What's information, for an organism or intelligent machine? How can a machine or organism mean?

Aaron Sloman

University of Birmingham
`http://www.cs.bham.ac.uk/~axs`

January 23, 2011

Abstract

Words and phrases referring to information are now used in many scientific and non-scientific academic disciplines and in many forms of engineering. This chapter suggests that this is a result of increasingly wide-spread, though often implicit, acknowledgement that besides *matter* and *energy* the universe contains *information* (including information about matter, energy and information) and many of the things that happen, including especially happenings produced by living organisms, and more recently processes in computers, involve information-processing. It is argued that the concept "information" can no more be defined explicitly in terms of simpler concepts than any of the other deep theoretical concepts of science can, including "matter" and "energy". Instead the meanings of the words and phrases referring to such things are defined *implicitly* in part by the structure of the theories in which they occur, and in part by the way those theories are tested and used in practical applications. This is true of all deep theoretical concepts of science. It can also be argued that many of the pre-scientific concepts developed by humans (including very young humans) in the process of coming to understand their environment are also implicitly defined by their role in the theories being developed. A similar claim can be made about other intelligent animals, and future robots. An outline of a theory about the processes and mechanisms various kinds of information can be involved in is presented as partial implicit definition of "information". However there is still much work to be done including investigation of varieties of information processing in organisms and other machines.

Contents

1 Introduction

The question “What is information?”, like “What is matter?” and “What is energy?”, cannot have a simple answer in the form of a non-circular definition. Answering such a question involves answering a host of related questions. Answers to the second and third cannot be given without presenting deep and complex theories about how the physical universe works. The theories, along with links to experimental methods, instruments and observation techniques, provide the only kind of definition possible for many of the concepts used in the physical sciences: implicit definition. Moreover, the answers are always subject to the possibility of being revised or extended, as the history of physics shows clearly: old concepts may be gradually transformed as the theories in which they are embedded are expanded and modified – sometimes with major discontinuities, as happened to concepts like “matter”, “energy” and “force” in the work of Newton and Einstein, for example. Lesser transformations go with improved instruments and techniques for observation, measurement, and testing predictions. So concepts, have a continuing identity through many changes, like rivers, growing organisms, nations, and many other things (Cohen, 1962; Schurz, 2009).

1.1 The need for a theory

“Information” (in its oldest, and still growing, use), is another such concept. So answering the question “What is information?” will require developing a deep and complex theory of how parts of the universe that use or interact with information work, for instance entities (information users) that do various things with information: acquiring, manipulating, combining, deriving, storing, retrieving, comparing, analysing, interpreting, explaining, indexing, annotating, communicating, and above all *using* information for practical purposes. Information cannot play a role in any process unless there is something that encodes or expresses the information: an “information bearer” (B), and some user (U) that takes B to express information I (i.e. interprets B). The same bearer B may be interpreted differently by different users, and the same user, U may interpret B differently in different contexts (C). We need a theory that explains the different ways in which *a bearer B can express information I for U in context C*, and what that means. I shall henceforth use “representation” to refer to any kind of information bearer, and will later criticise some alternative definitions, in Section 2.3.

Such a theory will have to mention different kinds of information-users and information-bearers (physical and non-physical), as well as different kinds of information content, and the different ways information-bearers can be related to the information they carry, often requiring several layers of interpretation, as we’ll see. The theory will also have to survey varieties of information users, with different sorts of information processing architectures, interacting with different sorts of environment, using information-bearers (representations) that have different structures, and use different media (physical and non-physical).

Questions to be addressed include: What are the requirements for U to treat B as expressing a meaning or referring to something? What are the differences between things that merely manipulate symbolic structures and things that also understand and make use of information they associate with those structures, for example, deriving new information from them, or testing the information for consistency? Compare Searle (1980).

1.2 Is biological information-processing special?

Many of the questions have a biological context. In what ways do organisms acquire, store, extract, derive, combine, analyse, manipulate, transform, interpret, transmit, and use information? Which of these are, or could be, replicated in non-biological machines? If not all of them, then why not? Is there something special common to all forms of biological information processing?

1.3 Questions seeking answers

More general questions of a more philosophical kind that need to be answered include, whether "information" is as important a concept for science as "matter" and "energy", or just a term that is bandied about, with changing meanings, by undisciplined thinkers? Is it reasonable to think of the universe as containing matter, energy and information, with interdependencies between all three, or is there only matter and energy, in various static and changing configurations?

Why is a simple explicit definition for "information" impossible? Is it like some older scientific concepts, not explicitly definable, but implicitly definable by developing powerful explanatory theories that use the concept? Is information something that should be measurable as energy and mass are, or are its features mainly structures to be *described* not *measured* (e.g. the structure of this sentence, the structure of a molecule, the structure of an organism)? How does this (centuries old) notion of information (or meaning) relate to the more recent concept of information as something measurable? (Shannon, 1948) Are there conservation laws for information, or is that idea refuted by the fact that one user can give information to another without losing any? Moreover, it is even possible for me to say something that gives you information I did not have. (Compare the role of relay switches in electrical power circuits.)

This document attempts to give partial answers to these questions, and to specify requirements for more complete answers. I shall attempt to sum up what I think many scientists and engineers in many disciplines, and also historians, journalists, and lay people, are talking about when they talk about information, as they increasingly do, even though they don't realise precisely what they are doing. For example the idea of information pervades many excellent books about infant development, such as (E. J. Gibson & Pick, 2000), without being explicitly defined. I shall try to explain how a good scientific theory can implicitly define its main theoretical concepts, and will sketch some of the main features of a theory of the role of information in our universe. A complete theory would require many volumes. In several other papers and presentations cited below, I have presented some of these ideas in more detail.

2 Uses of the word "information"

2.1 Confusions

Unfortunately, there are many confusions about both the *content* of the notion of "information" (what we mean by it, how it should be defined, whether it can be given any precise definition) and its *status* (e.g. as a theoretical term in scientific theories, or as a loose and ill-defined, though currently fashionable, concept). The word may be a source of so much confusion that a better one is needed, but it is too late to propose a replacement, and there is no obvious candidate. “Meaning” is just as bad, or worse, since it often refers to an intention (what did you mean to do?), or the importance of some event or object (what’s the meaning of the election result?), whereas information does not have to be the content of anyone’s intention, and can be devoid of importance.

Some philosophers talk about “propositional content” but the normal interpretation of that phrase rules out information expressed in non-propositional forms, such as the information in pictures, maps, videos, gestures, and perceptual systems. So I shall stick to the label “information”, and attempt to explain how it is used in many everyday contexts and also in scientific (e.g. biological) contexts. The word is also used in this sense in engineering, in addition to being used in Shannon’s sense, discussed further in Section 2.2.

The phrase “semantic information” is as pleonastic as the phrase “young youths”, since information, in the sense under discussion, is semantic. It is sometimes useful to contrast syntactic information with semantic information, where the former is about the form or structure of something that conveys information, whereas the semantic information would be about the content of what is said. (“Content” is metaphorical here.) For instance, saying that my sentences often have more than eight words gives syntactic information about my habits, whereas saying that I often discuss evolution or that what I say is ambiguous or unoriginal gives semantic information, or, in the latter case, meta-semantic information. Likewise, we provide syntactic information about a programming language (e.g. how it uses parentheses) or semantic information (e.g. about the kinds of structure and transformations of structure that it can denote). We can distinguish the “internal” semantics of a programming language (the internal structures and processes the programs specify) from its “external” semantics, e.g. its relevance to a robot’s environment, or to a company’s employees, salaries, jobs, sales, etc.

