Actual Possibilities

Aaron Sloman*
School of Computer Science
The University of Birmingham
Birmingham, B15 2TT, England
A.Sloman@cs.bham.ac.uk

Abstract

This is a philosophical 'position paper', starting from the observation that we have an intuitive grasp of a family of related concepts of "possibility", "causation" and "constraint" which we often use in thinking about complex mechanisms, and perhaps also in perceptual processes, which according to Gibson are primarily concerned with detecting positive and negative affordances, such as support, obstruction, graspability, etc. We are able to talk about, think about, and perceive possibilities, such as possible shapes, possible pressures, possible motions, and also risks, opportunities and dangers. We can also think about constraints linking such possibilities. If such abilities are useful to us (and perhaps other animals) they may be equally useful to intelligent artefacts. All this bears on a collection of different more technical topics, including modal logic, constraint analysis, qualitative reasoning, naive physics, the analysis of functionality, and the modelling design processes. The paper suggests that our ability to use knowledge about "de-re" modality is more primitive than the ability to use "de-dicto" modalities, in which modal operators are applied to sentences. The paper explores these ideas, links them to notions of "causation" and "machine", suggests that they are applicable to virtual or abstract machines as well as physical machines. Some conclusions are drawn regarding the nature of mind and consciousness.

1 Introduction: possibilities everywhere

This paper is a conceptual exploration and does not pretend to offer any formal results. It suggests some new ways of looking at old problems which may be relevant both to AI researchers designing robots or systems that need to reason about complex mechanisms, and to cognitive scientists investigating human and animal capabilities.

The orientation is primarily ontological, i.e. concerned with the sorts of things whose existence we (or animals or future robots) take for granted both in our perception and in our problem solving and reasoning. Exactly how we (and other animals or robots) represent and reason about the sort of ontology described here is a topic for future research, which will have to formalise the ideas and embody them in useful mechanisms. Much vagueness is inevitable at this stage and not all questions raised will be answered nor all the conjectures justified.

The key idea is that the world contains not only objects with properties and relationships, but also *possibilities* and *links* between or *constraints* on possibilities. The latter are the causal powers of objects. Some of the possibilities are things we like or dislike, hope for or fear, and we label them risks, dangers, opportunities, prospects, temptations and so on. Others, described more neutrally, are of interest because of the role they play in explanatory and predictive theories, for instance, possible velocities or accelerations of an object, possible voltages across a conductor, possible pressures in a gas. Often the functional role of an object is concerned with possible states or processes that object can enter into, which, in turn, will change the possibilities for other things. An old example from planning is an action whose effects influence the preconditions for another action.

1.1 Perception of possibilities

In Gibson's theory of perception as the acquisition of affordances (Gibson 1986), the key idea is that for an organism merely to be given information about the structures of objects in the environment is not sufficient for its needs. Gibson's claim is that many organisms have evolved the ability to perceive what he calls "affordances", which involve sets of possibilities and constraints on possibilities. For instance, a surface affords support for the animal, an

* In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifth International Conference (KR '96)*, Morgan Kaufmann Publishers, 1996

opening in a wall affords passage, a berry affords picking and eating. He makes a second claim that these are not inferences made by central cognitive mechanisms on the basis of structural information delivered by the visual system. Rather the detection and representation of affordances happens deep within the perceptual mechanisms themselves. The first claim may be true even if the second one is not.

I tried to elaborate both claims in Sloman (1989). Compare Pryor (1996) for related ideas regarding "reference features" used by reactive planners. I'll return below to questions about the nature of the various types of affordances that animals can perceive and use in their actions, and about how their knowledge is represented and manipulated.

1.2 Possibilities, causation and change

Some features of an object are permanent while other features and relationships can change[1]. Objects may change their size, their colour, their temperature, their orientation, their distance apart, their containment relationships and so on. The properties and relationships of an object that are capable of changing may be thought of as "selections" from ranges of possible properties and relationships. (I don't mean that selections are conscious or deliberate). When things change different selections are made.

Some of these sets involve global features, such as size, weight, or orientation. Some involve relations of the whole object to other things, e.g. containment, distance and direction from other objects, motion towards something, attraction between objects, and many more.

Some of the sets of possibilities inherent in an object are analysable in terms of possibilities involving *parts* of the object. For example whether an object is striped or not depends on whether parts have the same colour. The shape of an object depends on spatial relations between parts of the object. The degree of compression of an object depends on both spatial relations between and forces between parts of the object. The degree of flexibility involves possible changes in spatial relations between parts.

The sets of possibilities inherent in an object help to define the nature of the object, or at least what it is about the object that needs to concern an animal or robot that interacts with it, or a designer who embeds it in a larger configuration. Some of these possibilities will be perceivable simply on the basis of sensory qualities (e.g. those that depend on shape, spatial relations, observed motion, resistance to pressure, tactile qualities, etc.) while others require inference or learnt associations, e.g. edibility, risk of being stung, what an agent can see, etc.

[1] Degrees of permanence are discussed later in connection with remoteness of possibilities

1.3 Constraints and causal links

Selections from different sets of possibilities cannot all be made independently: there are *laws* or *constraints* linking different ranges of possibilities. These links, which may depend on context as well as the nature of the object, are themselves higher level properties, and insofar as the links or constraints can change they too may be selections from more abstract (higher order) ranges of possibilities, i.e. possible constraints between possibilities.

Certain properties of an object can be directly altered from outside, for instance, its temperature, or the voltage across it. This involves a selection from one of the ranges of possibilities. Typically this will *cause* the selections from one or more other sets of possibilities to change. For example, changing the temperature may cause the length to change. Exactly how changes in one set cause changes in another depends on the object (and possibly its environment). We can think of the object as a "transducer" linking sets of possibilities.

1.4 An example: electrical resistance.

Many physical properties of objects involve transduction relations between possibilities, for instance electrical resistance of a conductor.

