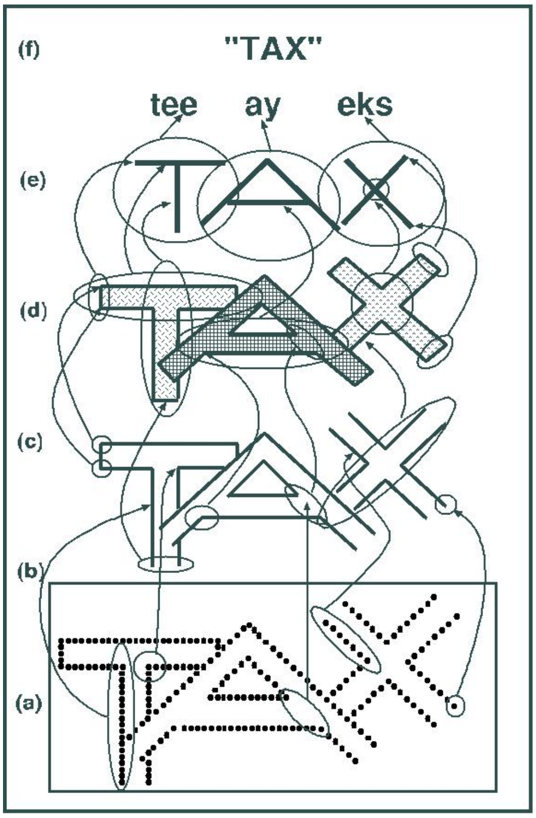


Success of that system depended crucially on finding a collection of intermediate layers of representation since the mapping between image structures and 'scene' structures (letters, words) was far too messy and complex to be used directly. The multi-layered processing system used is crudely depicted in the figure below, which also shows some of the important multi-strand relationships between components within and between the different levels of abstraction.

Conjecture: *Humans process different layers of interpretation in parallel.* This is obvious for language. What about vision?

Concurrently processing bottom-up and top-down helps constrain search. There are several ontologies involved, with different classes of structures, and mappings between them.

- At the lowest level the ontology may include dots, dot clusters, relations between dots, relations between clusters. All larger structures are *agglomerations* of simpler structures.
- Higher levels are more abstract – besides *grouping* (agglomeration) there is also *interpretation*, i.e. mapping to a new ontology.
- Concurrent perception at different levels can constrain search dramatically (POPEYE 1978) *(This could use a collection of neural nets.)*
- Reading text would involve even more layers of abstraction: mapping to morphology, syntax, semantics, world knowledge



These ideas of 'hierarchical synthesis' and 'analysis by synthesis' came out of cognitive psychology e.g. [Neisser, 1967] and played an important role in AI about 30 years ago but went out of fashion.

Is it time to revive them? If so what are the intermediate ontologies that should be used in the hierarchies and what forms of representation should be used? Our preliminary conclusion is that old ideas about representing multiple levels of static structure need to be enhanced by allowing for the need to represent multiple levels of multi-strand processes, which multi-strand relationships change, some continuously, some discretely. This could be done by means of concurrent simulations running at different levels of abstraction, but suitably linked to one another and to the sensory data (and in some cases also to motor outputs).

(Ref geons? Explain flaws in Marr's proposals for generalised cylinders. Need for representations relative to different frameworks: the image or optic array, the perceiver's body, an object-centered framework, various scene-centered frameworks – e.g. group of people around a table.)

Further examples of processes involved in different parts of complex visual systems can be found in these recent presentations by members of CoSy

> Henrik Christensen, 'Active Vision Systems', Tutorial presentation at EC Vision Summer School on Cognitive Vision, August 2005
> http://www.cas.kth.se/~hic/active-vision.pdf
>
> Ales Leonardis, 'Problems of Representation and Learning in Machine Vision', Tutorial presentation at IJCAI05 Tutorial *Representation and learning in robots and animals* Edinburgh, July 30th-31st 2005,
> http://www.cs.bham.ac.uk/research/projects/cosy/conferences/ijcai-booklet/tutorial-notes-leonardis.pdf

4.1 Ontologies and representations in concurrently active sub-systems

As explained in section 2.3 of the Workplan we conjecture that an intelligent human-like robot will have several concurrently active sub-systems – not merely subsystems performing different tasks such as visual processing, linguistic processing and control of walking, but also subsystems that operate in fundamentally different ways doing very similar tasks, e.g. low-level visual mechanisms and high-level visual mechanisms starting from the same visual data (the same samples of the optic array) but performing different visual tasks, for instance, controlling posture while walking, and providing information about whether an object ahead is small enough to step over. These different levels may use quite different kinds of representation – for instance only implicit representations in the reactive posture control mechanism.