2.2 This is not “information” in Shannon’s sense

There is another, more recent, use of the word “information” in the context of Shannon’s “information theory” (Shannon, 1948). But that does not refer to what is normally meant by “information” (the topic of this paper), since Shannon’s information is a purely syntactic property of something like a bit-string, or other structure that might be transmitted from a sender to a receiver using a mechanism with a fixed repertoire of possible messages. If a communication channel can carry N bits then each string transmitted makes a selection from 2^N possible strings. The larger N is, the more alternative possibilities are excluded by each string actually received. In that syntactic sense longer strings carry more “information”. Likewise the information capacity of a communication channel can be measured in terms of the number of bits it can transfer in parallel, and the measure can be modified to take account of noise, etc. Shannon was perfectly aware of all this. He wrote

> “The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. *These semantic aspects of communication are irrelevant to the engineering problem.*” (Shannon, 1948). [My emphasis.]

It is worth noting that although he is talking about an engineering problem of reproducing a message exactly, doing that is not what most human communication is about. If you ask me a question, my answer may fill a gap in your information, allowing you to make inferences that

I could not make. Both of us may know that, and that could be the intention of my answer. On a noisy phone line that could happen if you knew in advance that the answer was either "elephant" or "fly". If I say "fly" and you hear "spy", the fact that my precise message was not transmitted accurately does not matter: you can tell that I did not say "elephant", and proceed accordingly. A pupil's questions or comments may give a teacher information that the pupil would not understand, e.g. about how to continue a lesson. So communication in intelligent systems depends on, but is far more than, mere signal transmission. It also uses context, general knowledge of the world, more or less sophisticated interpretation mechanisms, and reasoning capabilities. Shannon's work is summarised, with strong warnings about extending it beyond the context of electromechanical signal transmission in (Ritchie, 1986).

Having a measurable amount of information in Shannon's sense does not, in itself, allow a string to express something true or false, or to contradict or imply something else in the ordinary senses of "contradict" or "imply", or to express a question or command. Of course, a bit string used in a particular context could have these functions. E.g. a single bit could express a "yes" or "no" answer to a previously asked question, as could a "continue" or "stop" command. In some contexts, that single bit may indirectly convey a great deal of information. "Is everything Fred wrote in his letter true?" "Yes."

2.3 Misguided definitions

Bateson (1972) describes "a bit of information" and later "the elementary unit of information" as "a difference that makes a difference".[1] This is widely misquoted as offering a definition of "information" rather than "a bit/unit of information". He seems to be thinking of any item of information as essentially a collection of "differences" that are propagated along channels. This is far too simplistic – and perhaps too influenced by low level descriptions of computers and brains. An alternative approach is to define "information" implicitly by a complete theory, as happens for many scientific concepts. This paper attempts to present substantial portions of such a theory, though the task is not completed. Section 3.2 explains how theories can implicitly define the concepts they use and 6.1 relates this to defining "information".

What it means for B to express I for U in context C cannot be given any simple definition. Some people try to define this by saying U uses B to "stand for" or "stand in for" I. For instance, Webb writes "The term 'representation' is used in many senses, but is generally understood as a process in which something is used to stand in for something else, as in the use of the symbol 'I' to stand for the author of this article" (Webb, 2006). This sort of definition of "representation" is either circular, if standing for is the same thing as referring to, or else false, if standing in for means "being used in place of". There are all sorts of things you can do with information that you would never do with what it refers to and vice versa. You can eat food, but not information about food. Even if you choose to eat a piece of paper on which "food" is written that is usually irrelevant to your use of the word to refer to food. Information about X is normally used for quite different purposes from the purposes for which X is used. For example, the information can be used for drawing inferences, specifying something to be prevented, or constructed, and many more. Information about a possible disaster can be very useful and therefore desirable, unlike the disaster itself.

So the notion of standing for, or standing in for is the wrong notion to use to explain information content. It is a very bad metaphor, even though its use is very common. We

---

[1] In at least two of the essays "The Cybernetics of 'Self': A Theory of Alcoholism" and in "Form Substance and Difference".

can make more progress by considering ways in which information can be used. If I give you the information that wet whether is approaching, you cannot use the information to wet anything. But you can use it to decide to take an umbrella when you go out, or, if you are a farmer you may use it as a reason for accelerating harvesting. The falling rain cannot so be used: by the time the rain is available it is too late to save the crops.

The same information can be used in different ways in different contexts or at different times. The relationship between information content and information use is not a simple one.

2.4 The world is NOT the best representation of itself

In recent years, an erroneous claim, related to confusing representing with standing in for, has found favour with many, namely the claim that "the world is its own best representation".

Herbert Simon pointed out long ago (Simon, 1969) that sometimes the changes made to the environment while performing a task can serve as reminders or triggers regarding what has to be done next, giving examples from insect behaviours. The use of stigmergy, e.g. leaving tracks or pheromone trails or other indications of travel, which can later be used by other individuals, shows how sometimes changes made to the environment can be useful as means of sharing information with others. Similarly if you cannot be sure whether a chair will fit through a doorway you can try pushing it through, and if it is too large you will fail, or you may discover that it can go through only if it is rotated in some complex way.

The fact that intelligent agents can use the environment as a store of information or as a source of information or as part of a mechanism for reasoning or inferring, does not support the slogan that the world, or any part of it, is always, or even in those cases the best representation of itself (a) because the slogan omits the role of the information-processing in the agent making use of the environment and (b) because it sometimes is better to have specific instructions, a map, a blue-print or some other information structure that decomposes information in a usable way, than to have to use the portion of the world represented, as anyone learning to play the violin simply by watching a violinist will discover.

In general, information about X is something different from X itself. Reasons for wanting or for using information about X are different from the reasons for wanting or using X. E.g. you may wish to use information about X in order to ensure that you never get anywhere near X if X is something dangerous. You may wish to use information about Xs to destroy Xs, but if that destroyed the information you would not know how to destroy the next one until you are close to it. It may then be too late to take necessary precautions, about which you had lost information.

Dreyfus (2002) wrote "The idea of an intentional arc is meant to capture the idea that all past experience is projected back into the world. The best representation of the world is thus the world itself." As far as I can make out he is merely talking about expert servo control, e.g. the kind of visual servoing which I discussed in (Sloman, 1982). But as any roboticist knows, and his own discussion suggests, this kind of continuous action using sensory feedback requires quite sophisticated *internal* information processing (Grush, 2004). In such cases "the world" is not nearly enough.

2.5 Disagreements about information bearers, representations

Brooks also wrote a series of papers attacking symbolic AI, including (Brooks, 1991). He repeatedly emphasises the need to test working systems on the real world and not only in simulation, a point that has some validity but can be over-stressed. (If aircraft designers find it useful to test their designs in simulation, why not robot designers?) Moreover, he disputes the need for representations (information bearers constructed and manipulated by information users), saying: “We hypothesize (following Agre and Chapman) that much of even human level activity is similarly a reflection of the world through very simple mechanisms without detailed representations,” and “We believe representations are not necessary and appear only in the eye or mind of the observer.” A critique of that general viewpoint is presented in (Sloman, 2009c), which mostly deals with (Brooks, 1990), in which he goes further:

> “The key observation is that the world is its own best model. It is always exactly up to date. It always contains every detail there is to be known. The trick is to sense it appropriately and often enough.”