This can be understood without knowing about deep explanations of electrical phenomena. It suffices to know that some objects are able to conduct electricity, that flow of an electric current can be produced by applying a voltage across the conductor, that both voltage and current can be measured and that there is always a fixed relationship between the two states.

More precisely, talking about resistance of a conductor presupposes that there is a set P1 of possible voltages across the conductor, a set P2 of possible currents in the conductor, and that the conductor has a property R that limits the selections from the two sets of possibilities so that the ratio between voltage and current is constant.

So a piece of wire can be seen as a transducer from one range of possibilities (possible voltages, P1) to another range of possibilities (possible currents, P2) with the ability to constrain the relationship between inputs (voltages) and outputs (currents). That ability is its electrical resistance.

Although we measure both the voltage and the current using numbers, the numbers are multipliers for something physical, unit voltage and unit current, which depend on the system of measurement in use. So resistance is not just a numerical ratio, though it can be represented by one.

Thus construed, resistance is an abstract property of conductors. Abstract properties typically are *implemented* in other properties. Some time after learning about resistance, physicists learnt more about the architecture of matter, including underlying mechanisms and properties in terms of which resistance (and the ability to conduct electricity) is "implemented". I shall not discuss the philosophical question whether the resistance IS those other things or not. Certainly physicists knew about resistance as characterised here before learning about electrons, quantum mechanics, etc. So there is at least an epistemological distinction between resistance and the underlying properties, whether there is an ontological difference or not.

There probably still are many engineers who know nothing about the underlying physics, and manage with concepts of properties defined solely in terms of ranges of possibilities and constraints linking those ranges of possibilities, like the constraint expressed in the equation **V = RC**. For engineering purposes, that sort of "surface" knowledge is often more useful than deeper knowledge about the underlying physical implementation. However, the deeper knowledge may be required for the task of designing new conducting materials. In that case we have to understand how features of the architecture of matter are relevant to the possibility of electrical currents.

2 Other physical properties

Electrical resistance is just one among many examples of properties linking sets of possibilities. A conducting piece of wire may also have a modulus of elasticity, which is a physical property associated with two ranges of possibilities, namely possible tensile forces that can be applied to the ends of the wire and possible changes in length of the wire. As with resistance, the wire links particular possibilities in the first range with selections in the second range, and within a sub-range of each set the ratio between the measures of the input possibility and output possibility is fixed. Beyond that subrange inelastic deformation occurs.

In some conductors changes in temperature produce changes in resistance. Here we have a range of possible temperatures P3, linked to a range of possible resistances P4, where each resistance is itself a particular link between sets P1 and P2, i.e. possible voltages and possible currents through the conductor. So some physical properties involve second-order possibility transducers. The extent to which temperature changes elasticity is another second-order link.

Some conducting devices are designed so that turning a knob or moving a slider changes the resistance between two terminals, by changing the length of wire used between the terminals. Here the set of possible orientations of the knob is mechanically linked to the set of possible lengths of wire through which the current flows, and this is linked to a set of possible resistances. Thus the relation between knob position and resistance is another second order link.

Not all causal links between possibilities involve linear relationships. In an amplifier's volume control the link between possible orientations and possible resistances might be logarithmic. In a gas held at a fixed temperature changes of pressure produce changes in volume and the first is inversely proportional to the second, at least within certain pressure and temperature ranges. However, a monotonic non-linear relation can often be transformed into a linear one by remapping either P1 or P2. (E.g. replace volume with its inverse.) But there are more interesting exceptions.

2.1 More complex interactions

Many of the possibility-linking properties are expressible in terms of mathematical functions linking numerical values, but we also have names for linkages between qualitatively different possibilities. For example a vase admits a variety of possible interactions with other objects: it could be struck by a fast moving bullet, hit by a slower moving sledge hammer, squashed between two plates moving together very slowly, or dropped on a hard floor. Call the set of possible damage-causing interactions P1. The vase is also capable of changing into a very wide variety of states in which its wholeness is lost: it may break into two pieces in many different ways, or it can break into large numbers of fragments, also in very many different ways. Call that set of possible states P2. Depending on the material and its shape, it will have a causal property linking the occurrence of any element of P1 to a selection from P2.

We have various names for such properties, including "fragile", "brittle", "breakable" and "delicate". (The latter is sometimes used to describe the type of *appearance* that is associated with the causal property of being delicate, i.e. easily broken because components are very slender.) If the vase is made of glass or china it will probably be both brittle and fragile. If the material is very thin it may also be delicate. It is interesting that we also use these words to describe more abstract properties of non-physical objects, for example a fragile personality, delicate health, fragile ecosystem, etc. because of strong analogies between different sorts of causal properties linking quite different sets of possibilities.

These are all standard examples of *dispositional* concepts, concerned with "what would happen if...", i.e. all referring to linkages between sets of possibilities. Many engineers, technicians and craftsmen have to become familiar with a wide range of such properties, including knowing how to recognise when they are present, knowing how to make use of them when designing, making or fixing things, knowing how to recognise the conditions that will trigger the linkages, knowing how to prevent such conditions arising, e.g. by choice of storage conditions or packing materials.

2.2 Combinations of inputs are sometimes relevant

So far I have written as if all possibility transduction involves a link between two sets of possibilities P1 and P2. However, there are many cases in physics where one measurable depends on two or more others. E.g. in every case where I talked about a second-order transducer, it is possible instead to talk of a first-order transducer with two inputs. E.g. current in a wire can depend on both voltage and temperature. We just don't happen to have a name for this relationship between three sets of possibilities, though perhaps we could find one useful.

Many of the properties linking sets of possibilities depend on a combination of factors such as the material used, the shape or structure of the object, and the environment. For instance a vase made of lead will behave very differently from a vase made of glass when dropped. Moreover the properties may change when the vase is at a very high or very low temperature, or immersed in a liquid, e.g. treacle. Thus in general the selection from the set P2 can depend not just on a selection from one other set P1, but on a host of selections from other sets, P1a, P1b, P1c, etc. some involving internal states of the object and others relationships with or states of the environment. Full replication of human abilities to use knowledge about objects in deciding what to do, making plans, controlling actions, would have to take account of all these different cases.