There may also be a different roles for architectural layers concerned with producing and controlling behaviour 'online' and separate layer concerned with generating and monitoring exploratory and playful behaviour in the previous layers, as part of an important kind of learning, not available to precocial species. We are investigating requirements for a simplified two-level system where the main innate competence instead of being manifested immediately in behaviour is manifested internally in a 'meta-level' deliberative learning mechanism which drives the behavioural system in such a way as to test hypotheses and expand behaviours using various kinds of 'syntactic' composition mechanisms, allowing much faster and more flexible learning than neural nets and statistics-based learning mechanisms on their own. This seems to be what happens in altricial species that are born helpless but achieve sophisticated and varied competences as adults.

Testing these ideas in a working robot will be quite difficult. Some preliminary ideas for a test domain are described in http://www.cs.bham.ac.uk/˜axs/polyflaps/

(This may to lead to a new interdisciplinary project working in parallel with CoSy, and collaborating closely with it.)

5 Some general notes on representations

It became clear very early in the history of AI that finding good ways of representing various kinds of information was important, and that there were many options to be investigated including choices between symbolic, neural, self-organising, discrete, continuous, statistical, fuzzy, structural, and hybrid systems of various sorts.

E.g. although it is often thought that statistics-based and neural-net based forms of representation were discovered as a reaction to problems in symbolic AI, in fact they have been known for a long time. Minsky wrote over 40 years ago:

> "Incidentally, in spite of the space given here for their exposition, I am not convinced that such incremental or statistical learning schemes should play a central role in our models. They will certainly continue to appear as components of our programs but, I think, mainly by default. The more intelligent one is, the more often he should be able to learn from an experience something rather definite; e.g., to reject or accept a hypothesis, or to change a goal. (The obvious exception is that of a truly statistical environment in which averaging is inescapable. But the heart of problem solving is always, we think, the combinatorial part that gives rise to searches, and we should usually be able to regard the complexities caused by noise as mere annoyances, however irritating they may be.)" [Minsky, 1963] http://web.media.mit.edu/ minsky/papers/steps.html

In parallel with Minsky's eclecticism some researchers were arguing that AI should focus on logic-based forms of representation because of their generality, expressive power, well defined semantics and amenability to formal manipulation in a principled way.

For an example see the 1969 logicist manifesto of McCarthy and Hayes: "Some philosophical problems from the standpoint of AI" [McCarthy and Hayes, 1969],

One of the oldest critiques of this emphasis on Logic points out the role of visual, spatial, diagrammatic reasoning in human intelligence, including much mathematical thinking. The point was made a hundred years ago by Poincare, and often raised as an objection to the logicist standpoint, e.g. in [Sloman, 1971] which analysed the role of 'analogical' representations in reasoning and proofs, in contrast with Fregean representations, attempting to show that both can be useful and, in their different ways, rigorous.

Over the years there have been many attempts to survey various kinds of forms of representation, e.g.[Brachman and Levesque, 1985] [Glasgow et al., 1995] [Peterson, 1996], and the regular conferences on Knowledge Representation (See www.kr.org).

In what follows, I have tried to summarise some of the issues in a way that abstracts from all the details (e.g. about how various forms of representation work and why some are more useful than others for specific tasks, etc.) and presents some high level principles that can guide our thinking about representations.

5.1 The concept of representation

Many people use the word 'representation' as if it were clear that there is a boundary between what is and what is not a representation. Unfortunately for them, people disagree on what is a representation, and the word 'representation', like the word 'information', is used with great generality both in ordinary conversation and in technical literature, including control engineering, neuroscience and biology.

In the face of this diversity of usage it is a complete waste of time to attempt to define some boundary that separates representations from other states and structures involved in natural and artificial biological systems.

Instead we need to understand the *dimensions* in which such things can vary.

There is a huge amount of literature on this topic, but the vast majority grossly over-simplifies the problems by assuming that there are at most one or two dimensions of variation, and in some cases further over-simplifying by treating dimensions as dichotomies, e.g. classifying representations as symbolic/sub-symbolic, or as verbal/pictorial, or as discrete/continuous.

Another problem is that some ideas keep getting reinvented because the subject is not yet taught in such a way that people have a deep understanding of its history. E.g. the notion that spatial or map-like representations can be more useful for certain tasks than logical or verbal descriptions is constantly being reinvented.

Yet another peculiarity of the field is that whereas many people in AI and cognitive science focus on representations *within* an agent others regard representations as essentially *external* entities used by humans for communication or other purposes such as mnemonic aids, or calculations.

A more subtle issue is that not everyone notices that internal states and processes do not need to be *brain* states and processes, any more than data-structures in a running computer need to be electronic structures: in both cases there are different implementation levels involving implementation hierarchies where virtual machines are implemented in other virtual machines which are ultimately implemented in physical machines.