That’s impossible when you are planning the construction of a skyscraper using a new design, or working out the best way to build a bridge across a chasm, or even working out the best way to cross a busy road, which you suspect has a pedestrian crossing out of sight around the bend. The important point is that intelligence often requires reasoning about what might be the case, or might happen, and its consequences: and that cannot be done by inspecting the world as it is. Recall that information bearers and things they represent have different uses (Section 2.3).

2.6 Computation and information

It is sometimes suggested, e.g. in Searle (1980), that computation is concerned only with syntax. That ignores the fact that even in the simplest computers bit patterns refer to locations and instructions, i.e. they have a semantic interpretation. An extreme view in the opposite direction is expressed by Denning (2009): “The great principles framework reveals that there is something even more fundamental than an algorithm: the representation. Representations convey information. A computation is an evolving representation and an algorithm is a representation of a method to control the evolution”. A position close to Denning’s will be developed here, though his view of computation (i.e. information-processing) is too narrow.

2.7 Not all information is true

Some people, for example the philosopher Fred Dretske, in his contribution to (Floridi, 2008), claim that what we ordinarily mean by “information” in the semantic sense is something that is true, implying that it is impossible to have, provide or use false information. False information, on that view can be compared with the decoy ducks used by hunters. The decoys are not really ducks though some real ducks may be deceived into treating the decoys as real – to their cost! Likewise, argues Dretske, false information is not really information, even though some people can be deceived into treating it as information. It is claimed that truth is what makes information valuable, therefore anything false would be of no value.

Whatever the merits of this terminology may be for some philosophers, the restriction of “information” to what is true is such a useless encumbrance that it would force scientists and robot designers (and philosophers like me) to invent a new word or phrase that had the same meaning as “information” but without truth being implied. For example, a phrase something like “information content” might be used to refer to the kind of thing that is common to my belief that the noise outside my window is caused by a lawn-mower, and my belief that the noise in the next room is caused by a vacuum cleaner, when the second belief is true while first belief is false because the noise outside comes from a hedge trimmer.

The observation that humans, other animals and robots, acquire, manipulate, interpret, combine, analyse, store, use, communicate, and share information, applies equally to false information and to true information, or to what could laboriously be referred to as the "information content" that can occur in false as well as true beliefs, expectations, explanations, and percepts, and moreover, can also occur in questions, goals, desires, fears, imaginings, hypotheses, where it is not known whether the information content is true. So in constructing the question "Is that noise outside caused by a lawnmower?", a speaker can use the same concepts and the same modes of composition of information as are used in formulating true beliefs like: "Lawnmowers are used to cut grass", "Lawnmowers often make a noise", "Lawnmowers are available in different sizes", as well as many questions, plans, goals, requests, etc. involving lawnmowers. Not only true propositions are valuable: all sorts of additional structures containing information are useful. Even false beliefs can be useful, because by acting on them you may learn that they are false, why they are false, and gain additional information. That's how science proceeds and much of the learning of young children depends heavily on their ability to construct information contents without being able to tell which are true and which are false. The learning process can then determine the answers. This will also be important for intelligent robots.

For the purposes of cognitive science, neuroscience, biology, AI, robotics and many varieties of engineering, it is important not to restrict the notion of "information" to what is true, or even to whole propositions that are capable of being true or false. There are information fragments of many kinds that can be combined in many ways, some, but not all, of which involve constructing propositions. Information items can be used in many other processes. The uses of information in control probably evolved before other uses of information in biological organisms, including, for example, microbes. Explaining how and why other uses evolved, such as forming memories, predictions, questions and explanations, along with increasingly sophisticated mechanisms to support them, is a task for another occasion. Some hypotheses are sketched in (Sloman, 2007a).

3 Is "information" as used here definable?

3.1 The inadequacy of explicit definitions

In order to understand how a concept like "information" can be used in science without being definable, we need to understand some general points from philosophy of science. Shannon's notion of information was defined precisely (Shannon, 1948) and has had important applications in science and engineering. Nevertheless, for reasons given above, that concept is not what we need in talking about an animal or robot that acquires and uses information about various things (the environment, its own thinking, other agents, future actions, etc.), even though Shannon's notion is relevant to some of the mechanisms underlying such processes. Can we define this older, intuitive, more widely used, notion of information?

After many years of thinking about this, I have concluded that "information" in this sense cannot be explicitly defined without circularity. The same is true of "mass", "energy" and other deep concepts used in important scientific theories. Attempts to define "Information" by writing down an *explicit* definition of the form "Information is ...." all presuppose some concept that is closely related ("meaning", "content", "reference", "description", etc.). "Information is meaning", "information is semantic content", "information is what something is about" are all inadequate in this sense.

This kind of indefinability is common in concepts needed for deep scientific theories. Attempts to get round this by "operationalising" theoretical concepts fail. For example, there are standard methods of measuring mass and energy, but those do not *define* the concepts, since the measuring methods change as technology develops, while the meanings of the words remain mostly fixed by their roles in physical theories. The measurement methods define what are sometimes called "bridging rules" or "correspondence rules", which link theories to observations and applications. Carnap (1947) called some of them "meaning postulates". All this was known to early 20th century philosophers of science, some of whom had tried unsuccessfully to show that scientific concepts are definable in terms of the sensory experiences of scientists, or in terms of "operational definitions" specifying how to detect or measure physical quantities (Bridgman, 1927).

The absence of any explicit definition does not mean either that a word is meaningless or that we cannot say anything useful about it. The specific things said about what energy is and how it relates to force, mass, electrical charge, etc., change over time as we learn more, so the concepts evolve. Newton knew about some forms of energy, but what he knew about energy is much less than what we now know about energy, e.g. that matter and energy are interconvertible, and that there are chemical and electromagnetic forms of energy. Growing theoretical knowledge extends and deepens the concepts we use in expressing that knowledge (Cohen, 1962; Schurz, 2009). That is now happening to our concept of information as we learn more about types of information-processing machine, natural and artificial.

3.2 Concepts implicitly (partially) defined by theories using them

If concepts are not all defined in terms of sensory experiences or measurement operations, how do we (including physicists) manage to understand the word "energy"? The answer seems to be: such a word mainly acquires its meaning from its role in a rich, deep, widely applicable theory in which many things are said about energy, e.g. that in any bounded portion of the universe there is a scalar (one-dimensional), discontinuously variable amount of it, that its totality is conserved, that it can be transmitted in various ways, that it can be stored in various forms, that it can be dissipated, that it flows from objects of higher to objects of lower temperatures if they are in contact, that it can be radiated across empty space, that it can produce forces that cause things to move or change their shape, etc. (All that would have to be made much more precise for a physics text book.)

If a theory is expressed logically, and is not logically inconsistent, and its undefined concept labels are treated as variables ranging over predicates, relations and functions, then there may be a non-empty set of possible models for the set of statements expressing the theory, where the notion of something being a model is illustrated by lines, points, and relations between them being a model for a set of axioms for Euclidean geometry, and also certain arithmetical entities being a model for the same axioms. This notion of model was first given a precise recursive definition by Tarski but the idea is much older, as explained in (Sloman, 2007c). I think the core idea can be generalised to theories expressed in natural language and other non-logical forms of representation including non-Fregean forms of representation, but making that idea precise and testing it are research projects (compare (Sloman, 1971)). The models that satisfy some theory with undefined terms will include possible portions of reality that the theory could describe. Insofar as there is more than one model, the meanings of the terms are partly indeterminate, an unavoidable feature of scientific theories. Sloman (1978, Chap 2) explains why it is not usually possible to completely remove indeterminacy of meaning. Compare Cohen (1962).

