

Discussion Paper:

## **Predicting Affordance Changes**

(Steps towards knowledge-based visual servoing)

**Aaron Sloman**

(Original title: Perceiving movements involving predictable affordance changes)

### **Fragments of a possible scenario for the CoSy Playmate robot In the final year of the project? 2004-2008**

## **CONTENTS**

- [Background information](#)
- [Hypotheses about how to make progress in machine vision](#)
- [Ontologies for intelligent visual systems](#)
- [Some examples of non-numerical height comparison](#)
- [Video demonstrations of challenges](#)
- [Semi-metrical relationships can be derived from topological relationships.](#)
- [Servoing vs 'Sense-decide-act' cycle](#)
- [Three linked sub-hypotheses](#)
- [How do actions change affordances?](#)
- [Movements that change action affordances](#)
- [Movements that change knowledge affordances \(epistemic affordances\)](#)
- [Seeing structure and understanding affordance changes](#)
- [How should the predicting be done?](#)
- [How to deal with noise and imprecision](#)
- [The relative rarity of hard cases: phase boundaries](#)
- [Examples: Changes that reduce uncertainty](#)
- [Changing spatial relations to make a prediction problem easier](#)
- [The importance of meta-management](#)
- [Snapshots of various possible motion scenarios: Predicting consequences of motion \(i.e. changing affordances, without dynamics\)](#)
- [Visual servoing](#)
- [Does reasoning about grasping have to be probabilistic?](#)
- [Note on nest-building birds](#)

- [How is the reasoning done?](#)
- [Other connections](#)
- [Implications for natural language interactions](#)
- [How much of the above could a robot learn?](#)
- [Related documents](#)

## Background information

### Note:

This document was originally written in 2007, during the EU-funded "CoSy" Cognitive Robotics project (summarised [here](#)). This was a proposal for re-designing the Robot's visual systems (especially for manipulation tasks, though similar arguments could be applied to the navigational tasks).

But the proposal was too remote from the available technology and the interests of most collaborators, and the proposal could not be pursued, especially as the project had only one more year to run.

Since then the ideas here have continued to develop, but not as part of any project, partly as a result of work on related topics presented in various discussions on this web site. I believe the ideas are still relevant to understanding natural visual systems and understanding why artificial visual systems remain far inferior in their versatility and robustness, even though highly trained specialized applications may be impressive within the scope of the benchmarks against which they are trained and evaluated.

Many of the ideas presented here arose in conversations with Jeremy Wyatt, before and during the CoSy project. This included the idea of a generalised aspect graph.

### NOTE added 18 Oct 2015

A separate document recently added to this web site explores relationships between evolution of abilities to perceive and reason about affordances (interpreted more widely than in Gibson's work) and evolution of mathematical abilities leading up to the discoveries reported in Euclid's *Elements*. Those abilities are discussed in relation to cognitive/perceptual abilities involved in perceiving differences between pictures of possible scenes and pictures of impossible scenes (e.g. by Reutersvard, Penrose and Escher, among others). The ability to detect that a picture depicts an impossible object is clearly related to the ability to detect that some action under consideration or some change in the environment is impossible, where that is not *empirical* impossibility, but *geometric* or *topological* impossibility (both of which are varieties of *mathematical* impossibility). See:

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/impossible.html>

---

This file is available in html and pdf formats

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/changing-affordances.html>

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/changing-affordances.pdf>

It is still liable to change, so please save links to the file rather than copies, which are likely to become out of date.

---

**WARNING: This file is still under development and liable to change.**

**This file was installed:** 18 Nov 2007

**THIS VERSION LAST UPDATED:**

20 Jan 2017 Added table of contents

3 Aug 2016 Some reformatting. Added PDF version; 4 Aug 2016: Minor edits.

## CONTENTS

- [Hypotheses about how to make progress in machine vision](#)
- [Ontologies for intelligent visual systems](#)
- [Some examples of non-numerical height comparison](#)
- [Video demonstrations of challenges](#)
- [Semi-metrical relationships can be derived from topological relationships.](#)
- [Servoing vs 'Sense-decide-act' cycle](#)
- [Three linked sub-hypotheses](#)
- [How do actions change affordances?](#)
- [Movements that change action affordances](#)
- [Movements that change knowledge affordances \(epistemic affordances\)](#)
- [Seeing structure and understanding affordance changes](#)
- [How should the predicting be done?](#)
- [How to deal with noise and imprecision](#)
- [The relative rarity of hard cases: phase boundaries](#)
- [Examples: Changes that reduce uncertainty](#)
- [Changing spatial relations to make a prediction problem easier](#)
- [The importance of meta-management](#)
- [Snapshots of various possible motion scenarios: Predicting consequences of motion \(i.e. changing affordances, without dynamics\)](#)
- [Visual servoing](#)
- [Does reasoning about grasping have to be probabilistic?](#)
- [Note on nest-building birds](#)
- [How is the reasoning done?](#)
- [Other connections](#)
- [Implications for natural language interactions](#)
- [How much of the above could a robot learn?](#)
- [Related documents](#)

---

## Hypotheses about how to make progress in machine vision and study of natural vision

The PlayMate robot, illustrated in [this video](#) prepared for the CoSy project review in November 2007, was very unreliable, and frequently failed to achieve its goal, like many robots under development then and now, though since then there are increasingly many robots that display high levels of manipulative skill as a result of a great deal of training. I think it is still the case that such robots do not know what they have or have not done, why they did not do something, or what would have happened if they had done something they did not do. Other limitations are discussed below.

Some failures of robot actions are due to inaccuracy in the production of movements specified by motor control signals. Since the CoSy robot used the Katana robot arm which provides very precise control, that was not the source of the PlayMate's problems. The difficulties arose to a large extent from the content and quality of the information the robot obtained about the current state of the environment, and its dependence on such information.

If the visual system were able to provide exact 3-D locations of every point of every surface of objects in the scene, including the robot's own hand, then in principle the planning and motor control subsystems could produce plans, and motor signals based on those plans, that enabled the robot to achieve its goals far more often (apart from problems such as objects slipping in the gripper when lifted, which are not a serious problem in the current scenario). However, we'll later see that even **perfect** metrical information about the current scene may not be the most useful basis for intelligent action planning and control.

This is related to the reasons why understanding a proof in Euclidean geometry or topology, based on understanding a diagram or collection of diagrams, does not depend on having perfect metrical information about the diagram, as illustrated in

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/triangle-sum.html>

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/trisect.html>

The poor quality of information available for planning and motor control has three main aspects:

1. Inadequacy of the visual sub-systems (which typically fail to find all the important detail in images, and fail to provide accurate 3-D information on the basis of current stereo algorithms),
2. Inadequacy of the forms of representation used, which do not allow important qualitative structures and qualitative relationships implied by the sensory information to be represented.
3. Inadequacy of information-processing architecture used, which does not allow the system to detect and deal with problems in what it knows.

The last two features would not be remedied by removing the first flaw. Having a perfect metrical model of the environment would still leave most of the problems of intelligent decision making and control.

There are two very different ways to try to improve the performance.

- Make all measurements and predictions more accurate and reliable, including using sophisticated probabilistic reasoning mechanisms to derive predictions and plans from uncertain information.
- Make the robot more self-aware and self-correcting: i.e. give it the ability to notice when its performance is likely to be degraded by uncertainty or by lack of information, and give it the ability to use the information it already has to work out how to get missing information or reduce uncertainty.