5.2 Some previous work and work in progress

- http://www.cs.bham.ac.uk/research/cogaff/Aaron.Sloman_towards.th.rep.pdf
  Towards a General Theory of Representations
  Talks about syntax, semantics and pragmatics of control states, and how they can vary. (written 1994, published 1996 – needs updating)
- http://www.cs.bham.ac.uk/research/cogaff/sloman-vis-affordances.pdf
  What the brain's mind tells the mind's eye
  (Unfinished, unpublished draft. Sections 3 and 4 are relevant, especially on varieties of implicit and explicit representation.)
- Talks on representations
  http://www.cs.bham.ac.uk/research/cogaff/talks/
  Talk 7: When is seeing (possibly in your mind's eye) better than deducing, for reasoning?
  Talk 30: Varieties of meaning
  Other things in http://www.cs.bham.ac.uk/research/cogaff/

5.3 Recommendation for CoSy

We should avoid taking sides on terminological and philosophical disputes about what is or is not a representation or whether representations are or are not necessary in intelligent systems.

Instead, if we use the notion of 'representation' as shorthand for a very general notion (as currently used in a huge amount of literature in science and engineering) something like

> 'a means of encoding, expressing, storing, or transmitting, information for current or future use'

(where information is not the mathematical Shannon/Weaver concept, but rather the more ordinary concept of semantic content, reference, meaning: i.e. information is about something, real or imaginary) then we can identify the following scientific and engineering assumptions on the basis of which quite specific research tasks can be defined:

- Robots, like organisms, need to do a variety of things with information, which can be about many different kinds of things including unperceived entities, the past, the future, spatially remote locations, specific entities or situations, generalisations, explanations, etc.
- We need to understand the variety of such tasks and what the implications are of performing them, the ways in which they can succeed, or fail, or performance can be improved. There will be many different trade-offs to be analysed.

5.3.1 Varieties of tasks

These tasks include (among others)

- controlling current behaviour using current sensory information
  where

  > *'behaviour'* refers to both external and internal actions,
  > including virtual machine actions
  > and
  > *'sensory'* refers to sensing of internal and external states and processes, including virtual machine states and processes

- saving current (internal or external) sensory information for future use
- using previously acquired information for some task (need to list variety of types of tasks)
- formulating goals, targets, or constraints
- that includes
  - expressing gaps in current information (i.e. questions that need to be answered. Betty seems to be able to do this when she goes to look at something from the side before returning to continue manipulating it.)
- monitoring behaviour and effects of behaviour and comparing with goals, targets, constraints, etc.
- (Note: this can be as simple as predictive control systems or as sophisticated as judging one's own behaviour on ethical grounds or on grounds of social adequacy.)
- deriving new information from old, including making predictions, forming generalisations, refuting hypotheses, identifying questions or problems, etc.
- transforming information to make it more useful in some way (e.g. switching between logical, spatial or neural forms or between declarative and procedural forms) etc.

- many kinds of learning
  - forming or extending ontologies One of the most important functions of science, often overlooked: http://www.cs.bham.ac.uk/research/cogaff/crp/chap2.html
  - developing new forms of representation (e.g. arabic numerals, cartesian coordinates, logics, grammars, programming languages, musical notations, are all relatively recent forms of representation invented by humans, but there are much older forms used in our brains and minds that we probably still don't know about.)
  - identifying faults in procedures or strategies and finding improvements.

5.3.2 Tradeoffs between varieties of forms of representation

Information can be expressed/encoded/stored in many different formats using many different mechanisms and media and the forms so far investigated in current science and engineering are probably only a subset of what is possible. We don't yet know how many different forms are used in biological systems (e.g. forms of representation of information during development of an egg into an animal, forms of representation in brains).

Where there are different forms of representation and mechanisms for manipulating we must not simply choose one or a subset: we need to understand the variety of options and the tradeoffs between them. (In general arguing that one sort is best is a waste of time.)

This includes the following sub-tasks:

- understanding the ways in which kinds of representations can differ
- e.g. finding dimensions of variation for forms of representation, extending the following list:
  - continuous vs discrete
  - forming a total ordering vs being partially ordered
  - being atomic (e.g. a number) vs being complex (e.g. a vector or tree, or graph, or map).
  - having fixed structure (e.g. N-dimensional vector) vs having variable structures, perhaps of unbounded complexity (e.g. parse trees)
  - being localised in some space (e.g. a node in a network) vs being distributed (e.g. distributed over many nodes in a network)
  - using physically separate items vs using physically superimposed or merged items E.g. radio and TV information is broadcast in patterns of electromagnetic radiation where different signals are superimposed and receivers require to do filtering, de-multiplexing, etc.
  - using chemical or other structures to encode information that apply the information by changing the structures and combining them with other structures vs using encodings that are not changed by their use, e.g. stored programs in a computer's memory, or the fixed weights in a neural net.

5.3.3 Varieties of criteria of assessment

We need to find detailed criteria for assessing/comparing different forms taking into account the fine structure of various tasks. E.g. in a vision system a form of representation for the same scene can be

- more or less easy to derive from sensory input

- more or less changeable as viewpoint changes
- more or less useful for planning long term actions
- more or less useful for controlling current actions
- more or less useful for
  - a beginner learning some task
  - an expert already fluent in performing tasks of that sort
- more or less useful for segmenting the scene into different physical objects
- more or less useful for predicting behaviour of objects in the scene
- more or less useful for forming or testing generalisations
- more or less useful for communication with others who can or cannot see the same scene, for different purposes.
- more or less useful for combining with information from other sensory modalities, e.g. touch, hearing, ...
- more or less useful for retaining in a ‘recent’ memory when looking at a different part of the environment.
- more or less useful for guiding
  - manipulation of an object by a hand or tool
  - approach to the object by the whole agent (e.g. a wall or door or table)
- .... etc.... ???? ....