Adding new independent postulates using the same undefined terms will further constrain the set of possible models. That is one way to enrich the content of a theory. Another way is to add new undefined concepts and new hypotheses linking them to the old ones. That increases the complexity required of a piece of reality if it is to be a model of the theory. Other changes may alter the set of models and increase the number of things that are derivable from the theory, increasing the variety of predictions. Some changes will also increase the *precision* of the derived conclusions, e.g. specifying predicted processes or possible processes in more detail. Adding new "meaning postulates", or "bridging rules", linking undefined terms to methods of measurement or observation, as explained above, can also further constrain the set of possible models, by "tethering" (label suggested in (Chappell & Sloman, 2007)) the theory more closely to some portion of reality. As science progresses and we learn more things about energy. the concept becomes more constrained – restricting the possible models of the theory, as explained in (Sloman, 2007c). This gradual increase in understanding would not be possible if the initial concepts were fully determinate. Far from requiring absolutely precise concepts, as normally supposed, some scientific advances depend on (partial) indeterminacy of concepts.

3.3 Evaluating theories, and their concepts

For concepts that are implicitly defined by their role in the theory, the evaluation of the concepts as referring to something real or not will go along with the evaluation of the theory. How to evaluate scientific theories is itself a complex and difficult question and there are many tempting but shallow and inadequate criteria. I think the work of Lakatos extending and refining Popper's ideas (Lakatos, 1980) is of great value here, in particular insofar as it draws attention to the difficulty of evaluating or comparing theories conclusively at a point in time. Instead it often takes time before we can tell whether the research programme associated with a theory is "progressive" or "degenerating". It always remains possible for new developments to resurrect a defeated theory, as happened to the corpuscular theory of light.

Doubt is cast on the value of a theory and its concepts if the theory does not enhance our practical abilities, if it doesn't explain a variety of observed facts better than alternative theories, if all its predictions are very vague, if it never generates new research questions that lead to new discoveries of things that need to be explained, if its implications are restricted to very rare situations, and if it cannot be used in making predictions, or selecting courses of action to achieve practical goals, or in designing and steadily improving useful kinds of machinery, In such cases, the concepts implicitly defined by the theory will be limited to reference within the hypothetical world postulated by the theory. Concepts like "angel" and "fairy" are examples of such referentially unsuccessful concepts, though they be used to present myths of various sorts, providing entertainment and, in some cases, social coercion.

These ideas about concepts and theories were elaborated in Sloman (1978, Chap 2), which pointed out that the deepest advances in science are those that extend our ontology substantively, including new theories that explain possibilities not previously considered. How concepts can be partly defined implicitly by structural relations within a theory is discussed further in (Sloman, 1985, 1987). These ideas can be extended to non-logical forms of representation, as discussed in (Sloman, 2008b).

3.4 The failure of concept empiricism and symbol-grounding theory

Because a concept can be (partially) defined implicitly by its role in a powerful theory, and therefore some symbols expressing such concepts get much of their meaning from their structural relations with other symbols in the theory (including relations of derivability between formulae including those symbols) it follows that not all meaning has to come from experience of instances, as implied by the theory of concept empiricism. Concept empiricism is a very old philosophical idea, refuted by Kant (1781), and later by philosophers of science in the 20th century thinking about theoretical concepts like "electron", "gene", "neutrino", "electromagnetic field". (For more on Concept Empiricism, see: (Prinz, 2005; Machery, 2007).)

Unfortunately, the already discredited theory was recently reinvented and labelled "symbol grounding theory" (Harnad, 1990). This theory seems highly plausible to people who have not studied philosophy, so it has spread widely among AI theorists and cognitive scientists, and is probably still being taught to unsuspecting students. Section 3.2 presented "symbol tethering" theory, according to which meanings of theoretical terms are primarily determined by structural relations within a theory, supplemented by "bridging rules". Designers of intelligent robots will have to produce information-processing architectures in which such theories can be constructed, extended, tested and used, by the robots, in a process of acquiring information about the world, and themselves.

Marvin Minsky in (2005) also talks about "grounding" but in a context that neither presupposes nor supports symbol-grounding theory. He seems to be making a point I agree with, namely that insofar as complex systems like human minds monitor or control themselves the subsystem that does the monitoring and controlling needs to observe and intervene at a high level of abstraction instead of having to reason about all the low level details of the physical machine. In some cases, this can imply that the information that such a system has about itself is incomplete or misleading. I.e. self-observation is not infallible, except in the trivial sense in which a voltmeter cannot be misled about what its reading of a voltage is, as explained in (Sloman, 2007b).

The rest of this paper attempts to outline some of the main features of a theory about roles information can play in how things work in our world. The theory is still incomplete but we have already learnt a lot and there are many possible lines of development of our understanding of information processing systems in both natural and artificial systems.

4 Information-bearers, information contents.

4.1 Users, bearers, contents, contexts – physical and virtual

As explained in Section 1.1, an information-bearer B (a representation) can express information I for user U in context C. The user, U, can take B to express information about something remote, past, future, abstract (like numbers), or even non-existent, e.g. a situation prevented, or a story character.

The expressed information can be involved in many processes, for instance: acquiring, transforming, decomposing, combining with other information, interpreting, deriving, storing, inferring, asking, testing, using as a premiss, controlling internal or external behaviour, and communicating with other information-users. Such processes usually require U to deploy mechanisms that have access to B, to parts of B, and to other information-bearers (e.g. in U's memory or in the environment).

The existence of information-bearers does not depend on the existence of what they refer to: things can be referred to that do not exist. Mechanisms for this were probably a major advance in biological evolution. Example information-bearers explicitly used by humans include sentences, maps, pictures, bit-strings, video recordings, or other more abstract representations of actual or possible processes. At present little is known about the variety of information bearers in biological systems, including brains, though known examples include chemical structures and patterns of activation of neurons. In some cases the information-bearers are physical entities, e.g. marks on paper or acoustic signals, or chemicals in the blood stream. But many information-bearers in computing systems, e.g. lists of symbols, the text in a word-processor, are not physical entities but entities in virtual machines (see Section 6.3). The use of virtual machines in addition to physical machines has many benefits for designers of complex information processing systems. Sloman (2009f) argues that evolution produced animals that use virtual machines containing information bearers, for similar reasons. The problem of explaining what information is includes the problem of how information can be processed in *virtual machines*, natural or artificial. (In this context, the word "virtual" does not imply "unreal"[2].)

The bearer is a physical or virtual entity (or collection of entities) that encodes or expresses the information, for that user in that context. Many people, in many disciplines, now use the word "representation" to refer to information-bearers of various kinds, though there is no general agreement on usage. Some who argue that representations are not needed proceed to discuss alternatives that are already classified as representations by broad-minded thinkers. Such factional disputes are a waste of time.

4.2 Changing technology for information-bearers

Early general purpose electronic computers used only abstract bit-patterns as forms of representation, though the physical implementation of the bit-patterns varied. Over the years since the 1940s many more information-bearers have been developed in computers, either implemented in bit patterns, or in something else implemented in bit-patterns, e.g. strings, arrays, lists, logical expressions, algebraic expressions, images, rules, grammars, trees, graphs, artificial neural nets, and many more. These are typically constructed from various primitive entities and relationships available in virtual machines though they are all ultimately implemented in bit-patterns, which themselves are virtual entities implemented in physical machines using transistors, magnetic mechanisms in disc drives, etc. The use of such things as error-correcting memories and raid arrays implies that the bits in a bit pattern are virtual entities that do not correspond in any simple way to physical components.