(Note: This kind of uncertainty is simply a matter of not having complete information. It need not involve any *measure* of degree of certainty or uncertainty. If I cannot tell whether a distant gap is too narrow for me to pass through I need not have a *measure* of my uncertainty, though I may know that I am less certain about a distant gap than about a nearby gap which I can more easily compare with my body width.)

A vast amount of work has been done (and will continue to be done) on the first method of increasing reliability, e.g. by using better cameras, better lighting, laser range finders, projected light stripes, better image processing algorithms, and other techniques.

The hypothesis explored here is that the second method is also worth exploring, and in the long run could turn out to be more important for explaining and replicating animal intelligence. That would include use of "cognitive visual servoing" or "knowledge-based visual servoing". This contrasts with servoing in which all the information used is numerical (including derivatives, etc.)

([NOTE: 29 Oct 2015] Since the above was written, work by Jeremy Wyatt and colleagues have made significant steps in that direction, but using methods that continue to depend on metrical details and probabilistic reasoning, unlike the methods suggested below. It is possible that the two approaches could be combined: a topic for another discussion.)

For a tutorial introduction to visual servo control see [Hutchinson \*et al.\*](#) The methods described in the tutorial depend on use of measurements in images and scenes, and their derivatives. According to Hutchinson *et al.* "... the design of stable, robust, image-based servoing systems ... has not been fully explored". Perhaps that's because researchers have focused on attempts to maximise precision and minimise risk of error: a pair of conflicting requirements. Things may have changed since that was written.

## Ontologies for intelligent visual systems

The methods I am suggesting (admittedly somewhat imprecisely) are mainly concerned with static and changing [topological](#) and [partial ordering](#) relationships (e.g. touching, overlapping, containing, getting nearer, between, more visible, partly obscuring, projecting to, being approximately aligned, becoming more constrained or less constrained, etc.).

At present I suspect there are no readily available good, useful algorithms for deriving such information from sensor and effector (e.g. proprioceptive) information. It may also turn out to be the case that getting such information from rectangular grids of numerical measures poses difficulties that are avoided by the physical design of biological visual sensors, which have a totally different structures from frame grabbers. the methods proposed would do better with cameras whose 'retinas' are not rectangular arrays but more like biological retinas, with resolution varying symmetrically around a fovea along with precise control of motion of the fovea when locked on a moving image structure. But there's a great deal of work to be done before any definite claims can be made.

The rest of this paper does not address the problem of improving camera design or the form of information produced by cameras.

Instead we focus on the requirement to use new kinds of information, both about the environment and about the robot's current information processing, to reason about whether and how to change what it is doing. That will require development of

1. new forms of representation to express the new information (some of it illustrated by the video in the next section)
2. new architectures to allow robots to have appropriate self knowledge about what they are doing.

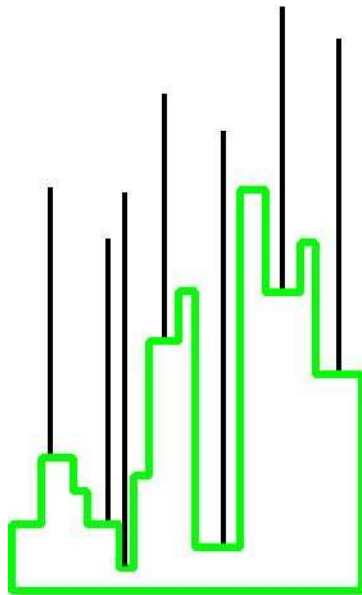
This hypothesis is expanded by drawing attention to the importance of the robot's being able to predict affordance changes that could be produced by its own actions, including predicting changes in *action affordances* and predicting changes in *epistemic affordances*.

This requires the ability to think and reason about sets of possibilities at a high level of abstraction, and to find new useful ways of **chunking** sets of possibilities: an important form of learning that seems not to have had enough attention. (See Sloman ([1996](#)))

The hypothesis is based on informal observation of some of the things humans and other animals can do. If we succeed in implementing these ideas in a working system, that could be at least a demonstration of the feasibility of the mechanisms. It may also suggest new experiments that could be done using children and other animals, including investigation of how cognitive visual servoing abilities develop. It is possible that related research could help us understand some intelligent behaviours in some other animals.

---

## Some examples of non-numerical height comparison (Added 5 Sep 2014)



It is easy to tell which of the horizontal portions of the image is lowest, and which is highest, without specifying a length for any of them. How? It is not so easy, but still possible, to determine which of the vertical black lines is longest and which shortest. How?

In the above figure you can find the highest horizontal green portion by moving your gaze from the top of the image down to find a horizontal fragment, then scanning to left and right to see whether any other green portion crosses your scanning line. A similar procedure makes it easy to find the lowest horizontal green portion. Neither requires any measurement or estimated measurement of height on any particular scale (inches, centimeters, etc.).

It is not so easy to tell which is the tallest black line and which the shortest. That's because there isn't a common base line in the image. However, if they were physical rods and the green surface was a solid object there are several possible physical manipulations that would make the comparison easy, e.g. moving all the rods into the hole with the lowest bottom, and then simply comparing the top ends of the rods, to find the one that projects above all the others, and the one that all the others extend beyond.

I suspect that in the first few years of life a typical human child discovers hundreds of such means of answering questions about metrical relationships without using any measurements, only topological relationships (e.g. touches, overlaps, extends beyond, etc.). This learning is done unconsciously, and the results are used unconsciously.

I suspect also that they develop ways of replacing actual *physical* motions to support comparisons (e.g. comparisons of length) with *imagined* or *visualised* motions, such as visualising the top of the left-most black line (line 1) moving to the top of the fourth black line from left (line 4), and seeing whether that would bring the bottom of line one to a location on line 4 or a location projecting below the bottom.

---

## Video demonstrations of challenges (Updated March 2014)

The original version of this paper made use of some video recordings made in 2005, that don't seem to be viewable using currently available tools. So I have prepared two new videos, demonstrating some of the points made above, using a mug and a pen held in various positions in relation to the mug, and asking questions about what can be seen regarding possibilities for motion.

Both videos have unscripted verbal commentaries.

The first video (.webm, about 30 secs),

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/pen-mug/pen-mug-intro.webm>

introduces the second video (.webm, about 6 mins 30secs):

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/pen-mug/pen-mug-demo.webm>

This shows a mug, a pen and a hand holding the pen and moving it in various positions and orientations relative to the mug, so as to change the affordances: e.g. some positions restrict some vertical motions and some positions restrict some horizontal motions, e.g. left-right horizontal movements, or front-back horizontal movements, or rotational movements.

You can easily do experiments yourself, holding a pen near, above, inside, a mug and moving it in various ways (including translations and rotations). Consider what predictions you can make about how further movements will or will not be constrained if you continue a particular movement. E.g. will the pen make contact with a part of the mug that will constrain further movement? Will continued motion bring the end of the pen into the mug so that further movement sideways and down is constrained by the mug. If motion of the pen is already constrained, what movements would alter the relationships so as to remove the constraint? Consider also what predictions you can make about what information will and will not be available to you.

**A task for a vision system is to be able see a movement and to predict that IF that movement continues THEN the relationships between the pen and the mug will (or will not) change in specific ways so as to restrict further movements.**

Some of the changes in relationships will be topological, some metrical, some continuous (getting closer), some discrete (coming into contact with, entering the volume directly above an object), some will be changes in action-affordances or epistemic-affordances, between something being possible and it being impossible, with a small, medium, or large region of uncertainty in which the phase change occurs.