2.3 The input and output sets may be hard to characterise

The sets of possibilities linked by a causal property may be very hard to characterise in a formal and general way. We have an intuitive grasp of the range of possibilities P1 that would trigger a manifestation of fragility in a vase but would find it difficult to list necessary and sufficient conditions. Similarly the set of possible outcomes P2 associated with notions like fragility or brittleness can be very large and varied, and hard to characterise.

Nevertheless people can learn about both sets as well as learning how to detect objects which link P1 and P2. That raises interesting questions about how the learning is done and the form in which the knowledge is stored and used. It seems that the expertise of a craftsman or engineer can include the ability to distinguish far more different sorts of cases than we have words for in our language. For instance the word "fragile" can describe both a vase and a spider's web, yet different classes of input and output possibilities are relevant to their fragility.

2.4 The causal links may have different forms

The knowledge we have concerning the causal links between P1 and P2 may be very different in form, depending on the case. For example, knowing the resistance of a conductor involves knowing a very precise relationship between possible voltages and possible currents. The mapping between members of P1 and P2 is one to one and easily formalised. In the case of a fragile and brittle vase all we know is that each item in P1 will realise *one* of the possibilities in P2.

Common sense knowledge includes the ability to make coarse-grained predictions, for instance about the difference between dropping a vase and squashing it between two sheets of metal, or the difference between striking it on the side with a hammer and striking it with a downward blow. An expert may, as a result of years of experience, learn more about the relationship between members of P1 and members of P2, and on that basis can at least partially control the manner in which something breaks, for instance the occupant of a house breaking a window in such a way as to make it look like the work of a burglar doing it from outside, or a demolition expert using explosives to bring down a large disused chimney safely.

2.5 Discrete outcome devices

Gambling devices, such as a roulette wheel or a device in which a descending ball bounces against pins until it falls into one of a set of slots also allow only limited prediction. The difference is that, unlike a fragile vase, these devices are constructed so as to have a small finite set of end states (so that there's a well defined set on which to place bets). It is a remarkable fact about such devices that although particular outcomes are unpredictable their statistical properties are predictable: they have well defined probability distributions and these can be adjusted, e.g. with hidden magnets and the like.

So here we have a set of initial possibilities P1 and outcomes P2 where we cannot find specific links between selections from P1 and P2, whereas we do find a very strong link between sequences of selections from P1 and global numerical properties of the resulting selections from P2: hence the concept of a *probability distribution* as a property of an object. There are many philosophical puzzles about the concept of probability but I shall not go into them here. (Popper's notion of "propensity" is relevant.)

2.6 Causal determinism

In the case of a shattered vase we do not have the vocabulary to describe all the possible initial events nor the possible outcomes in precise detail. In the case of precisely engineered gambling devices we may have the vocabulary, but we still have no way of predicting precisely the outcome of an interaction, even if we use very precisely engineered machinery to replicate initial conditions (which is harder when we have to manufacture multiple vases under identical conditions).

A causal determinist will argue that in all these cases there is a precise and rigid linkage between initial and final states, but it is merely our knowledge that is incomplete: knowledge of the fine details of the initial conditions and knowledge of all the physical laws involved. The assumption is that if our knowledge were complete, predictions could be precise. That assumption is often taken to be a feature of classical physics.

Where chaotic phenomena are involved only *infinite* precision in the initial states could uniquely determine future outcomes. Clearly we cannot, even in principle, obtain infinitely precise measures of physical states (especially if those states can vary continuously, in which case most of the states could not be described using any finite description). Whether the actual states of physical systems could have infinite precision is not clear. It is normally taken to be an assumption of classical physics that they do, but since no classical physicist ever had any reason to believe that infinitely precise measurements or descriptions were possible, it is not clear how they could have believed that the states themselves were infinitely precise.

This opens up a new line of thought: even if the *laws* of physics are totally determinate, e.g. expressed in simple differential equations, it might still be the case that the *states* and *boundary conditions* cannot be infinitely precise. In that case systems with non-linear feedback such as produces chaos might be intrinsically unpredictable even within classical physics.

That makes it even more remarkable that when the geometry of a chaotic gambling device forces a finite set of possible outcomes the long run frequencies should be determinate. This line of thought reduces some of the differences between classical physics and quantum physics: both can involve non-deterministic selections among finite sets of possibilities where the selections obey only statistical laws. The difference is that in these devices the classical possibilities are constrained by the geometry. This may not hold for all classical processes, e.g. the breaking of a vase may be unlike the process on a roulette wheel.

We've seen that engines that use no principles of quantum mechanics involve linked networks of possibilities. Any particular state of such an engine, or any particular extended piece of behaviour over a time interval is but one among many possibilities inherent in the design. All those possible voltages, currents, rotations, velocities, forces, etc. are real possibilities in the sense that the configuration of the machine allows them to occur, while other combinations of possibilities occurring in particular spatio-temporal relations are ruled out by the configuration. We'll look at different sorts of dynamics later.

As explained above, even when the details are unpredictable we may still have useful qualitative knowledge, which can guide our decisions regarding how to transport objects, how to predict types of breakage in a vase, how to make a broken window suggest a burglary, etc.

3 Types of knowledge and contexts of use

When a craftsman makes an object, some of the properties of the material in an object are particularly important during the process of construction and manufacture, others most important in the finished product.

E.g. the fact that wood can be cut, planed, sanded, etc. is important when an article is being made, whereas in the finished article hardness, rigidity, and low thermal conductivity may be more important.

Similarly some aspects of the shape of a building may be important during the process of construction, for instance making a partially completed structure stable, thereby reducing the need for scaffolding.

A particularly important sort of knowledge that the craftsman uses is knowledge of how to combine or shape materials so as to constrain the causal linkages. This may include grasping ways in which causal powers of some parts of an object combine with or constrain causal powers of other parts to produce properties of the whole. The global shape of an object, for example, involves various parts interacting with other parts to produce not only global geometric and topological features, but also global causal powers.