This use of bit-patterns as a form of representation is relatively recent, although Morse code, which is older, is very close. Long before that, humans were using language, diagrams, gestures, maps, marks in the sand, flashing lights, etc. to express information of various kinds (Dyson, 1997). And before that animal brains used still unknown forms of representation to encode information about the environment, their motives, plans, learnt generalisations,

---

[2] As explained in and in various papers and presentations available online (Sloman, 1985, 1987, 2008b, 2008c, 2009e)

etc.(Sloman, 1979, 2008b). It is arguable that all living organisms acquire and use information, both in constructing themselves and also in controlling behaviour, repairing damage, detecting infections, etc.[3]

Information-bearers need not be *intentionally* constructed to convey information. For example, an animal may hear a sound and derive the information that something is moving nearby. The original information-bearer is a transient acoustic signal in the environment produced unintentionally by whatever moved. The hearer constructs an enduring information-bearer (representation) that may be retained long after the noise has ended. The physical signal does not *intrinsically* carry that information, though for a particular user it may do so as a result of prior learning. However, in a different context, the same noise may be interpreted differently. So the association between bearer and information content can depend not only on user but on context: information (or meaning) involves at least a *four-termed relation* involving B, I, U, and C.

4.3 A common error about bit patterns and symbols

It is sometimes claimed that in Shannon's sense "information" refers to physical properties of physical objects, structures, mechanisms. But not all bit-strings are physical. For example, it is possible to have structures in virtual machines that operate as bit-strings and are used for communication between machines, or for virtual memory systems, especially when bit-strings are transmitted across networks in forms that both use data-compression and error correcting mechanisms based on redundancy. A similar mistake was made by Newell and Simon (Newell, 1980) when they proposed that intelligent systems need to use "physical symbol systems", apparently forgetting that many symbols used in AI systems are not physical entities, but entities in virtual machines (see Section 6.3).

4.4 Many forms of representation

There are many forms in which information can be expressed. Some are very general, including logic, human languages, and various structures used in computer databases. They are not completely general insofar as there may be some things, e.g. information about irregular continuous spatial or temporal variation, that they cannot express fully. Other forms of representation are more specialised, e.g. number notations, notations for differential and integral calculus, musical notation, and various styles of maps. What characterises a form of representation is a collection of primitives, along with ways of modifying them, combining them to form larger structures, transformations that can be applied to the more complex items, mechanisms for storing, matching, searching, and copying them, and particular uses to which instances of the form can be put, e.g. controlling behaviour, searching for plans, explaining, forming generalisations, interpreting sensory input, expressing goals, expressing uncertainty, and communication with others. The representing structures may be physical objects or processes, or objects or processes in virtual machines. The use of virtual machine forms of representation allows very rapid construction and modification of structures without having to *rearrange* physical components. In computers instead of physical rearrangements there are merely banks of switches that can be turned on and off, thereby implementing

---

[3] This is discussed in a presentation arguing that there is a sense in which life presupposes mind (informed control) `http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#lifemind`

changes to virtual network topology and signals transmitted, in terms of which higher-level virtual machine representations can be implemented.

Humans often use forms that are Fregean (Sloman, 1971) insofar as they use *application of functions to arguments* to combine information items to form larger information items. Examples include sentences, algebraic expressions, logical expressions and many expressions in computer programs. Purely Fregean forms of representation use *only* function application, whereas impure forms also use spatial or temporal order, and other relationships in the bearer's medium, as Bateson (1972) noted. For example, the programming language Prolog uses ordering of symbols as well as the function-argument relationship, as significant.

The 1971 paper argued, against McCarthy and Hayes (1969), that non-Fregean forms of representation, e.g. analogical representations, are often useful, and should be used in AI alongside logic and algebra. For example, information may usefully be expressed in continuously changing levels of activation of some internal or external sensing device, in patterns of activation of many units, in geometrical or topological structures analogous to images or maps, in chemical compounds, and many more. Despite some partial successes, this has proved easier said than done.

Exactly how many different forms exist in which information can be encoded, and what their costs and benefits are, is an important question that will not be discussed further here. One of the profound consequences of developments in metamathematics, computer science, artificial intelligence, neuroscience and biology in the last century has been to stretch our understanding of the huge variety of possible forms of representation (Peterson, 1996), including some forms that are not decomposable into discrete components, as sentences, logical expressions, and bit strings are, and some which can also change continuously, unlike Fregean representations.

Besides analogical and Fregean forms of representation many others have been explored, including distributed neural representations and forms of genetic encoding. Minsky (1992) discusses tradeoffs between some symbolic and neural forms. There probably are many more forms of representation (more types of information-bearer) than we have discovered so far. Some philosophers use the misleading expression "non-conceptual content" to refer to some of the non-Fregean forms of representation – misleading because it presupposes that concepts (units of semantic content) can only be used in propositional formats. We can achieve greater generality by using the label "concept" wherever there are re-usable information components that can be combined with others in different ways whether in propositions, instructions, pictures, goal specifications, action-control signals, or anything else.[4]

Obviously, a representation may convey different information to different users, and nothing at all to some individuals (e.g. humans listening to a foreign language). Moreover, the very same information-bearer can convey different information to the same user at different times, in different contexts, for example, indexical expressions, marks in the sand, shadows, etc. (Further examples and their implications are discussed below in Section 5.9 and in (Sloman, 2006b).)

The continued investigation of the space of possible forms of representation, including the various options for forming more complex information contents from simpler ones, and the tradeoffs between the various options, is a major long term research project. This paper is

---

[4] See also the discussion of alternatives to logical representations in Sloman (1978, Chap7). Sloman (2008b) argues that non-communicative "languages" used for perception, learning, planning, etc., evolved before human languages, some of them using non-Fregean forms of representation.

mostly neutral as regards the precise forms in which information can be encoded.

4.5 “Self-documenting” entities

It is normally assumed that we cannot talk about the information expressed by or stored in a bearer B without specifying a user (or type of user) U. However, it is arguable that *any* object, event, or process is intrinsically a bearer of information about itself (a “self-documenting” entity), though not all users are equally able to acquire and use the information that is available from the entity. So a twig lying in the forest is a bearer (or potential bearer?) of information about its size, shape, physical composition, location, orientation, history, and relationships to many other things. Different information users can take in and use different subsets or impoverished forms of that information, depending on their sensory apparatus, their information processing architecture, the forms of representation they are able to use, the theories they have, and their location in relation to the twig. (Compare the notion of “intrinsic information” in (Reading, 2006).)

Besides the “categorical” information about the parts, relationships, properties, and material constitution of an object or process that can be discovered by an appropriately equipped perceiver, there is also less obvious “dispositional” information about processes it could be part of, processes that it constrains or prevents, and processes that could have produced it. These are causal relationships. Intelligent perceivers make a great deal of use of such information when they perceive affordances of various kinds. Gibson’s notion of “affordance” (J. J. Gibson, 1979) focuses on only a subset of possible processes and constraints, namely those relevant to what a perceiver can and cannot do: action-affordances for the perceiver. We need to generalise that idea if we are to describe all the different kinds of information a perceiver can use in the environment, including *proto-affordances*, concerned with which processes are and are not physically possible in the environment, *epistemic affordances*, concerned with what information is and is not available and *vicarious affordances*, concerned with affordances for other agents, all described in Sloman (2008a). Some animals are able to represent *meta-affordances*: information about ways of producing, modifying, removing, or acquiring information about, affordances of various kinds.