None of this implies or requires use of probabilities or conditional probabilities. The most important changes are between what is and is not possible, or what will necessarily be the case (e.g. without a change of direction, the motion of the pen *must* produce a collision with the wall of the mug).

In some situations there are partial orderings of probability/likelihood but these are pre-mathematical "common sense" notions, not the concepts used in the theory of probability. In this sense, in most circumstances, the more your speed on a motorway exceeds the speed-limit the more likely you are to have a crash. In exceptional circumstances the opposite may be true, e.g. if someone is attempting to crash into you from behind in a vehicle whose maximum speed is lower than yours.

## **Semi-metrical relationships can be derived from topological relationships.**

### **Added 3 Sep 2014 (Still being revised)**

Some particularly important points probably not made clearly enough in the above videos are related to abilities

- to form intentions that involve metrical relationships,
- to devise plans for achieving those intentions, and then
- to execute those plans,

but without ever requiring perception of metrical values.

That's because visual comparisons of length, thickness, distance apart, curvature, and other features can, in many cases, be made on the basis of perceived topological relationships (e.g. containment, or overlap), without the use of any measurement operation producing an absolute value.

So animals (and future machines) with such capabilities will be able to have and to achieve intentions to find, or make something that is longer or shorter or the same length, thicker or thinner or the same thickness, straighter or more curved than or curved as much as something else.

More complex intentions can then make use of the results of those comparisons, e.g. making something using the found objects -- such as an archway with two sides the same height, or an item of clothing that fits the wearer well.

1. Perception of topological relationships in a scene, and topological changes that occur during a process, do not depend on abilities to acquire numerical (scalar) measures of location, size, distance apart, angles, curvature, etc., and then derive the non-numerical relationships by comparing numerical values, as is often assumed.

That's unnecessary because the ability to detect discontinuities in contents in static scenes, and discontinuities during perception of motion do not require use of such measures. The temporal order in which certain changes are sensed can provide comparative information about spatial measures. So, for at least a subset of cases, it may suffice to be able to detect certain discontinuities in a visual array, and to notice occurrences of something increasing or decreasing that occur in sensor values as eyes move across a scene (or hands in the case of haptic perception), such as something becoming visible, or ceasing to be visible, or a feature (e.g. reflection, or highlight) moving from one side of another feature, e.g. a crack or mark on a surface)

2. Perception of relative size, relative distance, relative curvature, instead of being derived from (possibly fuzzy or noisy) metrical sensor values, can be based on static topological relations, or



the temporal order in which sensing events occur.

For example, if moving the centre of attention linearly across an image causes the centre to rest on item A, then item B, then item C, where A, B and C are static objects, then it can be reliably inferred that A is further from C than B is. Likewise C is further from A than B is.

The videos were recorded using a Logitech Webcam B500, and the 'cheese' program on linux.

## Servoing vs 'Sense-decide-act' cycle

Some AI researchers assume (and some critics believe all AI researchers assume) that intelligent systems have to make use of a repeated three stage sequence of processes

1. acquire information about the environment via sensors (including checking predictions of effects of actions last performed);
2. process information and decide what to do next and make predictions about the consequences of doing it;
3. perform the selected action.

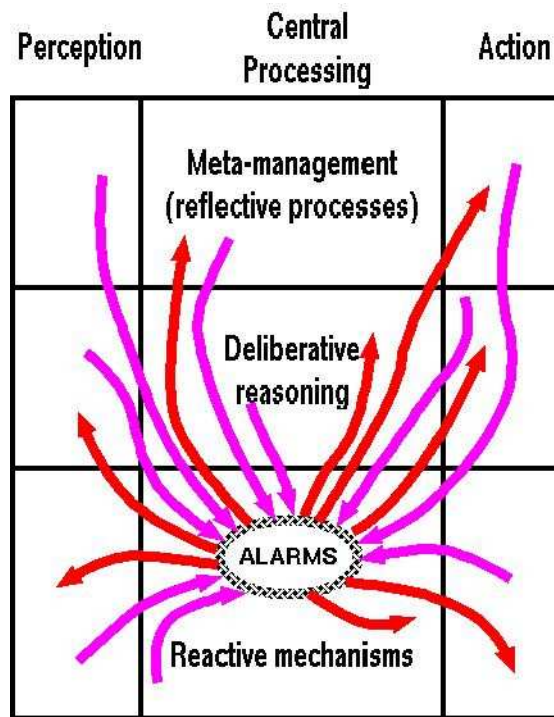
However, many control systems in which sending control signals and sensing are performed *continuously* are incompatible with this model. Such control systems require that the components of the alleged sequence are actually performed *in parallel*, for instance as you move your hand to pick up a small object while watching the movement, or paint a thin strip of a surface.

In other cases the monitoring continues until the action has advanced to the point where it can be completed ballistically (e.g. grasping a mug) and visual attention then moves to the next location at which action is needed, e.g. the jug to be grasped by the other hand, which requires the hand to be moved towards the jug using continual adjustments to the approach until it is close enough to complete the action without monitoring.

It is obvious that humans and many animals do not fit the sense-decide-act model in their everyday life, and instead do many things, including sensing, deciding and acting concurrently.

Since the early 1990s, the Birmingham CogAff project members have argued that an architecture is needed in which at least 9 different types of process are performed concurrently -- though without making the control-engineer's assumption that those processes are all continuous and of a type for which differential equations form a good representation.

Some may be continuous and some not, including possibly 'alarm' mechanisms that monitor mechanisms of other sorts and have the ability to freeze, modulate, redirect, or abort processes of all sorts.



The CogAff Architecture Schema allows for interactions between many different sorts of concurrently active processes, some continuous some discrete, including fast-acting 'alarm' mechanisms triggered by trainable pattern recognition processes.

In particular the notion of servo control, which normally assumes continuous (analogue) information processing can be generalised to include visual servoing which includes *discrete* processes of high level perception, goal-generation, goal processing, planning, decision making, self-monitoring, learning, and initiating new actions along with *continuous* control of movements and sensing of actions and environmental changes.

This paper assumes that such an architecture is available, and outlines a hypothesis that some of the information processing that could be useful for a robot (or animal) manipulating objects in the environment uses visual servoing and other kinds of servoing, partly on the basis of predicting changes in at least two kinds of affordances.

### Three linked sub-hypotheses

The hypothesis can be subdivided into the following:

1. The robot's reliability in performing manipulative tasks can be increased substantially by giving it the following new "cognitive servoing" competences (and probably others of the same general kind, still to be specified):
  - The ability to detect what it is and is not sure about
    - whether it is sure about properties and relations perceived in the scene
    - whether it is sure about predictions it makes about effects of its actions.
  - The ability to detect that performing certain actions will provide missing information, e.g. moving a block to one side will allow the full width of another block to be seen, or moving the camera to one side will allow the full width to be seen.

- The ability to move out of regions of uncertainty when it is on a "phase boundary" between being sure that something is true and being sure that it is false, for example, boundaries between:
    - being sure that it can estimate the size of something, sure that it cannot;
    - being sure that its hand is currently moving in the right direction to achieve a sub-goal, or sure that it is moving in the wrong direction;
    - being sure that an object is narrow enough in a certain dimension for it to be graspable by the robot, or sure that it is not narrow enough.
  - The ability to use 2-D projections of scene structures to reason qualitatively about which way to move in order to move away from a phase boundary into a region of certainty -- i.e. the ability to "reason with imagined diagrams" in order to solve problems related to planning and controlling actions. (Examples are given below.)
  - The ability to use all of the above as part of the process of "visual servoing" so as to detect and correct slight mis-alignments or mis-locations of the hand while moving in order to perform some task.
2. This requires the robot to be able not only to predict physical and geometrical changes that will result from its actions but also to predict and reason about something more abstract: *changes in affordances*.