6 Sources of meaning: symbol grounding and symbol tethering

6.1 Interdisciplinary inspiration

Our research has been inspired in part by analysis of types of development of competence in humans and other animals, in collaboration with a biologist who has been working very closely with the Birmingham group, Dr. Jackie Chappell. This collaboration led to an IJCAI05 paper [Sloman and Chappell, 2005b] and a UK Grand Challenge workshop presentation (abstract in [Sloman and Chappell, 2005a]).

This work has clarified options for sources of meaning (‘symbol grounding’ vs structure-based meaning combined with ‘symbol tethering’[46]), and some of the nature-nurture tradeoffs, indicating conditions under which most capabilities need to be innate (apart from minor calibration and adaptation) and conditions under which new capabilities need to be generated rapidly by the individual animal or robot. Biological examples provide existence proofs at both extremes (in precocial and altricial species).

The existence of precocial species born with sophisticated visual and action competences, such as deer that can walk, find the mother’s nipple, and run with the herd within hours of birth, demonstrates that sophisticated visual and other apparatus need not be learned, implying that the semantic content of the information structures is somehow determined by pre-existing structures and how they are used. This eliminates, at least for such precocial species, one of the supposed explanations of where meaning comes from (“There is really only one viable route from sense to symbols: from the ground up” [Harnad, 1990]).

How then can innate abstract structures determine meaning, as we claim happens for precocial species? Our answer[47] is inspired by ideas about models in meta-mathematics and theories in philosophy of science about scientific terms that could not possibly get their meaning by abstraction from perceptual processes, (e.g. ‘electron’, ‘quark’, ‘neutrino’, ‘gene’, ‘economic inflation’, ‘grammar’).

A (consistent) formal system determines a non-empty class of possible (Tarskian) models, which may be either abstract mathematical objects or collections of objects, properties, relations, events, processes etc. in the world. For a given formal system not every such collection can be a model, though it may have many different models. As (non-redundant) axioms are added the set of possible models decreases. So the meaning becomes increasingly specific. A formal system can determine precise meanings for its theoretical terms, by ruling out the vast majority of possible things in the universe as referents.

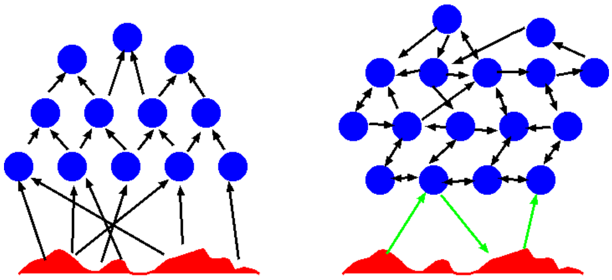
There is typically residual ambiguity, but, as [Carnap, 1947] and other philosophers have shown, this residual indeterminacy of meaning may be eliminated, or substantially reduced, by the use of bridging rules or ‘meaning postulates’. As [Strawson, 1959] noted, further ambiguity in reference to individuals can in some cases be eliminated or reduced by causal relations between the event of use of symbols and the things referred to.

We can call this ‘symbol tethering’[48], because it explains how a complex symbolic structure with rich structurally-determined meaning can be pinned down to a particular part of the world. This is very different from the theory of symbol grounding which requires all meaning to be derived bottom up by abstraction from sensory experience of instances, which is actually a very old philosophical

[46]We previously referred to the latter as ‘symbol attachment’. These ideas are explained further in http://www.cs.bham.ac.uk/research/cogaff/talks/#meaning

[47]Elaborated in http://www.cs.bham.ac.uk/research/cogaff/talks/#meanings, and also in [Sloman and Chappell, 2005b]

[48]This is a change of terminology from ‘symbol attachment’, suggested by Jackie Chappell.



There are many intermediate cases between pure symbol grounding and extreme symbol-tethering. Arrows on the left indicate abstraction of concepts from instances or sub-classes. On the right arrows indicate a mixture of constraints that help to determine possible interpretations (models) of the whole system and/or make predictions that enable the system to be used or tested.

theory (concept empiricism) put forward hundreds of years ago by empiricist philosophers and refuted by Kant around 1781. Our impression is that many of the AI researchers who now use the phrase 'symbol grounding' have ignored its original meaning (arising out of concept empiricism) and actually use it, unwittingly, to refer to something like what we have called 'symbol tethering'. We suggest that in altricial species, which start off far less competent than precocial species but after an early period of development tend to learn much faster and acquire skills requiring higher levels of cognitive sophistication, such as manipulative skills (e.g. crows, chimpanzees, humans)

- symbol tethering, in which structures produced by a combination of evolution, and exploratory 'play' and problem solving, acquire most of their meaning from their formal structure, with perceptual/action bridging rules to reduce their ambiguity, is far more important than what is normally called 'symbol grounding', and
- learning of chunks and combinatory procedures tailored to the current environment but constrained by powerful innate boot-strapping mechanisms that evolved to cope with general features of the environment (e.g. concerning space, time, motion, causation, etc.) is far more powerful than most forms of learning now being studied.