For instance, some shapes will make an object less liable to break under stress. Some shapes, such as boxes and bowls, will enable an object to hold other objects in place, which might otherwise roll around e.g. while being transported in a bumpy vehicle. Some shapes will contain a liquid. Some, but not all, will do both. I'll return to the role of knowledge in design later.

3.1 What do other animals know?

It is an interesting question to what extent animals that build nests or use tools or select where to walk or dig, or which branch to leap onto, have some grasp of the causal powers of different materials and of different structures. Do nest building birds have any grasp of the difference between ways of assembling materials that will produce a rigid structure and those that will not?

If Gibson is right about perception and affordances, then perhaps the magpie that uses its beak to insert a twig (which may be tens of centimetres in length) into its partly built nest, selects the location and direction of movement at least partly on the basis of perceiving the affordances for insertion and relative motion. The twig is held roughly at right angles to the beak, and even flying back to the nest with it, avoiding collision with branches on the way, is no mean achievement.

Insertion then requires first moving forward to one side of the insertion point (e.g. to the right of it), then inserting the end of the twig with a sideways motion of the head (e.g. to the left), and then repeatedly releasing the twig and grasping it nearer the free end and pushing it deeper into the nest with a sideways motion. Because of the intrinsic variability of shapes and sizes of twigs, partly built nests, and the configuration of branches in the supporting tree, no rigidly predetermined sequence of movements could achieve the task. So it looks as if the magpie needs some understanding of what it is doing.

Whether it also understands the consequences for the *long term* properties of the nest is doubtful. But even the ability to assemble the nest seems to involve using knowledge and abilities in a way that goes beyond what we currently know how to program into robots. This includes (a) the ability to perceive the affordances in the environment, (b) the ability to use the affordances to select actions (including selecting items for the nest, choosing route details while flying back without crashing into branches, deciding where to insert new bits into the nest) and (c) the ability to perform the actions, e.g. controlling fine details of the movement using visual and tactile feedback of changing affordances. In humans studies of how different sorts of brain damage can limit sorts of sub-skills may give clues as to which abilities to perceive, understand and use possibilities are used for different tasks.

4 Possibility transducers in virtual machines

So far all my examples have involved physical objects with physical properties which link and constrain sets of physical possibilities. Interactions with other information processing agents (not all of which are friendly) make new classes of possibilities and possibility transduction important. For example it may be important to think about possible items of information another agent can access (e.g. what it might see or hear), what information it might have gained previously, and how it could use different items of information, e.g. in taking decisions or making predictions. Our perception of faces uses mechanisms designed not only to discriminate and recognize individuals but also to detect a variety of types of *internal* states. These must have evolved together with mechanisms for displaying those states. Similar capabilities are found, though in simpler forms, in many other animals, though it is very hard to tell exactly which possibilities they detect and reason about.

Machines can also have information processing capabilities. In computer science and software engineering, it is now commonplace to think about abstract non-physical entities as having causal powers.

For instance, a word processor includes characters, words, sentences, paragraphs, lines, pages, page numbers, and other abstractions. It also allows a host of possible changes in any configuration of a document, for instance, inserting, deleting moving, changing font sizes, changing line spacing, etc. Each such change has potential consequences, such as altering the length of a line, changing line breaks, changing the contents of a page, changing page numbering, triggering a spelling checker and so on. How a particular change affects the configuration may be controlled by other aspects of the state of the program that can be changed by different sorts of actions. For instance setting the line length or the page size will alter the effects of inserting a word. So here too there are higher order possibility transducers. But most of the possibilities are not physical.

Although the wordprocessor may display part of its current state in a physical image on the screen, the program primarily constructs and manipulates *abstract* structures. They are datastructures in a virtual machine, like the structures in operating systems, compilers, theorem provers, planners, plant control systems, office information systems, etc. Some of these will be entirely concerned with processes in virtual machines. Others are part of a larger structure that includes physical states and processes of complex systems linked to the software.

One of the features characteristic of such systems is that typically the changes that occur are not continuous changes in the values of a set of numerical variables. In general there are discrete changes in more or less complex structures and many of the events involve changes in complexity, or structural changes such as switching two sub-trees in a tree. A formal characterisation of such a set of possibilities would therefore typically require something more like a grammar defining a class of possible structures than a fixed list of variables whose values can change. I.e. the dynamics of such structures cannot easily be construed in terms of motion in a high dimensional vector space.

Of course, the underlying *implementation machine* will typically use a large boolean state vector: but that's normally irrelevant to understanding the design or principles of behaviour of a complex piece of software, just as the precise configuration of atoms in a car may be irrelevant both to the driver and the car mechanic, since both operate best at higher levels of abstraction. However, some of the high level behaviour of the car may be best characterised in terms of equations linking a fixed set of numerical variables, e.g. speed, acceleration, torque, fuel consumption, air resistance, coolant temperature, and many more.

Describing and explaining changes in a computer will normally require quite different sorts of mathematics.

To illustrate: understanding how to use a word processor or compiler or operating system involves acquiring a grasp of the ontology used by the software. That in turn involves not only learning which objects and configurations of objects can exist but also grasping the additional possibilities and constraints on or links between possibilities inherent in various configurations. In a software system, like a word processor, these will constantly be changing during use, just as the physical affordances constantly change during the building of a nest or some other physical structure. Expert users of software systems have to develop the ability to perceive, represent, and make use of these affordances, including the high level affordances.

A software design environment is one in which the affordances involve very high orders in multiple domains. For example there is typically some sort of editor which has its own affordances for text construction and manipulation, along with the file management system. Then there may be one or more compilers which afford transformations of the text being constructed so as to produce new software systems with their own affordances. The process of editing not only changes the structures and the affordances in the source text configuration but also the affordances for the compilation process and structures and affordances in the resulting software. Moreover as the program text develops not only are there new possibilities for textual changes there are also new possibilities at the level of the program design, for instance, introduction of a macro provides opportunities to simplify syntax both in existing text and in new text, and development of a procedure changes possibilities for use of that procedure as a subroutine in future procedures. If the development environment involves an interpreter or an incremental compiler and the editor is part of the same software system, as in many AI development environments, all these possibilities are linked in very intricate ways, which can be hard to learn, but once learnt can speed up development of sophisticated software considerably.