In particular we distinguish the ability to predict

- changes in *action affordances*
  - changes in *epistemic affordances*
3. These new competences will require the "meta-management" capabilities of the robot to be extended.

I.e. it will need to have additional internal self-observation capabilities in order to detect states in which it lacks information or is uncertain about the information it has, and it will need the ability to use the results of such self monitoring in order to control subsequent planning, decision making, and actions. There are several presentations on varieties of architectures, using the CogAff schema as a framework for comparing alternatives and presenting H-CogAff as a conjectured architectural schema suitable for human-like minds, available here <http://www.cs.bham.ac.uk/research/projects/cogaff/talks/>

I shall later try to provide a summary presentation focusing on issues relevant to this discussion paper.

**The rest of this document elaborates on and illustrates the above.**

The document has been changing frequently since work began on it in mid November 2007, and it is likely to continue to change and develop. Comments, criticisms and suggestions welcome.

---

## How do actions change affordances?

Many people studying affordances have noticed that they are related to actions (or more generally to processes) that can produce some physical change in the environment. What is not so often discussed is that there are many changes in the environment that change the affordances in the environment. It is also not always noticed that whereas the main focus of investigations of affordances has been on what physical actions can be performed there are also important issues concerned with what might be called "epistemic affordances" or "cognitive affordances" i.e. affordances for an animal or robot concerned with information that is or is not available to that individual, which might potentially be useful in some context (e.g. perceiving, planning, controlling action, designing something new, predicting,

understanding the action of another agent, etc.)

Some people have discussed this, e.g.

<http://research.cs.vt.edu/usability/projects/uaf%20and%20tools/affordance.htm>

Physical and cognitive affordances help users perform physical and cognitive actions, respectively. We agree with Norman that these two kinds of affordance are not the same. They are essentially orthogonal concepts, but we think they both play very important roles. The reason for our giving them new names is to provide a better match to the kinds of actions they help users make during their cycle of interaction. A physical affordance is a design feature that helps, aids, supports, or facilitates physically doing something, and a cognitive affordance is a design feature that helps, aids, supports, or facilitates thinking and/or knowing about something.

Author not specified, but probably H.R. Hartson, who also wrote [Hartson \(2003\)](#).

So, physical actions or processes can change not only the available *action affordances*, they can also change *epistemic affordances* -- e.g. what can be perceived, felt, heard, etc. allowing the individual to obtain new information or, in the case of negative affordances, obstructing access to information.

So both action affordances and epistemic affordances can be changed when something moves in the environment and that means that the possibilities for those movements are related to possibilities for adding, removing or modifying action and epistemic affordances.

We can refer to the affordances to produce or modify affordances as "meta-affordances". This paper introduces examples and discusses ways in which meta-affordances can be used in predicting how actions or other events will change affordances.

A particularly important class of actions that can affect epistemic affordances is the set of changes of view point or view direction, but there are many others, including moving an object to make something more visible. Besides epistemic affordances related to vision, there are others related to other sensory modalities, but not much will be said about that here.

I believe that this discussion is closely connected to other CoSy discussion papers concerned with the need for exosomatic, amodal ontologies and limitations of the use of sensorimotor contingencies as a means of representation, but that will have to be discussed in another paper. For more on that topic see <http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0601>  
COSY-DP-0601 (HTML file): Orthogonal Recombinable Competences Acquired by Altricial Species

<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0603>

COSY-DP-0603 (HTML): Sensorimotor vs objective contingencies

This is a first draft discussion of some of the ways in which the PlayMate scenario might be extended to include acquisition and use of meta-affordances, concerned especially with predicting affordance changes.

There are some proposals for using these ideas for dealing with uncertainty by identifying "phase boundaries" between regions of certainty regarding affordances, and keeping away from those phase boundaries to avoid uncertainty.

## **Movements that change action affordances**

Consider holding a pen in the vicinity of a mug resting on a table with nothing else nearby on the table. Depending on where the pen is, what its orientation is, and how you are holding it, there will be different possibilities for motion of the pen, with different consequences. There will also be different possibilities for obtaining information about some or all of the pen, or the mug, or about the relationship between them.

For example, if you are holding the pen horizontally above the mug, centred on the mug's vertical axis then, if you try moving the pen down, the motion will be limited by the rim of the mug. However there are several actions that will make it possible to move the pen to a lower level including these:

- moving the pen horizontally in various directions until no part of the pen is above the mug, after which it will be possible to move the pen down to the table.
- rotating the pen until its axis is vertical, after which it will be possible to move the pen vertically downwards until it hits the bottom of the mug.

Those are examples where an action (horizontal movement, or rotation about a horizontal axis) produces a new state in which changed affordances allow additional actions (downward vertical movement).

Other movements will restrict the actions possible. E.g. if the pen is pushed horizontally through the handle of a mug and the mug is fixed, that will restrict possibilities for movement of the pen in any direction perpendicular to its long axis.

## **Movements that change knowledge affordances (epistemic affordances)**

There are also changes that will alter the information-gaining affordances. For example, if the pen is oriented vertically and only the portion projecting above the mug is visible there are many questions you will not be able to answer on the basis of what you can see, e.g.

- How long is the pen?
- Is the invisible part of the pen inside the mug?
- If the pen is moved horizontally to left or right, or moved horizontally further away from you, will its movement be obstructed by part of the mug?

Such unavailable information can often be made available either by moving something in the scene or by changing the viewpoint.

For example, lifting the pen vertically can change the situation so that the first question can be answered. The second and third questions could be answered either by moving to look down from a position above the mug or by moving the viewing position sideways horizontally and viewing the mug and pen from some other positions.

The problem discussed in this paper is: what are the ways in which by performing an action an agent can change not just the physical configurations that exist in the environment, but also the affordances that are available to the agent, including both *action* affordances and *epistemic* affordances (i.e. affordances for gaining information).

## **Seeing structure and understanding affordance changes**

Humans (though probably not infants or very young children), and also, I suspect, some other animals, are able to perceive scene structure in such a way as to support reasoning about how to change things so as to alter affordances. This competence includes

- Seeing what movements are possible in a given situation.
- Predicting how a particular movement will lead to affordance changes e.g. adding or removing constraints on further motion, or adding or removing constraints on what information can be acquired from the environment.
- Planning a movement or combination of movements in order to achieve some desired end state.

The pictures below illustrate some of the constraint changes that can be predicted.

If all this is correct, then one of the previously unnoticed (?) functions of a vision system is to be able, when seeing a movement of an object the vicinity of another object to predict that IF that movement continues THEN the relationships between the two objects will (or will not) change in specific ways so as to restrict or allow further movements (seeing changing action affordances), or so as to restrict or allow further information acquisition (seeing changing epistemic affordances).

Similar reasoning should be applicable to reasoning about consequences of *possible* motions as opposed to actual motions. This is relevant to both [the CoSy PlayMate scenario](#) and [the CoSy Explorer scenario](#).

[See also the KR'1996 paper "[Actual possibilities](#)"]

## How should the predicting be done?

Current AI systems, if they can do such things at all, will probably either use some sort of logical formalism to represent states of affairs and actions, and will perform the tasks by manipulating those representations, e.g. as a planner or theorem prover does, or use some probabilistic mechanism such as forward propagation in a neural net or some sort of Markov model.