These, admittedly still tentative, ideas are being explored further in connection with learning in a specially designed domain, the 'polyflap' domain mentioned earlier, described in http://www.cs.bham.ac.uk/~axs/polyflaps. It is not yet clear whether this will turn out to be a feasible task within the timescale of CoSy. A possibly simpler domain involving objects pushed on a tabletop is being explored in PhD work by Marek Kopicki. See http://www.cs.bham.ac.uk/~msk/report3/

6.2 Biologically inspired 'altricial' learning systems

The rapid, automatic, non-need-driven collection/creation of a store of labelled, reusable, perceptual and action chunks, along with 'syntactic' mechanisms for combinatorial extensions of those 'basic' chunks, provides a rich and extendable store of rapidly deployable cognitive resources, using mechanisms that will already sound familiar to many AI researchers who have worked on planning, problem solving, reasoning, 3-D vision, and language understanding, though we have said very little about how they might work.

It may be that if some of this happens while the brain is still growing the result can be 'compiled' into hardware structures that support further development in powerful ways, explaining why the process needs to start while the infant is still physically under-developed. (This is very vague.)

We are not saying that precocial species cannot learn and adapt but that the amount of variation of which they are capable is more restricted and the processes are much slower (e.g. adjusting weights in a neural net trained by reinforcement learning, as contrasted with constructing new re-usable components for combinatorially diverse perceptual and action mechanisms).

So we suggest that in addition to the experiments where we pre-design working systems with combinations of different sorts of competence in order to understand ways in which such competences might work and cooperate, we should also begin the longer term exploration of architectures and mechanisms for a robot towards the altricial end of the spectrum described above, growing its architecture to fit opportunities, constraints and demands provided by the environment.

7 Meanings of ‘ontology’ and some history

(These notes arose out of email discussions in the CoSy project during the first half of 2005).

The multiplicity of the meanings of the word ‘ontology’ can generate confusion. There’s no *right* meaning. This section presents a ‘potted’ history (very much over-simplified) to draw out a few main types of use of the word ‘ontology’, showing (crudely) their conceptual and historical relationship. I end with the question whether CoSy needs ontology tools.

7.1 Meanings of the word ‘ontology’

A: The oldest sense of the word, going back to Aristotle and perhaps earlier, as indicated by the suffix ‘ology’, is as a name for a subject of study or expertise or discussion (Greek: ‘logos’ = ‘word’, ‘meaning’, ‘thought’....). In that sense ontology is a branch of philosophy close to metaphysics (in the same way as these are names of subjects of study or areas of expertise: zoology, biology, geology, philology, oncology etc.)

In that sense ‘ontology’ refers to an investigation of the nature of being, what exists, why there is anything rather than nothing, constraints on possible worlds (e.g. could causation exist without space and time?), and perhaps whether what exists could be improved on or is the best of all possible worlds, etc. etc. Our project will not need to spend much time on doing ontology in that old sense.

Note that in that sense ‘ontology’ is not a count noun (e.g. it has no plural and you can’t easily refer to ‘an ontology’ or ‘different ontologies’).

However, like other ‘ologies’ the word developed different uses including uses as a count noun. E.g. ‘geology’ can be a name for an area of investigation, but can also refer to features of a part of the world, e.g. the Alps and the Himalayas can have different geologies. The word ‘ecology’ has also developed that way quite recently (damaging the ecology of X isn’t damaging the science but damaging the features of X described by ecology the science).

B: Recent philosophical usage of the word ‘ontology’ changed, to refer to specific conceptual frameworks. (I think this was a 20th century change, but I may be misinformed.)

This change was partly inspired by Strawson’s 1959 book *Individuals: an essay in descriptive metaphysics*, and also by the discovery by anthropologists and others (developmental psychologists?) that different people have different views about what exists or can exist e.g. some do and some don’t include souls of dead ancestors, tree spirits, transfinite ordinals, uncountable infinities, quarks, etc.). I.e. they have different ontologies. Strawson distinguished

- *Descriptive metaphysics*: the task of expounding the conceptual/metaphysical framework *actually* in use by some community or by various communities (we could add, or various individuals at different stages of development, or various species)

  from

- *Revisionary metaphysics*: the task of arguing about which is the *correct* framework, usually including the claim that we’ve got it wrong so far.

The products of descriptive metaphysics are offered as accounts of ontologies actually in use, not as accounts of how things have to be.