Information-users will typically be restricted in the kinds of information they can obtain or use, and at any time they will only process a subset of the information they could process. They will typically not make use of the majority of kinds of information potentially available. For instance, detailed, transient, metrical information about changing relationships will be relevant during performance of actions such as grasping, placing, catching or avoiding, but only more abstract information will be relevant while future actions are being planned, or while processes not caused by the perceiver are being observed (Sloman, 1982). States of an information-processing system (e.g. the mind of an animal or robot) are generally not just constituted by what is actually occurring in the system but by what would or could occur under various conditions – a point made long ago in Ryle (1949).

The information-processing mechanisms and forms of representation required for perceivers to acquire and use information about actual and possible processes and causal relationships are not yet understood. Most research on perception has ignored the problem of perceiving *processes*, and *possibilities for* and *constraints on* processes, because of excessive focus on perceiving and learning about *objects*.

5 Aspects of information

5.1 Information content and function

Items of information can have different aspects that need to be distinguished, of which three important examples are *content*, *function* (or use, or causal role) and the *medium* in which information is expressed, or represented, where each of those can be further subdivided.

It is possible for the same information content (e.g. that many parents abuse their children by indoctrinating them) to be put to different uses. E.g. it can be stated, hypothesised, denied, remembered, imagined to be the case, inferred from something, used as a premiss, used to explain, used to motivate political action, and many more. Those could all be labelled "declarative" uses of information. An item of declarative information can be true or false, and can imply, contradict, or be derived from, other items of factual information. It can also provide an answer (true or false) to a question, or a description of what needs to be achieved for an item of control information to be successful, e.g. for a command to be obeyed.

The same content can also occur in other information uses, e.g. "interrogative" and "imperative" uses: formulating requests for information and specifying an action to be performed (or modified, terminated, suspended or delayed, etc.), for instance asking whether it is the case or exhorting people to make it false by changing their ways. An important use that is hard to specify is in conditionalising some other information content, which could be a statement, intention, command, question, prediction. Examples: "If it's raining take an umbrella", "If it's raining, why aren't you wet?" There is usually no commitment regarding truth or falsity of the condition, in such uses.

Like questions, imperative uses of information are not true or false, though particular processes can be said to follow or not follow the instructions. Just as some declarative information contents are inconsistent, and therefore incapable of being true, likewise, some instructions are inconsistent, and therefore impossible to execute (e.g. "Put seven balls into an empty box and, put red marks on ten of them").

5.2 Medium used for information bearer

From the earliest days of AI and software engineering it was clear that choice of form of representation could make a large difference to the success of a particular information-processing system. Different expressive media can be used for the various functions: vocal utterances, print, internet sites, use of sign language, political songs, etc. The same content expressed in print could use different fonts, or even entirely different languages. But some information contents cannot be adequately expressed in some media, e.g. because, as J.L.Austin once quipped: "Fact is richer than diction" (Austin, 1956). Some kinds of richness are better represented in a non-Fregean medium, e.g. using static or moving images, or 3-D models.

A pre-verbal child, or a non-human animal, can have percepts whose content specifies a state of affairs in the environment; and can have intentions whose content specifies some state of affairs to be achieved, maintained or prevented. It is unlikely that toddlers, dogs, crows, and apes use only linguistic or Fregean forms of representation, though there are many unanswered questions about exactly which other forms or media are possible.

Many information-bearers use static media, like sentences, pictures, or flowcharts, whereas some use dynamic media, in which *processes* are information-bearers, e.g. audio or video recordings, gestures, play acting, and others. If the dynamic representation is repeatedly produced it may be represented by some enduring static structure that is used to generate the dynamic process as needed – e.g. a computer program can repeatedly generate processes. I suspect the role of dynamic information-bearers and static encodings of dynamic information-bearers, in animal intelligence, and future intelligent robots, will turn out to be far more important than anyone currently realises, not least because much information about the environment is concerned with processes occurring, and processes that could occur.

Earlier, in 4.5, we mentioned self-documenting entities, which potentially express information for various kinds of information user simply in virtue of their structure, properties and relations. These information bearers do not depend for their existence on users. They can be contrasted with the sensory signals and other transient and enduring information bearers constructed by information users. An element of truth in the view of Brooks criticised above (2.5) is that in some cases the presence of self-documenting entities reduces (but does not eliminate) the need for an information user to construct internal representations. Moreover, during performance of actions, force-feedback and visual feedback can be used to provide fine-grained control information that reduces the reliance on ballistic control, which may be inaccurate.

Another way of putting the point about control using feedback is that *the changing relationships* to external objects produced when performing physical actions can be useful *self-documenting* aspects of the environment, helping with control. They can also be useful for other observers (friendly or unfriendly!) who can perceive the actions and draw conclusions about the intentions and motives of the agent – if the viewers have appropriate meta-semantic information-processing capabilities. In that sense, intentional actions can serve as unintended communications, and it is conjectured in Sloman (2008b) that fact played a role in evolution of languages used intentionally.

5.3 Same content, but different function

Items of information with the same declarative content can be given different functional roles in an information user. For example, the same thing can be stated to be true and either asked about or commanded to be made or kept true. It can also be wondered about, hypothesised, imagined regretfully, treated as an ideal, etc.

The philosopher R.M. Hare (Hare, 1952) introduced the labels “Phrastic” and “Neustic” to distinguish the semantic content of an utterance and the speech act being performed regarding that content, e.g. asserting it, denying it, enquiring about its truth value, commanding that it be made true, etc. The concept of “information content” used here is close to Hare’s notion of a “Phrastic”, except that we are not restricting semantic content to what can be expressed in a linguistic or Fregean form: other media, including maps, models, diagrams, route-summaries, flow-charts, builders’ blue-prints, moving images, 3-D models, and other things, can all encode information contents usable for different functions. Moreover, not all uses are concerned with communication between individuals: information is processed in perceiving, learning, wanting, planning, remembering, deciding, etc. (Sloman, 1979, 2008b). We therefore need to generalise the Phrastic/Neustic distinction to contrast content and function in many different information media, including information expressed in diagrams, maps, charts (Sloman, 1971), and also whatever forms are used in animal brains or minds. In many cases the “neustic” is not expressed within the representation but simply by its role in an information processing architecture, as explained in Sloman (2009a), or in some aspect of the context, e.g. the word “Wanted” above a picture of a human face.

Questions, requests, commands, desires, and intentions, can all be described as examples of "control information", because their information-processing function (the *neustic* aspect), involves making something happen, unlike factual information, which, in itself, has no implications for action, although it can have implications in combination with motives, conditional plans, etc. Control information (and what should be done) is commonly found in kitchen recipes, computer programs, knitting patterns, legal documents, etc. There must be many forms implemented in animal brains.

Summing up: When information is used we can distinguish the content of the information (phrastic) from the use that is being made of it (neustic). The latter may be explicitly indicated in the medium, or implicitly determined by the subsystem of the user that the bearer is located in, or the context. We can also distinguish different information media, e.g. linguistic, Fregean, pictorial, hybrid, static, dynamic, etc. Each of these can be further subdivided in various ways, only some of which have already been explored in working artificial systems.

5.4 Processing requirements for different media

One of the achievements of AI research in the last half-century has been the study of different information media, and analysis of different information processing mechanisms required for dealing with them, including sentences, algebraic expressions, logical expressions, program texts, collections of numerical values, probability distributions, and a variety of analogical forms of representation, including pictures, diagrams, acoustic signals, and more. There are many ways in which information media can vary, imposing different demands on the mechanisms that process them.

One of the most important features of certain media is their "generativity". For example, our notations for numbers, sentences, maps, computer programs, chemical formulate, construction blue-prints, are all generative insofar as there is a subset of primitive information bearers along with ways in which those primitives can be combined to form more complex bearers, where the users have systematic ways of interpreting the complex bearers on the basis of the components and their relationships. This is referred to as a use of "compositional semantics", where meanings of wholes depend on meanings of parts and their relationships, and sometimes also the context (Sloman, 2006b).