Either way, states will be represented by a logical or algebraic structure, such as a predicate applied to a set of arguments, or a vector of values, and predictions will involve constructing or modifying such structures.

The abilities described and illustrated below seem to involve the use of a different sort of mechanism: one that makes use of 'analogical' representations in the sense defined in ([Sloman 1971](#)), discussed as an example of the use of an internal GL (Generalised language) in [this presentation on evolution and development of language](#).

This ability to reason about how affordances change as a consequence of changing locations, orientations, and relationships of objects also provides illustrations of the notion of Kantian causal competence, contrasted with Human causal competence in presentations by Chappell and Sloman [here](#).

The important point about such reasoning, apart from the fact that it is visual reasoning that uses analogical representations, is that the reasoning is geometric, topological and deterministic, in contrast with mechanisms that are logical or algebraic and probabilistic.

## How to deal with noise and imprecision

Detecting whether motion restrictions are present or not, or whether a continued motion will produce new restrictions or remove old ones, can be done with considerable confidence in VERY MANY cases even when images/videos are noisy and when accurate metrical information cannot be extracted from them.

That is because the nature of such restrictions, e.g.

*A prevents the motion of B from continuing*

does not depend on *precise* metrical relationships between objects their surfaces and their trajectories. Instead, much coarser-grained relationships, using relatively abstract spatial information, especially *topological* information and *ordering* information (e.g. A is between B and C), suffices for most configurations.

For example, if the point of a pen is within the convex hull of an upward facing mug then the material of the mug will eventually constrain horizontal and downward motion if the pen moves, but not upwards motion.

The word 'eventually' is used in order to contrast predicting *exactly* how much the object can be moved before contact occurs with predicting *that* contact will occur e.g. before the pen point has reached a target location outside the mug. I.e. the prediction is that a boolean change will occur (some relationship between objects will change from holding to not holding), but not exactly where or when it will change. That prediction does not involve high precision, but is sufficient to indicate the need to lift the pen before moving it horizontally far beyond the width of the mug.

If the mug is lying on its side, and the pen is horizontal with the point in the mug, then the mug constrains vertical movements and some, but not all, horizontal movements. For example, a horizontal movement bringing the pen out of the mug is not constrained, whereas a horizontal movement in the opposite direction into the mug will eventually be constrained -- when the pen hits the bottom of the mug. (The bottom surface is vertical because the mug is lying on its side.)

A robot that understands its environment needs to be able to perceive such constraints and use both in planning future actions and in controlling current actions: e.g. ensuring that the movement will bring about a desired change in constraints by adjusting the direction of motion or the orientation of one of the objects.

## **Requirements for precision can vary**

In very many cases there is no need for very precise control (e.g. below a few cm., or within a few degrees). The actual precision required depends on the task: predicting whether a ball thrown towards a bin at the far end of the room will go into the bin requires far more precision than predicting whether letting go of the ball when it is held close to a mug will cause it to enter the mug.

## **The relative rarity of hard cases: phase boundaries**

Predicting some of the changing affordances that will result from continuation of a perceived movement is very often quite easy because they depend only on topological relationships or very crude metrical relationships.

The exceptions occur when objects are close to 'phase transitions' e.g. close to the boundary of a convex hull of a complex object, or close to a plane through a surface or edge. In those special cases it is often hard to make binary classifications that are easy in the vast majority of cases. But it is usually easy to make a small movement that will turn a hard problem into an easy one.

## Examples: Changes that reduce uncertainty

This is now illustrated with some examples. The diagram represents various possible configurations involving a pencil and a mug on the side, along with possible translations or rotations of the pencil indicated by arrows.

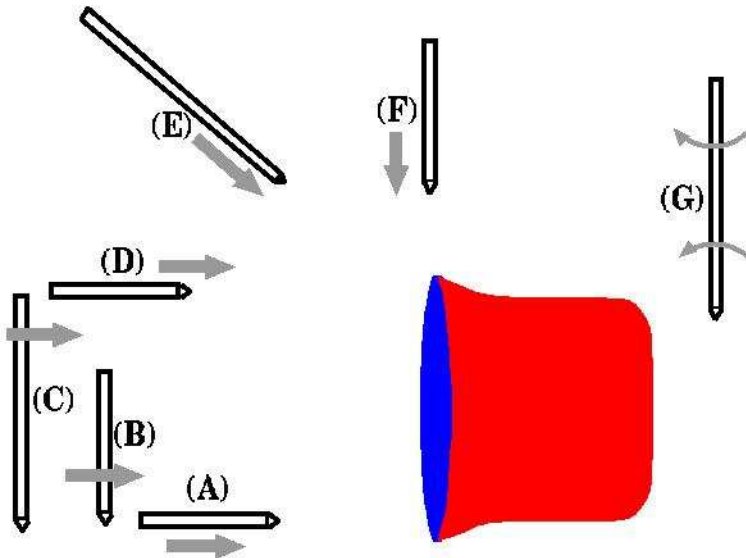


Figure 1

### Questions relating to Figure 1

Assume that all the pencils shown in the figure lie in the vertical plane through the axis of the mug. So they are all at the same distance from the viewer, as is the axis of the mug.

For each starting point and possible translation or rotation of the pencil we can ask questions like: will it enter the mug?, will it hit the side of the mug?, will it touch the rim of the mug?

In some cases the answer is clear. In cases where the answer is uncertain, because the configuration is in the "phase boundary" between two classes of configurations that would have clear answers we can ask how the pencil could be moved or rotated to make the answer clear. (Compare being unsure whether you are going to bump into something while walking: you can either try to look more carefully, use accurate measuring devices, etc. compute probabilities, etc. or you can alter your heading to make sure that you miss the object.)

#### • Questions about consequences of translations

- If pencil A moves horizontally to the right, will it enter the mug? If the answer is not clear, what vertical change of location of the pencil will make the answer clear?
- Similar questions for pencils B, C, and D.
- Is there a vertical motion of pencil C that will make it easier to tell whether the horizontal motion will enable the pencil to go into the mug? Is there some other motion of pencil C that will change it so that subsequent horizontal motion will obviously take it into the mug.



- If pencil E moves along its length towards the mug will the point go into the mug or will it hit the rim or outer surface?
- Is there a rotation or translation of pencil E after which the prediction becomes easier?
- Similar questions can be asked about F moving downwards.
- **Questions about consequences of rotations**
  - If pencil G is rotated in the vertical plane about its top end will it hit the mug? If the answer is not clear what translations will make the answer clear?
  - If pencil G is rotated in the vertical plane about its bottom end will it hit the mug?
  - What translations of pencil G will bring it about that it can rotate about the bottom end without hitting the mug?
- **Questions about alterations that will make the hard questions easy to answer.** (Several examples are given above.)

TO BE EXTENDED

**NB:**

The ability to answer such questions is required for PlayMate's ability to plan movements. The same comment applies to questions below.

## **Changing spatial relations to make a prediction problem easier**

As illustrated above, when predictions need to be made, an intelligent agent can move the object away from the 'difficult' position or trajectory so that it is far enough from the phase transition for fine control or precise predictions not to be required.

In some cases where being close to a phase transition makes a perceptual judgement difficult (e.g. will an object's motion lead to a collision?) it is possible to resolve the ambiguity by a change of viewpoint. Moving to one side, for example, may alter one's view of a gap so that it becomes clear whether the gap is big enough for an object to fit in it with space to spare. Some simple examples of problems requiring a change of viewpoint are given below.