But what exactly does the user learn? How is the information about all these abstract and rapidly changing affordances represented? How is that representation used both in the process of high level design and the relatively low level choice of text manipulation operations while coding, running tests, inspecting test output, etc.?

These examples show that virtual machines in software systems, like vases, admit of complex and varied sets of possibilities. However, like gambling devices, the underlying machinery is designed to allow only discrete sets of possibilities, and *unlike* both breaking vases and gambling devices the machinery constrains the dynamics to be totally predictable and repeatable (provided that the external environment and initial conditions do not change). It may be that these features are essential for reliable long term memory stores and intricate intelligent behaviour. Some people think the digital nature of such systems is a disadvantage, compared with neural nets allowing continuous variation of weights or excitation levels.

5 When are causal links mathematically expressible?

Both the electrical resistance of a wire and the fragility of the vase are features of the causal power of the object to "transduce" elements from one set of possibilities to another. In the former case we can express the relationship in a simple and precise mathematical formula, whereas the second case is far less rigid and precise and we do not have an appropriate mathematical representation. E.g. we have no set of mathematical descriptions of either range of possibilities (P1 and P2), even though we have an intuitive grasp of both since (a) we learn which forms of treatment to avoid so as to protect the vase and (b) we know which sorts of behaviours would be surprising and which would not, if the vase is struck, or dropped, or squeezed hard, etc. How our brains represent those classes and whether similar representations could work on computers is an interesting research question. Even if we had a precise way of characterising the two ranges of possibilities it is not clear that we could specify which element of the second range would be caused by an occurrence in the first range.

There are other cases where the causal link cannot be expressed as a simple mathematical formula linking numerical values. Computing systems are full of examples. For example, the behaviour of an interpreter for a simple programming language may be best expressed as a multi-branch conditional expression embedded in a looping construct of some sort. Here the range of input possibilities and the range of output possibilities may be expressible using languages with well defined grammars, and the transduction capability can be formally expressed in terms of rules for transforming input "sentences" into output "sentences".

In more complex cases the computer also stores records of what it has done, and these records can change future behaviour, in arbitrarily complex ways, for example if some inputs are programs that are compiled into internal stored instructions that can be invoked by later inputs. Organisms that learn also keep changing the affordances that some components offer to others.

From our present viewpoint these are just special cases of the general fact that as objects realise their possibility transducing potential they can change their future powers, for instance, by wearing out, becoming more flexible, storing information, losing energy, gaining energy, hysteresis, and so on.

6 Combining linked sets of possibilities

So far I have described, albeit in a sketchy and shallow way, some cases where we naturally see an object as having a collection of properties, where those properties are possibility transducers. Further work is needed to classify different cases, e.g. according to the structure of the set of possibilities, and according to the form of the relationships constraining possibilities.

One of the uses of a grasp of possibilities is in thinking about designs for new objects. This requires understanding the implications of combining objects into more complex objects. There are many instances of this in everyday life, some of them very simple, others more complex. Simple examples involve mere agglomerations with cumulative influences.

For example, a sheet laid on the ground has many possible motions, some of which will be produced by gusts of wind of various types. By placing a stone on each corner we can transform the links between possible gusts and possible motions, essentially by reducing the set of gusts that will result in the sheet blowing away: only the stronger gusts will produce that effect. By adding stones we can reduce the set even further. There are many examples that are very similar in form to this. If I stand against a wall with shelving containing books or other articles there is a range of postures I can take up that will enable me to reach objects on the shelves. If the shelving is high, many objects will be out of reach. By placing a block of wood on the floor so that I can stand on it, I can extend the range of possibilities. By adding further blocks, stacked vertically I can continue to extend the range of possibilities (though the blocks may also eliminate a subset by obstructing access to objects on lower shelves).

Combining objects into more complex wholes does not always have an additive effect on sets of possibilities. An interesting example is the case of two identical electrically conducting wires W1 and W2, which can either be joined at one end of each to make a new conductor of double the length of the original (joined in series) or joined at both ends to make a conductor of the same length (joined in parallel). Each piece of wire initially has a collection of properties, each linking sets of possibilities. But they are separate sets: possible currents in W1 are different things from possible currents in W2. Normally, the properties of W1 do not constrain the possibilities associated with W2, only the possibilities associated with W1. However if wires W1 and W2 are joined in series their currents are forced to be the same (at least in standard conditions). Similarly if their ends are joined in parallel, the voltages across them are constrained to be the same. Moreover in those two cases there are new relationships linking possible voltages to possible currents: i.e. the combination has a different electrical resistance from either of its components, and the resistance will depend on the form of combination. More complex interactions occur when two conductors are adjacent and the current is changing in one, or one is moving relative to the other. These are facts that can be used in the design of dynamos and electric motors.

6.1 Towards a theory of design?

All this suggests a further research topic: investigation of forms of combination of objects into more complex objects and the effects of different sorts of combinations on the possibility-transducing powers of the new structures. Particular subsets of cases are to be found in books on circuit design, and the design of various sorts of mechanical structures, e.g. linkages which allow parts to move relative to one another but with constraints on the possibilities. Car boot lids and engine bonnets are familiar examples.

Many physical mechanisms, including both mechanical devices and electronic machines, both digital and analog, consist of collections of physical objects each of which has properties that associate different ranges of properties and limit combinations of possibilities from those ranges. We have discovered how to make the “output possibilities” of one physical object the “input possibilities” for another, and to construct networks of such such possibility transducers in such a way that the complete configuration does something useful, which might not happen naturally.