In this new usage (sense B) the word ‘ontology’ became a count noun referring (roughly) to the most general conceptual assumptions underlying a specific system of beliefs.

In that sense there could be *different* ontologies, e.g. the ontology of the ancient Greeks, the ontology of Buddhists, a modern scientific ontology that includes genes, ecosystems, quarks, economic inflation, etc., an intuitionist or a platonist mathematical ontology, the capitalist economic ontology, etc.

This sort of ontology can be a complex structure including several sub-ontologies (as our current scientific ontology does).

This second sense (B) is the one that I have mostly been using recently in talking about the kind of ontology required for CoSy (as described in my previous message.)

C: Software-engineering uses of the word ‘ontology’ are much more recent, as a result of people in software engineering and AI coming to realise the following:

(a) Engineers designing complex systems need to think clearly about what sorts of things they are designing and what sorts of things their machines have to interact with, prevent, produce, maintain, etc., so that they need to think about and if possible make explicit the ontology they use as designers – the *design ontology*. This became increasingly important as engineering moved beyond systems that can be characterised by sets of differential equations and the like.

(b) Insofar as these complex systems process information, designers need to specify *what* kinds of information a system can be expected to process, which semantic contents it may have, how the information will be represented, how it will be manipulated and used, etc. So they started using ‘ontology’ to refer to the ontology used by their products, the *application ontology*

Because developing either kind of ontology (design or application ontology) can be a difficult and in some cases very complex process, it was soon realised that designers need tools, techniques, and formalisms to specify *what* kinds of information a system can be expected to process, which semantic contents it may have, how the information will be represented, how it will be manipulated etc. Some of the tools might also be used by the systems produced, e.g. ones that don’t have a fixed ontology, but go on extending their ontology.

(Outside AI this was connected with the growth of object-oriented programming languages and so-called object-oriented design – both of which generated much confusion as well as useful programming techniques).

So, engineers started moving away from informal descriptions and started producing formally specified ontologies for themselves to use (i.e. design ontologies) and for their machines to use (i.e. application ontologies), where the two could, of course, overlap in some cases.

D: Ontologies as formal structures

Ontologies of both the above two types are often expressed informally using familiar human forms of representation such as natural language, diagrams of various sorts, tables, etc. But as part of the drive towards theoretical clarity and elegance, and the need for more automated tools for producing working systems there was also a move towards requiring ontologies, even when produced by humans, to use a specific formalism with well defined conventions.

So there is now a sub-community of researchers whose only experience of the word ‘ontology’ is in the context of that kind of science and engineering, so for them the word usually refers to a formal structure specifying an ontology. (E.g. they may say ‘My ontology uses XML and requires a megabyte of filestore'.)

In parallel with all those developments, some people started producing more or less formal ontologies specific to their field of study or research, e.g. an ontology to describe the various stages, techniques, formalisms, products, processes, etc. involved in doing software engineering, or the ontology of biology. This might be called a *discipline meta-ontology*. Such meta-ontologies can be very important for teaching a discipline, or for preventing reinvention of wheels among researchers, etc.

That contrasted with producing ontologies for specific application domains, e.g. banking, weather forecasting, process control, etc.

7.2 Form vs content of some part of the world

Some ontologies are referred to as 'meta-models', because people sharing such an ontology (a very general set of concepts) can use it to produce different competing models of the same domain: the ontology merely specifies types of things that are conceptually possible, leaving open alternative theories specifying what actually exists and happens. This distinction (which is sometimes described as a distinction between the form of the world and its content) is very important in science, and often ignored in philosophy of science.[49]

It may be that one way to view the kind of ontological development that occurs within an individual of an altricial species learning about the environment through play and exploration is that it starts with a genetically determined meta-ontology which is used by mechanisms for generating hypothetical ontologies that can be tested by using them in playful exploration and experimentation. Of course some of those ontologies will be mistaken and that is one of the reasons why this process of learning and development needs to happen under the watchful gaze of more experienced individuals.

7.3 Ontology Tools

Because the tasks can be very difficult, especially for people not well trained in philosophy, software engineers and AI researchers have started developing (and re-inventing) more and more sophisticated methodologies and tools to aid the process of ontology development especially collaborative development of domain ontologies.

However, most of them (as far as I know) tend to assume a uniform representation (e.g. some kind of logic) and, as pointed out in our email discussions, that may be too restrictive for a project like CoSy. If we don't find suitably general tools that already exist we may need to develop our own tools for producing, maintaining and using ontologies suited to multi-level multi-functional robot architectures. In part this should also be a by-product of designing new kinds of learning mechanisms that develop or extend ontologies.