Similar comments apply to relations not between objects but between their trajectories. The exceptions are hard to deal with, but very many cases are easy, without requiring great precision, because they concern topological or ordering relations rather than metrical information, and a change of viewpoint or slight modification of a trajectory may turn a difficult prediction into an easy one.

Another type of exception is related to the fact that in the 'easy' cases discussed above movements can be visualised in advance with accuracy sufficient for the task of deciding what will happen, and they can also be performed ballistically, without fine-grained feedback control. A different sort of situation occurs when the object being acted on is very small (e.g. it takes up a relatively small portion of the visual field, and relatively small changes in motor signals will always make a difference to whether a finger does or does not make contact with the object). Using a small tool e.g. small tweezers to manipulate such objects requires additional competences beyond those discussed above. But for now we can ignore such actions: they require expertise that probably develops later involving fine-grained visual servoing to control very precise small movements. Such cases are ignored here.

## **The importance of meta-management**

Much work in the Birmingham [Cognition and Affect project](#) has been concerned with the role of a 'meta-management' layer in an agent architecture, namely a layer of mechanisms providing various kinds of self-monitoring and self-control of internal states and processes.

There are several presentations on varieties of architectures, explaining such ideas, [here](#). A relatively simple tutorial is included in this [presentation on robotics and philosophy](#).

See also the remarks about fully deliberative architectures [here](#).

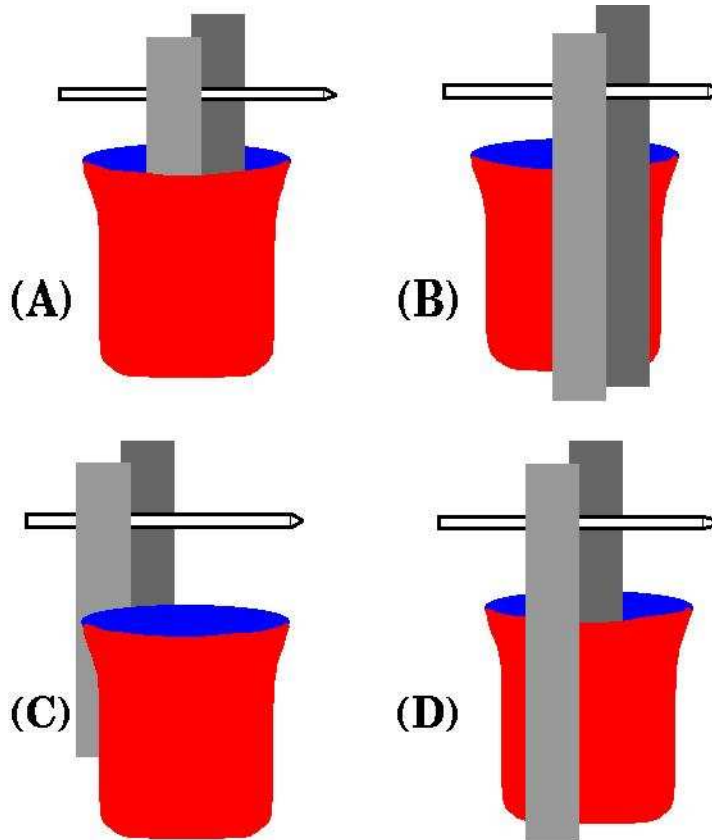
It is worth mentioning that meta-management capabilities are required for dealing with the problems of uncertainty mentioned above. The individual trying to predict how affordances will be changed if an action is performed, needs to be able to detect when that prediction is hard because the objects and trajectories are close to a 'phase boundary' so that only if precise, noise-free information is available can the prediction be made reliably. If such situations are detected, using a meta-management mechanism to evaluate the quality of current information, then working out how to change the situation so that the problem is removed, e.g. by moving an object or rotating it so as to move it further from the phase boundary can use a deliberative mechanism if the situation is unfamiliar, or a learnt reactive behaviour, if the situation is familiar.

## **Snapshots of various possible motion scenarios:**

### **Predicting consequences of motion**

**(i.e. changing affordances, without dynamics)**

The pictures below are somewhat idealised 'hypothetical' snapshots of situations in which motion can occur. Questions are asked about the pictures to illustrate some of the requirements for visual understanding of perceived structures. The examples add a requirement that was not included in the previous examples, namely a requirement to understand implications of things being at different distances from the viewer. However the scenes involve 2.5D configurations, i.e. the depth relations are merely orderings, without any metric.



**Figure 2**

What should a vision program be able to say about the above images (A), (B), (C), (D), each involving a mug, a horizontal pen, and two rigid vertical cards, if asked the following questions in each case:

- What will happen if the pen retains its orientation and is moved vertically downwards?
- What will happen if the pen retains its orientation and height and is moved horizontally left or right?
- What will happen if the pen retains its orientation and height and is moved horizontally towards or away from you (the viewer)?
- What will happen if the pen retains its orientation and height and is rotated (either way) about its centre in a horizontal plane?
- What will happen if the pen retains its orientation and height and is rotated (either way) about its centre in the vertical plane through the pen?
- If the visual information provided is too ambiguous or incomplete, is there a movement that the robot could make to get the information required to answer the question? (E.g. move camera left, move camera right, move camera over an object and look down, etc.)

---

Snapshots of slightly different possible motion scenarios:

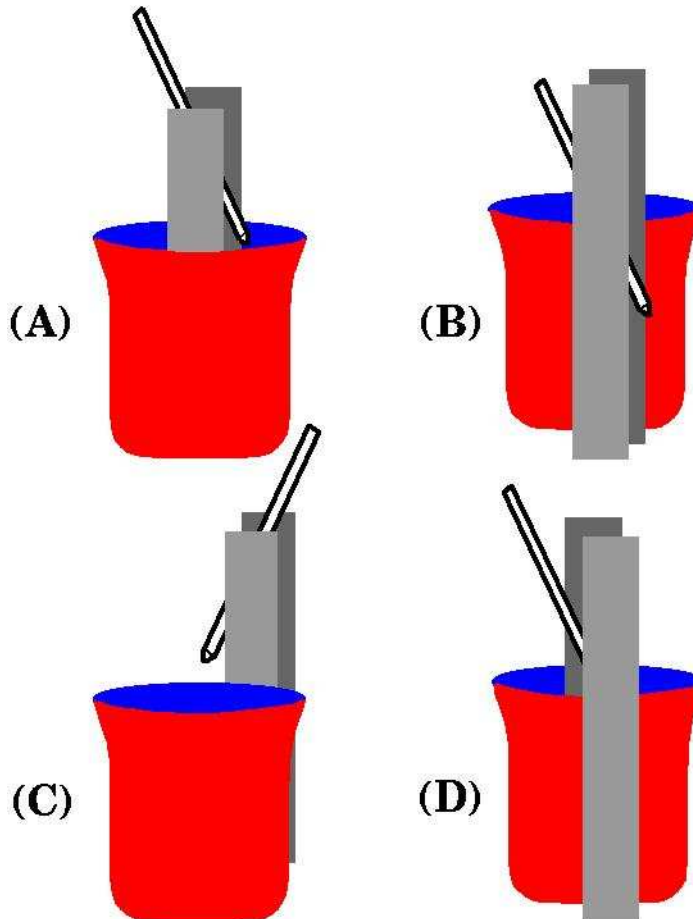
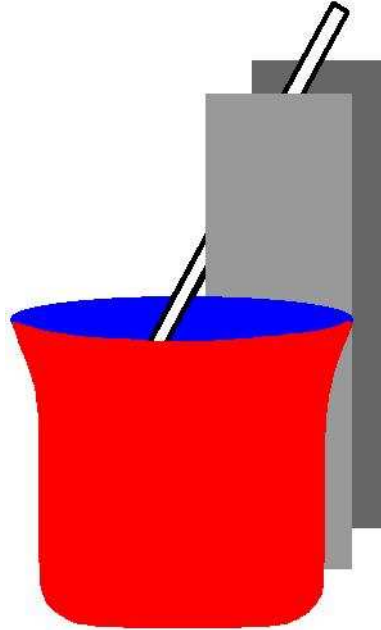