That is the task of a designer, and there are many very sophisticated designers, assembling fragments of various kinds into larger structures, including cars, gas turbines, aeroplanes, houses, football teams, commercial organisations and software systems. Walls and doors of houses link and limit very complex networks of possibilities: including transmission of sounds, of heat, of temperature, of people, of furniture, and, with the help of glue and nails, also constrain motion of pictures, wallpaper, etc. They also link and limit sets of possible psychological states of occupants.

In an old fashioned clock all the possibilities involve clearly visible state changes, such as rotation of wheels, downward movements of weights, oscillation of pendulums and ratchets, etc. In an electronic clock with a digital display the collections of interlinked possibilities are of a very different kind, and the laws associated with the various possibility transducers are very different, involving discrete rather than continuous changes, for example. But the general principle is the same in both cases: a complex machine is created from simple components by linking them in such a way that the associated possibilities are also linked, and thus we get large scale constrained behaviour that is useful to us, and emergent capabilities such as telling the time and providing shelter.

It seems that some people and some animals can only grasp the *perceivable* possibilities (Gibson’s affordances) and can understand a complex design only insofar as it links together perceivable structures. However, good human designers can also think about more abstract cases including both unobservable physical structures (chemical structures, sub-microscopic digital circuits) and also abstract structures in virtual machines. Are very different forms of representation used in these two sorts of cases?

7 Damage vs change: more and less remote possibilities.

We have talked about complex systems where some combinations of the possibilities associated with components are permitted and others ruled out, at least while the configuration is preserved. Breaking a machine in some way destroys the configuration and allows some new possibilities to come into existence, and perhaps removes others. Can we make a principled distinction between changes that should be regarded as damage and changes that preserve the integrity of the system?

This depends on whether we can identify certain sorts of objects as having some sort of self-maintaining capability. Holland (1995) discusses many cases, including individual organisms, cities and social systems. For now it is not important for us whether there is a well defined set of cases. There may be a range of different sorts of objects with different combinations of capabilities, including for example ocean waves, tornados, the planetary system, as well as plants, animals, ecosystems, and many sorts of machines, all of which maintain some sort of coherence over time despite considerable physical changes and despite external perturbances and internal development. These systems exhibit different sorts of stability depending on the variety and intensity of disturbing influences that they can resist.

Where parts cooperate to maintain global relations we can talk of those parts having a *function*, whether they were designed for that function or evolved biologically, or not. Many possible states and processes are consistent with the normal state of such a system, but when a change irreversibly interferes with one or more functions we may speak of "damage". The damaged system has a different set of inherent possibilities. In the normal undamaged configuration both the old and the new sets of possibilities exist. However some sets are "more remote" than others: they cannot be realised without a change in the configuration, whereas the "less remote" possibilities are all able to be actualised without changing the configuration and without damage. A less remote possibility might include a lever changing its orientation. A more remote possibility might include the lever breaking.

The notion of an undamaged configuration needs to be made more precise. In the case of an artefact there will be a set of possible states that will be described as undamaged states of the machine and others that are described as damaged states, the difference being related to the intended use of the machine. For now I wish to disregard such purposive notions and consider only classes of states that are identifiable by their physical properties.

Instead of talking about damage we can describe a configuration as preserved as long as the components continue to have certain properties and stand in certain relationships. Thus, for a clock, rotation of the hands and the cogwheels to which they are connected preserves a certain sort of configuration, whereas removal of a tooth from a cog does not.

We could clamp two parts of the clock together and define a new configuration as one that includes all the previous relationships and also the clamping relationship. The new configuration will directly support a smaller range of possibilities. Some of the possibilities that were directly supported by the previous configuration are remote relative to the new configuration. However the restriction may also enable new possibilities, such as the unattended clock remaining wound up for a long time.

A configuration then is defined not just by its physical components, but by a particular set of properties and relationships between those components (e.g. those in which all components of the clock retain their shape, their pivot points, the distances between pivot points, etc.).

I shall say that the configuration *directly* supports the collection of possibilities that require no change in the configuration (e.g. no change of shape of the rigid components) and *indirectly* supports larger classes of possibilities, which would require more or less drastic changes in the configuration. We could define a technical notion of "damage" to a configuration as removal of one of its defining relationships. Then achieving one of the more remote possibilities requires damage to the configuration.

What sorts of configurations are worth considering as having "defining" properties is a large question I shall not consider. Living organisms and machines designed by intelligent agents would be obvious examples. Others might be cities, rampaging crowds, galaxies, the solar system, stable complex molecules, tornados, clouds and other naturally occurring complex behaving structures with some degree of integration or coherence. For now I shall merely assume there are some interesting cases without attempting to delimit the whole class.

8 Do we need new forms of representation? De-dicto and de-re modality

So far I have used informal language to talk about possibilities. Can we represent them precisely so that a machine can manipulate them? Modal logic might look like an option. It is concerned with operators that can be applied to sentences to produce new sentences, e.g. "It is possible that P", "It is necessarily the case that P". However, in ordinary language, the adjectives "possible" and "impossible" can be applied directly to objects, events or processes, and not only to propositions. This "de-re" form of words might be construed as an abbreviation for an assertion in which a modal operator is applied to a complete proposition (de-dicto modality).

We should at least consider the alternative hypothesis that there is a more basic notion of possibility than possible truth or falsity of a proposition, namely a property of objects, events or processes (de-re modality). I am making both an ontological claim about the nature of reality, and an epistemological claim about our information processing capabilities. For all the sorts of reasons given above, it seems that de-re modality plays an important role in our perception (e.g. perceiving affordances), thinking and communication, that it is used by animals that cannot construct and manipulate complete propositions, and that it will figure in the internal processes of intelligent robots. However, this remains a *conjecture* open to empirical refutation.

When Gibson claimed that the primary function of biological perceptual systems is to provide information about positive and negative affordances, such as support, obstruction, graspability, etc. I don't think he meant that animal visual systems produce propositions and apply modal operators to them. Is there a different form of representation which captures in a more powerful fashion information which enables the animal to interact fruitfully with the environment?