7.4 Further information

A lot more information relevant to this discussion can be found by giving search engines phrases like

ontology "software engineering"

or

ontology engineering

[49]This point is developed at greater length in Chapter 2 of *The computer revolution in philosophy* http://www.cs.bham.ac.uk/research/cogaff/crp/chap2.html

8 Future Work

As far as CoSy is concerned, by analysing some of the more ambitious scenarios in great detail we can work out more detailed requirements for kinds of information content that will be required, analyse the ways in which the information will be used in various sub-systems within the whole architecture, and from there proceed to investigate options and tradeoffs between different solutions. It is important to be aware that different forms of representation will be needed for different purposes and especially in different parts of the architecture. In particular, we may have to invent forms of representation that are more powerful than any currently in use.

A conjecture arising from our analysis (especially analysis of the need to represent multi-strand processes in which there are continuous and discrete changes in multi-strand relationships) is that a particularly important advance will come from development of perceptual and reasoning mechanisms that use representations of changing spatial structures of various kinds (including structures with multi-strand relationships), within *simulations of multi-strand processes in the environment*, at different levels of abstraction, running concurrently in registration with one another and with sensory input (and in some cases with motor output). Some of the processes need to be continuous (with different resolution levels) and some discrete. It is possible that new rich planning formalisms, such as those being developed in Work-package 6, will have a useful interpretation in this context if multi-agent plans with conditionals and loops can be interpreted as programs for running discrete high level simulations. All of this can be seen as a generalisation of a notion proposed by the late Max Clowes that "perception is controlled hallucination", which we now expand as "multiple, concurrent, synchronised hallucination".

This is not an entirely new idea: it builds on ideas also found, for example in [Craik, 1943] and more recently in [Berthoz, 2000] and [Grush, 2004], among many others. We are not aware of any earlier presentation of the "multiple-concurrent simulation hypothesis", however, which links together many past problems and proposed solutions.

We conjecture that if such mechanisms can be made to work, they will add powerful new capabilities for use in perceiving, reasoning about, and acting in a richly structured environment, including both reasoning about the actions of others and planning or analysing the robot's own actions. A simulation mechanism with different levels of abstraction is potentially a powerful basis for discovering affordances, which are concerned with 'what would happen if'.

Implementations of such ideas will be very difficult, and will ultimately stress both neuroscience and AI. Such implementations, if they are ever produced, can be tested by showing how, as more and more features of the design are added, more and more complex scenarios become demonstrable, in a partially ordered collection of scenarios, ordered by various kinds of difficulty[50]. There are also bound to be empirical conjectures for biology, neuroscience, psychology and linguistics that come out of such theories, and which might be testable empirically in those fields.

[50] The proposed methodology based on a partially ordered network of scenarios is described in the notes for the IJCAI05 tutorial http://www.cs.bham.ac.uk/research/projects/cosy/papers#tr0503

9 References

What follows is a small sample of relevant literature on this topic.

References

[Anderson et al., 2001] Anderson, M., Meyer, B., and Olivier, P., editors (2001). *Diagrammatic Representation and Reasoning*. Springer-Verlag, Berlin.

[Barkowsky et al., 1996] Barkowsky, T., Röhrig, R., and Freksa, C. (1996). Operationalizing Diagrammaticity. In *ECAI-96 Workshop on Representations and processes between vision and natural language*,, Budapest.

[Barrow and Tenenbaum, 1978] Barrow, H. and Tenenbaum, J. (1978). Recovering intrinsic scene characteristics from images. In Hanson, A. and Riseman, E., editors, *Computer Vision Systems*. Academic Press, New York.

[Barwise and Etchemendy, 1998] Barwise, J. and Etchemendy, J. (1998). A Computational Architecture for Heterogeneous Reasoning. In Gilboa, I., editor, *Theoretical Aspects of Rationality and Knowledge*,, pages 1–27. Morgan Kaufmann,, San Francisco, CA.

[Berendt et al., 1998] Berendt, B., Barkowsky, T., Freksa, C., and Kelter, S. (1998). Spatial representation with aspect maps. In Freksa, C., Habel, C., and Wender, K. F., editors, *Spatial cognition – An interdisciplinary approach to representing and processing spatial knowledge.* Berlin: Springer.

[Berthoz, 2000] Berthoz, A. (2000). *The Brain's sense of movement*. Perspectives in Cognitive Science. Harvard University Press, London, UK.

[Brachman and Levesque, 1985] Brachman, R. and Levesque, H., editors (1985). *Readings in knowledge representation*. Morgan Kaufmann, Los Altos, California.

[Brooks, 1991] Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47:139–159.

[Carnap, 1947] Carnap, R. (1947). *Meaning and necessity: a study in semantics and modal logic*. Chicago University Press, Chicago.

[Chomsky, 1965] Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press,, Cambridge, Mass.

[Christensen, 2005] Christensen, H. (2005). Active Vision Systems, Tutorial presentation at EC Vision Summer School on Cognitive Vision.

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

[Glasgow et al., 1995] Glasgow, J., Narayanan, H., and Chandrasekaran, B., editors (1995). *Diagrammatic Reasoning: Computational and Cognitive Perspectives*. MIT Press, Cambridge, Massachusetts.

[Grush, 2004] Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27:377–442.