Figure 3

What should a vision program be able to say about the above images (A), (B), (C), (D), each involving a mug, a pen, and two rigid vertical cards, if asked the following questions in each case:

- What will happen if the pen retains its orientation and is moved along its length in the direction of its point?
- What will happen if the pen retains its orientation and height and is moved horizontally left or right?
- What will happen if the pen retains its orientation and height and is moved horizontally towards or away from you (the viewer)?
- What will happen if the pen retains its orientation and height and is rotated (either way) about its centre in the vertical plane through the pen?

- If the visual information provided is too ambiguous or incomplete, is there a movement that the robot could make to get the information required to answer the question? (E.g. move camera left, move camera right, move camera over an object and look down, etc.)
  - Would knowing the length of the pen enable the question to be answered without moving to get a different view?
- 



**Figure 4**

What should a vision program be able to say about the scene depicted in Figure 4?

Are there any actions a robot could take to shed light on what's going on?

Compare:

<http://www.cs.bham.ac.uk/research/projects/cosy/photos/penrose-3d/>  
Pictures based on the work of Oscar Reutersvärd (1934)

---

## **Visual servoing**

When a robot or animal is controlling its own motions, there are many examples of prediction of consequences of movement that are related to but different from the examples given.

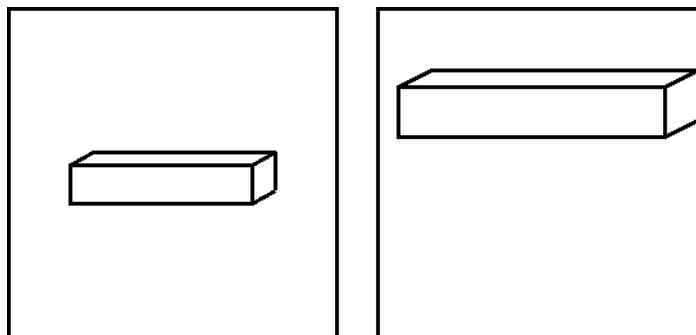
E.g. as the eye or camera moves forward the location of some object within the visual field indicates whether continued motion in a straight line will cause the eye to come into contact with the object or move past it.

Slightly more complex reasoning is required to tell whether a mouth or beak that is rigidly related to the eyes will be able to bite the object. That situation is analogous to the camera mounted on the PlayMate's arm, near its wrist, as shown here:



For example consider the problem of using camera images to control the motion of the hand with a wrist-mounted camera, when an object is to be grasped, or using eyes mounted above a mouth, when an object is to be grasped with the mouth.

Here are two schematic (idealised) images representing a pair of snapshots that might be taken from a camera mounted vertically above the wrist and pointing along the long axis of the gripper.



One of the images is taken when the gripper is still some way from the block to be grasped and the other is taken when the gripper is lower down, closer to the block. It should be clear which is which. Now, if the camera is mounted above the gripper is the gripper moving in the right direction?

For the robot to use the epistemic affordance here it has to be able to reason about the effects of its movements on what it sees and how the effects depend on whether it is moving as intended or not. It is possible that instead of explicit reasoning (of the sort you have probably had to do to answer the question) the robot could simply be trained to predict camera views and to constantly adjust its movements on the basis of failed predictions.

In one case it needs explicit self knowledge, which can be used in a wide variety of circumstances, and in the other case it needs implicit self knowledge, produced by training, which is applicable only to situations that are closely related to the training situations.

A human making use of the epistemic affordance by reasoning about the information available from the differences between the two views, may make use of logic, a verbal language, and perhaps some mathematics. A less intelligent animal or robot may have that information pre-compiled (e.g. into neural control networks) by evolution or previous training and available for use only in very specific control tasks.

Is there some intermediate form in which the information could be represented and manipulated that could be used by an intelligent animal to deal with novel situations, and which does not depend on knowing logic or a human-like language, but might make use of what we have been calling a GL (a Generalised Language), which has structural variability and compositional semantics and may involve manipulation of representations of spatial structures?

See: <http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0703>

Computational Cognitive Epigenetics

(Sloman and Chappell, to appear in BBS 2007)

In all cases visual servoing requires what could be described as 'self-knowledge' insofar as it involves explicit or implicit knowledge about the agent's situation and actions that can be used to make predictions and to interpret discrepancies between predicted and experienced percepts, and to use those discrepancies to alter what it is doing.

But this does not require an *explicit* sense of self if that implies that the robot (or animal or child learning how to bite things or grasp things) is able to formulate propositions about its location, its actions, its percepts, its goals, etc.

## **Does reasoning about grasping have to be probabilistic?**

Video input from a real camera will be far more complex, noisy and cluttered than the idealised line drawings depicted above. As a result it will be difficult to locate the edges, corners, axes, centroids, etc. of image components accurately, or to compute distances or angles between them accurately.

One way of dealing with that is to attempt to estimate the uncertainty, or the probability distributions of particular measures, and then to develop techniques for propagating such information in order to answer questions about what is going on in the scene, where the answers will not use precise measures but probability distributions.

Another way is to find useful higher level, more abstract descriptions, whose correctness transcends the uncertainty regarding the noisy image features. So for example, the change between the left and right images above could be described something like this (though not necessarily in English):

In the second picture, the image of the target object is larger and higher in the field of view.

The uncertainty and noise in the image can be ignored at that level because all the uncertainty in values in the images is subsumed by the above the description. The description does not say what the exact sizes of the of the images are in the two pictures, or the exact locations, or the exact amount by which it is larger or further from the bottom edge.

So since the gripper is below the camera, the fact that the image is moving up the field of view means that the direction of motion of the gripper is towards a point below the target, requiring the motion to be corrected by moving the wrist up. Exactly how much it move up need not be specified if the motion is slow enough and carefully controlled to ensure that the target object moves towards a location that has previously been learned is where it should be for the gripper to engage with it. If the gripper fingers are moved far enough apart the location need not be precise, and if there are sensors on the inner surface of the fingers they can provide information about when the object is between the fingers and the grip can be closed.

This description is over-simplified, but will suffice to illustrate the point that there is a tradeoff between precision of description and uncertainty and that sometimes the more abstract, less precise, description is sufficiently certain to provide an adequate basis for deciding what to do.

## Note on nest-building birds

Birds that build nests out of twigs, leaves and similar materials need to be able in some sense to understand and use changing affordances as they move twigs and other objects around during the construction process.

Future domestic robots will also need to have such competences.

The abilities to predict changing affordances form a special case of understanding causal relationships, in particular Kantian causal relationships, as discussed in <http://www.cs.bham.ac.uk/research/projects/cogaff/talks/wonac>

---

## How is the reasoning done?

When humans solve the prediction problems described above we seem to be making use of manipulable models of 2-D structures, containing parts that can be moved and rearranged, along with the ability to detect new contact points arising.