8.1 Possible world semantics for modal terms

Readers familiar with possible world semantics for modal operators involving notions of degrees of "accessibility" between possible worlds will see obvious links with the notion of more or less remotely supported sets of possibilities. However there is no simple ordering associated with a degree of remoteness of possibilities supported by a configuration.

For example the sets of possibilities that become accessible when one of the levers is broken, when some of the teeth are broken off a cog wheel, when a belt is removed from a pulley need not form a set ordered by inclusion. If we consider different *combinations* of kinds of damage or other change to the configuration, we get a very wide variety of sets of possibilities with at most a partial ordering in degree of remoteness from the set of possibilities directly supported by the original configuration. If there are different sequences of kinds of damage leading to the same state there need not even be a well defined partial ordering. In one sequence getting to state A requires more changes than getting to state B. In another sequence it could be the other way round.

Thus there need not be any well-defined ordering or metric that can be applied to notions of relative degree of remoteness of possibilities, and these ideas will not correspond exactly to a possible world semantics which requires degrees of accessibility to be totally or partially ordered. (Where there is no such requirement the two views of modality may turn out to be equivalent.)

There is a deeper problem about sticking to modal logic and current semantic theories, namely that we have no reason to presuppose that an adequate ontology for science or engineering would have the form of a model for a logical system, with a well defined set of objects, properties and relationships, so that all configurations can be completely described in the language of predicate calculus, and all changes can be described in terms of successions of logical formulae.

For example the wings of a hovering humming bird or a gliding seagull are very complex flexible structures whose shape is constantly changing in such a way as to change the transduction relationships between various possible forces (gravity, wind pressure, air resistance, etc.) and possible changes of location, orientation and speed. Is there some useful decomposition of the wings into a fixed set of objects whose properties and relationships can be described completely in predicate calculus (augmented if necessary with partial differential equations linking properties, etc.) Or do we need a different more "fluid" sort of representation for the dynamics of such flexible structures. An attempt to represent changing fields of influence diagrammatically can be found in Lundell (1996).

When a child learns how to dress itself can its understanding of the processes of putting on a sweater or tying shoe laces can be expressed in terms some articulation of reality into a form that is expressible in terms of names of individual objects (e.g. well defined parts of the sweater besides the obvious global parts such as sleeves, head opening, etc.), predicates, relations and functions? Or is some totally different form of representation still waiting to be discovered that is required for such tasks (Sloman 1989). Could the neural mechanisms in animal brains teach us about new powerful forms of representation.

9 Possibilities causes and counterfactual conditionals

Physical properties such as electrical resistance, tensile strength, fragility, flexibility, rigidity, all involve relationships between ranges of possibilities, as described above. In some cases the range of possibilities forms a linear continuum (e.g. possible voltages, possible currents) and in those cases the property linking them imposes a constraint that may be expressible in a particularly simple form, such as an equation linking two or more numerical measures. Other sets of possibilities may have more complex structures.

For example, a fragile vase may be struck or crushed in many different ways, and the resulting decomposition into myriad fragments can happen in many different ways. Here the two ranges of possibilities do not form linear sets, and therefore the links between them cannot be expressed as a single simple equation linking two numerical variables. In the case of a digital circuit or a software system there may be ranges of possibilities, both for inputs and outputs, that are not even continuous, for example possible inputs and outputs for a compiler, or possible startup files for an operating system and possible subsequent behaviours.

Despite the diversity, in all the cases there is a relationship between the sets of possibilities which can at least loosely be characterised by saying that the properties of the machine or configuration ensure that IF certain possibilities occur THEN certain others occur. Moreover some of these possibilities may not actually occur: the fragile vase sits in the museum forever without being hit by a sledge hammer. We can still say that the IF-THEN relationship holds for that vase. In that case we are talking about counterfactual conditionals. However, there is no difference in meaning between counterfactual conditions with false antecedent and other conditionals with true antecedent. In both cases the conditionals assert that there is a relationship between elements of certain sets of possibilities. I.e. the truth of a conditional and the existence of a possibility transducer may come to the same thing.

When we say that one thing causes another, as opposed to merely occurring prior to it we are referring implicitly to a set of such conditionals. However, this is a large topic on which huge amounts have been written (e.g. Taylor 1992) and I do not intend to address the problem in this paper, merely to point out the connection. If the previous point about expressions in predicate logic being insufficient to express all the types of possibilities that an intelligent agent needs to perceive and understand is correct, then the notion of a conditional has to be construed as something that does not merely link assertions in familiar logical formalisms. We may have to generalise it to cover new types of representations. (Some feed-forward neural nets can already be seen as an example of something different: namely a large collection of of parallel probabilistic IF-THEN rules. (Sloman 1994))

10 Levels of virtual machines

I previously remarked that a property like electrical resistance may be *implemented* in lower level physical structures, whose properties and relationships would be defined in the language of contemporary physics, or perhaps physics of the future. This is a simple example of a very general phenomenon. There are many properties of complex systems that are best thought of as "implemented" in other properties. This is commonplace in computer science and software engineering, where we often talk about virtual machines, as in the wordprocessor example. Similarly, phenomena typically studied by biologists (including genes) are implemented in mechanisms involving physics and chemistry, and phenomena studied by chemists (e.g. in drug companies) are implemented in mechanisms of quantum mechanics which the chemists may not understand. Social, political, and economic phenomena are also implemented in very complex ways in the physical world, but with many levels of intermediate virtual machines.

The points made previously about physical objects having properties which are essentially causal linkages between ranges of possibilities apply also to objects in virtual machines, including social mechanisms and abstract or virtual machines running on computers. In other words, it is not only the "bottom" level physical constituents of the universe, whatever they may be, that have causal powers: all sorts of more or less concrete objects also do. In fact a great deal of our normal thinking and planning would be completely impossible but for this fact, for we think about causal connections (e.g. connections between poverty and crime) without having any idea about the detailed physical implementation. Sometimes without knowing the details we can still interact very effectively with relevant bits of the world, like engineers thinking only about current and voltage but not the underlying quantum mechanics. As the vase and gambling machines show, we don't even need to presuppose complete causal determinism.