If an organism had only six basic actions, and could only process bearers of information about complex actions made up of at most three consecutive basic actions, then it would have restricted generativity, allowing for at most 216 complex actions. Some organisms appear to have sensor arrays that provide a fixed size set of sensor values from which information about the environment at any time can be derived. In contrast, humans, and presumably several other species, do not simply record sensor values but interpret them in terms of configurations of entities and processes in the environment, e.g. visible or tangible surface fragments in various orientations changing their mutual relationships.

If the interpretation allows scale changes (e.g. because of varying distances) and sequential scanning of scenes, both of which are important in human vision, the user can construct and interpret information bearers of different kinds and degrees of complexity. The mechanisms involved may have physical limits without being limited in principle, in which case the animal or machine may have "infinite competence" (explained more fully in Sloman (2002)). Even when the competence is not infinite, compositionality implies the ability to deal with novelty, a most important feature for animals and robots inhabiting an extremely variable environment.

Closely related to this are the ability to plan complex future actions and the ability to construct new explanations of observed phenomena.

A more complete exposition would need to discuss different ways in which information bearers can be combined, with different sorts of compositional semantics. One of the major distinctions mentioned in Section 4.4 is between and Fregean and other forms of composition. As explained in (Sloman, 1971), the systematic complexity of forms of representation can provide a basis for reasoning with information-bearers: deriving new conclusions from old information by manipulating the bearers, whether Fregean or not. Logical inference and geometric reasoning using diagrams two special cases among many.

5.5 Potential information content for a user

The information in B can be potentially usable by U even though U has never encountered B or anything with similar information content. That's obviously true when U encounters a new sentence, diagram or picture for the first time. Even before U encountered the new item, it was *potentially* usable as an information-bearer. In some cases, though not all, the potential cannot be realised without U first learning a new language, or notation, or even a new theory within which the information has a place.

You cannot understand the information that is potentially available to others in your environment if you have not yet acquired all the concepts involved in the information. For example, it is likely that a new-born human infant does not have the concept of a metal, i.e. that is not part of its ontology (Sloman, 2009b). So it is incapable of acquiring the information that it is holding something made of metal even if a doting parent says "you are holding a metal object". In humans a lengthy process of development is required for the information-processing mechanisms (forms of representation, algorithms, architectures) to be able treat things in the environment as made of different kinds of stuff, of which metals are a subset. Even longer is required for that ontology to be extended to include the concepts of physics and chemistry. In part that is a result of cultural evolution: not all our ancestors were able to acquire and use such information.

5.6 Potential information content for a TYPE of user

It is possible for information to be potentially available for a TYPE of user even if NO instances of that type exist. For example, long before humans evolved there were things happening on earth that could have been observed by human-like users using the visual apparatus and conceptual apparatus that humans have. But at the time there were no such observers, and perhaps nothing else existed on the planet that was capable of acquiring, manipulating, or using the information, e.g. information about the patterns of behaviours of some of the animals on earth at the time. (This is related to the points made about self-documenting entities in 4.5.)

There may also be things going on whose detection and description would require organisms or machines with a combination of capabilities, including perceptual and representational capabilities and an information-processing architecture, that are possible in principle, but have never existed in any organism or machine and never will – since not everything that is possible has actual instances. Of course, I cannot give examples, since everything I can present is necessarily capable of being thought about by at least one human. Weaker, but still compelling, evidence is simply the fact that the set of things humans are capable of thinking of changes over time as humans acquire more sophisticated concepts, forms of representation and forms of reasoning, as clearly happens in mathematics, physics, and the other sciences. There are thoughts considered by current scientists and engineers that are beyond the semantic competences of any three year old child, or any adult human living 3000 years ago. If the earth had been destroyed three thousand years ago, that might have relegated such thoughts to the realm of possible information contents for types of individual that never existed, but could have.

5.7 Information content shared between users

It is sometimes possible for a bearer B to mean the same thing (convey the same information content I) to different users U and U′, and it is also possible for two users who never use the same information-bearers (e.g. they talk different languages) to acquire and use the same information.

This is why relativistic theories of truth are false. It cannot be true for me that my house has burned down but not true for my neighbour. In principle we have access to the same sources of information in the world.

5.8 Ambiguity, noise, and layers of processing

Media can also vary in the extent to which they allow information to be expressed ambiguously. For example, some cases are totally unambiguous, e.g. the association between bit patterns and CPU instructions or memory addresses in a computer. In a virtual memory system, a bit pattern uniquely identifies a location in a virtual memory, but the mapping to physical memory locations is context sensitive. In natural languages and many forms of pictorial or map-like representation, local details are ambiguous and finding a global interpretation for a complex information-bearer can include searching and problem solving, possibly using constraint propagation and background knowledge, illustrated below in 5.9.

In some cases the medium requires several layers of interpretation, using different ontologies, to be coordinated, e.g. acoustic, phonetic, morphemic, syntactic, semantic and social, in the case of speech understanding systems. Other layers are relevant in visual systems, such as edge features, larger scale 2-D features, 3-D surface fragments, 3-D structures, layers of depth, 3-D processes involving interacting structures, intentions of perceived agents, etc. Trehub (1991) offers a theory about how such layers might be implemented neurally, but there remain many unknowns about how vision works.

In some cases, the requirement for layers of interpretation is the result of engineering designs making use of compression, encryption, password protection, zipping or tarring several files into one large file, and many more. In other cases, the layers are natural consequences of a biological or engineering information-processing task, e.g. the layers in visual information processing.

Some information-bearers include various amounts and kinds of noise, clutter, and partial occlusion, sometimes causing problems that require collaboration between interpretation processes at different levels of abstraction. Where multiple layers of processing are coordinated, ambiguities in some layers may be resolved by interpretations in other layers, possibly using background knowledge (Sloman, 1978, Chap 9). This is sometimes described as “hierarchical synthesis”, or “analysis by synthesis” (Neisser, 1967). A related view of layers of interpretation is presented in H. Barrow and Tenenbaum (1978).

Although there has been much research on ways of extracting information from complex information-bearers, it is clear that nothing in AI comes close to matching, for example, the visual competences of a nest-building bird, a tree-climbing ape, a hunting mammal catching prey, a human toddler playing with bricks and other toys. In part, that is because not even the requirements have been understood properly (Sloman, 2008a).

5.9 Information content for a user determined partly by context

There are lots of structures in perceptual systems that change what information they represent because of the context. E.g. if what is on your retina is unchanged after you turn your head 90 degrees in a room, the visual information will be taken to be about a different wall even if retinal images are unchanged because the two walls have the same wallpaper. The new interpretation uses the information that the head was turned. Many examples can be found in (Berthoz, 2000). (Sloman, 1971) showed how a particular line can represent different things in a 2-D image of a 3-D scene, depending on its relationships to other fragments. Determining whether a vertical line in a picture represents a horizontal mark on the floor or a vertical line on a wall generally requires use of context. Similar problems arise in language processing, e.g. determining whether "with" introduces a prepositional or adverbial phrase in "He watched the boy with binoculars".

Some information-bearing structures express different information for the same user U in different contexts, because they include an explicit indexical element (e.g. "this", "here", "you", "now", or non-local variables in a computer program).