[Harnad, 1990] Harnad, S. (1990). The Symbol Grounding Problem. *Physica D*, 42:335–346.

[Hayes, 1985] Hayes, P. (1985). The second naive physics manifesto. pages 1–36. Ablex, Norwood, NJ. Also in [Brachman and Levesque, 1985], pp. 468–485.

[Leonardis, 2005] Leonardis, A. (2005). Problems of Representation and Learning in Machine Vision, Tutorial presentation at IJCAI05 Tutorial *Representation and learning in robots and animals* Edinburgh, July 30th-31st 2005,.

[McCarthy, 1995] McCarthy, J. (1995). Making robots conscious of their mental states. In *AAAI Spring Symposium on Representing Mental States and Mechanisms*. Accessible via http://www-formal.stanford.edu/jmc/consciousness.html.

[McCarthy and Hayes, 1969] McCarthy, J. and Hayes, P. (1969). Some philosophical problems from the standpoint of AI. In Meltzer, B. and Michie, D., editors, *Machine Intelligence 4*. Edinburgh University Press, Edinburgh. (Accessible as http://www-formal.stanford.edu/jmc/mcchay69/mcchay69.html).

[Minsky, 1963] Minsky, M. L. (1963). Steps towards artificial intelligence. In Feigenbaum, E. and Feldman, J., editors, *Computers and Thought*, pages 406–450. McGraw-Hill, New York.

[Minsky, 1978] Minsky, M. L. (1978). A framework for representing knowledge. In Winston, P. H., editor, *The psychology of computer vision*, pages 211–277. McGraw-Hill, New York.

[Minsky, 1987] Minsky, M. L. (1987). *The Society of Mind*. William Heinemann Ltd., London.

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

[Peterson, 1996] Peterson, D., editor (1996). *Forms of representation: an interdisciplinary theme for cognitive science*. Intellect Books, Exeter, U.K.

[Sloman, 1971] Sloman, A. (1971). Interactions between philosophy and AI: The role of intuition and non-logical reasoning in intelligence. In *Proc 2nd IJCAI*, London. Reprinted in *Artificial Intelligence*, vol 2, 3-4, pp 209-225, 1971, and in J.M. Nicholas, ed. *Images, Perception, and Knowledge*. Dordrecht-Holland: Reidel. 1977.

[Sloman, 1978] Sloman, A. (1978). *The Computer Revolution in Philosophy*. Harvester Press (and Humanities Press), Hassocks, Sussex. Online at http://www.cs.bham.ac.uk/research/cogaff/crp.

[Sloman, 1992] Sloman, A. (1992). Prolegomena to a theory of communication and affect. In Ortony, A., Slack, J., and Stock, O., editors, *Communication from an Artificial Intelligence Perspective: Theoretical and Applied Issues*, pages 229–260. Springer, Heidelberg, Germany.

[Sloman, 1998] Sloman, A. (1998). Diagrams in the mind. In *in Proceedings TwD98 (Thinking with Diagrams: Is there a Science of Diagrams?)*, pages 1–9, Aberystwyth.

[Sloman and Chappell, 2005a] Sloman, A. and Chappell, J. (2005a). Altricial self-organising information-processing systems. *AISB Quarterly*, (121):5–7. Available at http://www.cs.bham.ac.uk/research/cogaff/05.html#200503.

[Sloman and Chappell, 2005b] Sloman, A. and Chappell, J. (2005b). The Altricial-Precocial Spectrum for Robots. In *Proceedings IJCAI'05*, page In press, Edinburgh. Available at http://www.cs.bham.ac.uk/research/cogaff/altricial-precocial.pdf.

[Sloman and Chrisley, 2005] Sloman, A. and Chrisley, R. L. (2005). More things than are dreamt of in your biology: Information-processing in biologically-inspired robots. *Cognitive Systems Research*, 6(2):145–174.

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

[Trehub, 1991] Trehub, A. (1991). *The Cognitive Brain*. MIT Press, Cambridge, MA.

[Ziemke, 2002] Ziemke, T. (2002). Special issue on Situated and Embodied Cognition (Ed. Ziemke). *Cognitive Systems Research*, 3. 3.

9.1 Reference documents

Related documents, publications...

Scientific papers:

These two papers are relevant to both this deliverable and to DR.1.1 Preliminary report on requirements for architectures and tools

[Sloman and Chappell, 2005b] http://www.cs.bham.ac.uk/research/cogaff/05.html#200502
[Sloman and Chappell, 2005a] http://www.cs.bham.ac.uk/research/cogaff/05.html#200503

IJCAI'05 Tutorial on *Representation and learning in robots and animals* Edinburgh, July 30th-31st 2005, organised by A.Sloman and B.Schiele. A.Leonardis was one of the tutorial presenters. The tutorial web site will remain online including the tutorial booklet, which will be expanded.

http://www.cs.bham.ac.uk/research/projects/cosy/conferences/edinburgh-05.html
http://www.cs.bham.ac.uk/research/projects/cosy/conferences/ijcai-booklet