Compare Sloman 1971 on the Fregean Analogical distinction:

<http://www.cs.bham.ac.uk/research/cogaff/04.html#200407>

Brian V. Funt, 1977

WHISPER: A Problem-Solving System Utilizing Diagrams and a Parallel Processing Retina  
IJCAI 1977, pp 459-464

<http://dli.iit.ac.in/ijcai/IJCAI-77-VOL1/PDF/077.pdf>

Usefully summarised in

Zenon Kulpa

Diagrammatic Representation And Reasoning

*Machine GRAPHICS & VISION*, Vol. 3, Nos. 1/2, 1994, 77-103

<http://www.ippt.gov.pl/~zkulpa/diagrams/Diagres.pap.pdf>

See also: Kulpa's

[Diagrammatics web page](#)

*Written circa 2007*

I think a relatively simple computer implementation could be built and used as part of a visual reasoner in CoSy, using techniques used in graphical software for making and editing diagrams, e.g. TGIF, XFIG, etc.

(Tgif saves all of its diagrams in a logical format, using Prolog.



<http://bourbon.usc.edu/tgif/>

It can generate 2-D displays from the Prolog specification, and mouse and keyboard interactions with the display can lead to a new Prolog specification of the display.)

The hard part will be parsing real visual images to produce the required 2-D manipulable representations.

***Comment added 18 Oct 2015***

It is clear that what I wrote above in 2007 was over-optimistic. The techniques for manipulation used in graphical tools mostly operate on metrically precise structures, and normally do not support reasoning about what is and is not possible. There may be more recent work on geometrical and topological theorem proving that is relevant, though I suspect everything done so far uses forms of representation and reasoning that are very different from those used in animal brains for reasoning about affordances. For further discussion of abilities to perceive and reason about possibilities and impossibilities (constraints) see the following:

**Some (Possibly) New Considerations Regarding Impossible Objects**

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/impossible.html>

Their significance for mathematical cognition, and current serious limitations of AI vision systems.

Slightly easier will be software to:

1. Manipulate the parsed 2-D images, e.g. by sliding one structure in a specified direction while leaving other structures unchanged, or rotating a structure around a specified point while preserving its shape.
2. Detect consequences of continuous movements of one or more parts of the diagram, e.g. detecting when a moving circle first comes into contact with a fixed triangle, or detecting when the bottom portion of a partially occluded rectangle behind a circle becomes visible as the rectangle is moved horizontally.

For affordance prediction and the avoidance of phase boundaries it may be useful to be able to grow a "penumbra" of specified thickness around the 2-D image projection of any specified object, and then when an object A moves in the vicinity of object B, d

- (a) detect when A's penumbra first makes contact with B's penumbra and where it happens;
- (b) detect when one of the penumbras first makes contact with the other object (inside its penumbra)
- (c) detect when A itself first makes contact with another object (inside its penumbra)

Choosing penumbra sizes to facilitate reduction of uncertainty will require programs that can analyse aspects of the structure of a scene and detect whether some relationship introduces uncertainty in predictions. Then choosing a penumbra size to use when selecting a movement that is certain not to produce a collision will be a task dependent problem.

[All this is closely related to Brian Funt's PhD. See reference below.]

NOTE: I suspect that a detailed analysis of the suggestions here could involve developing some interesting new mathematics.

## Other connections

Arnold Trehub's retinoid mechanism may be useful:

The Cognitive Brain (MIT press, 1991)

<http://www.people.umass.edu/trehub/>

As mentioned above this work on predicting affordance changes is related to my recent work with Jackie Chappell on GLs (Generalised Languages) evolved for 'internal' use in precursors of humans as well as many other mammals, e.g. chimpanzees and possibly hunting mammals, and in some bird species. GLs are also required by pre-verbal children. See

<http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#glang>

<http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#talk111>

What evolved first: Languages for communicating, or languages for thinking (Generalised Languages: GLs)

---

## Implications for natural language interactions

If a robot can perform actions in order to change affordances, whether action affordances or epistemic affordances, then this provides a natural topic for situated dialogue.

Examples:

Why are you hesitating?

To check whether my hand will bump into the cube

Why did you move your head left?

To get a better view of the size of the gap between the cube and the block

Can your hand fit through the gap between the two blocks?

I am not sure, but I'll try

Can your hand fit through the gap between the two blocks?

I am not sure, but I can move them apart to make sure it can.

Is the block within your reach

Yes because I just placed a cube next to it.

How can you get the cube past the block?

Move it further to the right to make sure it will not bump into the block then push it forward.

etc. etc.

There is a wide variety of propositions, questions, goals, plans, and actions, dealing with a collection of spatial, causal and epistemic relationships that can change. If we choose a principled, but non trivial subset related to what the robot can perceive, plan, reason about, and achieve in its actions, then that defines a set of questions, commands, assertions, explanations, that can occur in a dialogue.

---

## How much of the above could a robot learn?

At a later date we could move back to an earlier stage and instead of building all the above competence in, enable the robot to learn some of it.

That will require working out a suitable initial state, including initial forms of representation, competences, and architecture that is able to support the development of a suitable altricial competence.

See

<http://www.cs.bham.ac.uk/research/projects/cogaff/07.html#717>

COSY-TR-0609 (PDF):

Natural and artificial meta-configured altricial  
information-processing systems

Jackie Chappell and Aaron Sloman

Invited contribution to a special issue of The International  
Journal of Unconventional Computing

Vol 2, Issue 3, 2007, pp. 211--239,

---

## Related documents

- H. Rex Hartson (2003), Cognitive, physical, sensory, and functional affordances in interaction design, in *Behaviour & Information Technology*, 22, 5, pp. 315--338,  
<http://research.cs.vt.edu/usability/publications/Hartson%20-%20affordances%20-%20BIT.pdf>
- <http://www-cvr.ai.uiuc.edu/~seth/ResPages/pdfs/HutHagCor96.pdf>  
S. Hutchinson and G. Hager and P. Corke,  
'A Tutorial on Visual Servo Control',  
in *IEEE Trans. on Robotics and Automation*, 12, 5, Oct, 1996,  
pp. 651--670,
- <http://www.cs.bham.ac.uk/research/projects/cogaff/06.html#0604>  
Aaron Sloman, (1982), Image interpretation: The way ahead?, In  
*Physical and Biological Processing of Images*  
Eds. O.J. Braddick and A.C. Sleight, pp. 380--401, Springer-Verlag, Berlin,  
(Proceedings of international symposium organised by The Rank Prize Funds, London, 1982.),
- <http://www.cs.bham.ac.uk/research/projects/cosy/photos/penrose-3d/>  
A Challenge for Vision: Seeing something impossible  
Episodic memory for impossible objects.
- <http://www.cs.bham.ac.uk/research/projects/cosy/photos/crane>  
A Challenge for Vision: Seeing a Toy Crane  
Crane-episodic-memory
- <http://www.cs.bham.ac.uk/research/cogaff/96-99.html#15> "Actual Possibilities", in *Principles of Knowledge Representation and Reasoning: Proceedings Fifth International Conference on Knowledge Representation (KR '96)*, Morgan Kaufmann Publishers, 1996, pp 627-638.

- <http://www.cs.bham.ac.uk/research/projects/cogaff/misc/vision>  
Aaron Sloman, 'How to look at what you see.'  
Discussion paper still under development, 2013-14  
A presentation of some hard, apparently unsolved, problems about natural vision and how to replicate the functions and the designs in AI/Robotic vision system
- 

**Maintained by:**

Aaron Sloman

<http://www.cs.bham.ac.uk/~axs/>