Admittedly social and economic engineering are very difficult, but we constantly interact with and bring about changes in people and their thought processes, desires, intentions and plans. Similar causal interactions can happen between coexisting abstract states entirely *within* a brain or an office information system.

11 Causal powers of computational states

An important implication of all this is that whereas many people have the intuitive feeling that somehow computational processes are incapable of having the properties required for mental states, events and processes, such as desires, pains, pleasures, emotions, experiences of colour, etc. this may be because they think of computations as if they were simply static structures like collections of symbols on a sheet of paper.

However, when symbolic information structures are implemented in a working system which is actually controlling some complex physical configuration, such as an airliner coming in to land, the limbs of a robot, or the machinery in a chemical plant, then it is a crucial fact about the system that the information processing states and events and the abstract data-structures, all have causal powers both to act on one another and also (through appropriate input and output devices) to change, and be changed by, gross physical structures in the environment. A further development of this line of thought would show how to defend AI theories of mind against charges that no computational system could have the right causal powers to support mental states and processes. But that is a topic for another occasion.

12 Grammars for things and for processes.

Can we move towards a scientific theory of sets of possibilities and how they are related, with sufficient precision to provide a basis for designing robots which perceive and think about possibilities? I suspect we are not yet ready to complete this task, though much work in AI can be seen as addressing it. (Recent examples are Chittaro & Ranon (1996), Lundell (1996), Stahovich et al. (1996) as well as much work on constraint manipulation).

Perhaps the notion of a formal grammar will turn out to be relevant. Grammars are specifications of sets of possibilities: usually sets of legal formulas within a formalism. However various attempts have been made to generalise this notion to accommodate, for example, grammars for images, grammars for 3-D structures, and grammars for behaviours (e.g. dances). I am not aware of any grammatical formalism that is able to cope with some of the kinds of continuous variability mentioned above (e.g. changes of configuration as a child dons a sweater.) Nevertheless it may be that some future development of some existing grammar formalism will suffice, perhaps combined with techniques for constraint propagation.

The main point for now is that the components of the grammar do not need to be propositions or constituents of propositions. So a grammar provides a different view of ranges of possibilities from that provided by a modal logic. Roughly one is object-centred (de-re) and the other fact-centred, or proposition-centred (de-dicto). However it is not clear whether these are trivially equivalent.

13 Conclusion

Many common sense and scientific notions of things are inherently modal (i.e. to do with possibilities and relationships between possibilities), including both explicitly dispositional concepts (e.g. "brittleness", "risk", "irritability"), and many others (e.g. "electrical resistance", "volume", "shape").

I have offered a view of objects (both physical objects and more abstract objects like data-structures or procedures in a virtual machine) as having properties that are inherently connected with sets of possibilities, some of the possibilities being causal inputs to the object and some outputs, and I have suggested that many of the important properties of the objects are concerned with the relationships between possibilities in different sets, i.e. causal links between possibilities. These properties are often *implemented* in lower level properties of different kinds. Moreover, by combining them in larger configurations we can use them to *implement* higher level "emergent" machines, producing many layers of implementation.

In some virtual machines we find causal powers linking events in ways that might later provide detailed models of how human mental processes work. Similar but simpler cases already exist in software systems.

One consequence of this way of thinking is that we don't have to go to quantum mechanics to be faced with issues concerning collections of coexisting possibilities from which reality makes selections.

Acknowledgements

I have benefited from interactions with many people including Natasha Alechina, Darryl Davis, Pat Hayes, Michael Jampel, Brian Logan, Mark Ryan, Henry Stapp, Bill Robinson, Toby Walsh, Dave Waltz, Ian Wright, Bill Woods, and the anonymous referees. My thinking on these topics is deeply influenced by the work of Gilbert Ryle. The ideas presented here overlap considerably with those in Bhaskar (1978). Chapter 2 of Sloman (1978) contains an early attempt of my own to address these issues.

References

R Bhaskar (1978) *A Realist Theory of Science* Sussex:The Harvester Press

L. Chittaro & R Ranon (1996) Augmenting the diagnostic power of flow-based approaches to functional reasoning, in *Proc 13th National Conference on AI (AAAI96)* Portland, Oregon, 1010–1015, AAAI Press/MIT Press

J.J. Gibson, (1986) *The Ecological Approach to Visual Perception*, Lawrence Earlbaum Associates, 1986 (originally published in 1979).

J. Holland (1995) em Hidden order: How adaptation builds complexity Reading, Mass: Addison Wesley

M. Lundell (1996) A qualitative model of physical fields, in *Proc 13th National Conference on AI (AAAI96)* Portland, Oregon, 1016–1021, AAAI Press/MIT Press

L. Pryor (1969) Opportunity recognition in complex environments, in *Proc 13th National Conference on AI (AAAI96)* Portland, Oregon, 1147–1152, AAAI Press/MIT Press

G. Ryle (1949) *The Concept of Mind,* Hutchinson.

A. Sloman (1978) *The Computer Revolution in Philosophy: Philosophy Science and Models of Mind* Hassocks: Harvester Press 1978.

A. Sloman (1989) 'On designing a visual system: Towards a Gibsonian computational model of vision' *Journal of Experimental and Theoretical AI* **1**(4), 289–337

A. Sloman (1994), Semantics in an intelligent control system, in *Philosophical Transactions of the Royal Society: Physical Sciences and Engineering*, **349**(1689), 43–58.

T. F. Stahovich, R. Davis & H. Shrobe (1996) Generating multiple new designs from a sketch, in *Proc 13th National Conference on AI (AAAI96)* Portland, Oregon, 1022–1029, AAAI Press/MIT Press

C.N. Taylor (1992) *A Formal Logical Analysis of Causal Relations* DPhil Thesis, Sussex University. Available as Cognitive Science Research Paper No.257