Another factor that makes it possible for U to take a structure B to express different meanings in different contexts can be that B has polymorphic semantics: its semantic function (for U, or a class of users) is to express a higher order function which generates semantic content when combined with a parameter provided by the linguistic or non-linguistic context. E.g. consider: "He ran after the smallest pony". Which pony is the smallest pony can change as new ponies arrive or depart. More subtly, what counts as a tall, big, heavy, or thin X can vary according to the range of heights, sizes, weights, thicknesses of Xs in the current environment and in some cases may also depend on why you are looking for something tall, big, heavy , etc.

There are many more examples in natural language that lead to incorrect diagnosis of words as vague or ambiguous, when they actually express precise higher order functions, applied to sometimes implicit arguments, e.g. "thin", "long", "efficient", "heap". Other examples include spatial prepositions and other constructs, which can be analysed as having a semantics involving higher order functions some of whose arguments are non-linguistic, discussed in (Sloman, 2006b).

A more complex example is: "A motor mower is needed to mow a meadow" which is true only if there is an implicit background assumption about constraints on desirable amounts of effort or time, size of meadow, etc. So a person who utters that to a companion when they are standing in a very large meadow might be saying something true, whereas in a different context, where there are lots of willing helpers, several unpowered lawnmowers available, and the meadow under consideration is not much larger than a typical back lawn, the utterance would be taken to say something different, which is false, even if the utterances themselves are physically indistinguishable. Moreover, where they are standing does not necessarily determine what sort of meadow is being referred to. E.g. they may have been talking about some remote very large or very small meadow.

The influence of context on information expressed is discussed in more detail in relation to Grice's theory of communication, in (Sloman, 2006b), along with implications for the evolution of language. The importance of the role of extra-linguistic context in linguistic communication can be developed in connection with indexicals, spatial prepositions, and Gricean semantics, into a theory of linguistic communications as using higher order functions some of whose arguments have to be extracted from non-linguistic sources by creative problem-solving. This has implications for language learning and the evolution of language. It also requires the common claim that natural languages use compositional semantics, to modified, to allow context to play a role. The use of non-local variables can have a similar effect in programming languages. It seems very likely that brain mechanisms also use context-modulated compositional semantics.

5.10 Information-using subsystems

An information-user can have parts that are information users. This leads to complications such as that a part can have and use some information that the whole would not be said to have. E.g. your immune system and your digestive system and various metabolic processes use information and take decisions of many kinds though we would not say that *you* have, use or know about the information.

Likewise there are different parts of our brains that evolved at different times that use different kinds of information, even information obtained via the same route, e.g. the retina or ear-drum, or haptic feedback. Input and output devices can be shared between sub-systems that use them for different purposes, possibly after different pre- or post- processing, as explained in (Sloman, 1993). Some sub-systems are evolutionarily old and shared with other species, some are newer, and some unique to humans.

An example is the information about optical flow that is used in humans to control posture, without individuals being aware of what they are doing (Lee & Lishman, 1975). More generally, it is likely that human information processing architectures include many components that evolved at different times, performing different functions, many of them concurrent, some of them surveyed in Sloman (2003). The subsystems need not all use the same forms of representation, and individual subsystems need not all have access to information acquired, derived, constructed or used by others. In particular, some will use transient information that is not transferred to or accessible by other subsystems.

That is why much philosophical, psychological, and social theorising is misguided: it treats humans as unitary and rational information users. That includes Dennett's intentional stance and what Newell refers to as "the Knowledge level". For example, the philosophical claim that only a whole human-like agent can acquire, manipulate and use information is false. To understand biological organisms and design sophisticated artificial systems, we need what McCarthy (2008) labels "the designer stance". Unfortunately education about how to be a designer of complex working systems is not part of most disciplines that need it.

5.11 Layers of interpretation in epigenesis

There is a different kind of use of information: when the user is constructing itself! In that process there are not sensors and motors transferring information and energy between the organism and its environment. The processes by which genetic information is used in organisms are very complex and varied. The use of information provided genetically can be very indirect, involving many stages, several of which are influenced by the environment (e.g. maternal fluids, or soil nutrients), so that the interpretation process required for development of an organism, is highly context sensitive.

In many cases, much of the information from which the processes start is encoded in molecular sequences in DNA, specifying, very indirectly, how to construct a particular organism by constructing a very complex collection of self-organising components, which themselves construct more self-organising components. The interpretation of those sequences as instructions depends on complex chemical machinery assembled in a preceding organism (the mother) to kick-start the interpretation process. The interpreting system builds additional components that continue the assembly, partly influenced by the genetic information and partly by various aspects of the environment. During development, the ability to interpret both genetic and environmental information changes, partly under the influence of the environment. So the standard concept of information encoded in the genome is over-simple theory. (Many details are discussed in (Jablonka & Lamb, 2005). The importance of cascaded development of layered cognitive mechanisms influenced by the environment is discussed in Chappell and Sloman (2007). See also (Dawkins, 1982).)

The problems of interpreting and using visual and genetic information show that the role of the user U in obtaining information I from a bearer B in context C may be extremely complex and changeable, in ways that are not yet fully understood. That kind of complexity is largely ignored in most discussions about the nature of information, meaning, representation, but it cannot be ignored by people trying to design working systems.

6 Conclusion

In Section 3 it was claimed that it is not possible to define explicitly, precisely, and without circularity, what we mean by "information", in the semantic sense that involves not merely having some syntactic or geometric form but also having the potential to be taken by a user to be *about* something. So subsequent sections presented an *implicit* definition in the form of a first-draft informal theory about the role of information in our world.

6.1 An implicitly defined notion of "information"

What was said above in Section 3.2 about "energy" applies also to "information". We can understand the word "information" insofar as we use it in a rich, deep, precise and widely applicable theory (or collection of theories) in which many things are said about entities and processes involving information. I suspect that we are still at a relatively early stage in the development of a full scientific theory of information, especially as there are many kinds of information processing in organisms that we do not yet understand.

Some of the contents of a theory of information have been outlined in previous sections, elaborating on the proposition that a user U can interpret a bearer B as expressing information I in context C. Among the topics mentioned include the variety of sources of information, the variety of information-bearing media (about which we still have much to learn), the variety of structures and systems of information-bearers (syntactic forms), the variety of uses to which information can be put (including both communicative and non-communicative uses), the variety of information contents, the variety of ways in which information contents can change (e.g. continuously, discretely, structurally, etc.), the different kinds and degrees of complexity of processes required for interpreting and using the information in particular bearers, the variety of information-using competences different users (or different parts of the same user) can have, the potential information available in objects not yet perceived by information users, and more. We already have broader and deeper understanding of information in this sense than thinkers had a thousand years ago about force and energy, but there is still a long way to go.

Unlike Shannon's information, the information content we have been discussing does not have a scalar value, although there are partial orderings of information content. One piece of information I1 may contain all the information in I2, and not vice versa. In that case we can say that I1 contains more information. I1 can have more information content than both I2 and I3, neither of which contains the other. So there is at most a partial ordering. The partial ordering may be relative to an individual user, because giving information I1 to an user U1, may allow U1 to derive I2, whereas user U2 may not be able to derive I2, because U2 lacks some additional required information. Even for a given user, the ordering can depend on context.

Information can vary both discontinuously (e.g. adding an adjective or a parenthetical phrase to a sentence, like this) or continuously (e.g. visually obtained information about a moving physical object). More importantly, individual items of information can have a structure: there are replaceable parts of an item of information such that if those parts are replaced the information changes but not necessarily the structure. Because of this, items of information can be extracted from other information, and can be combined with other information to form new information items, including items with new structures. This is connected with the ability of information users to deal with novelty, and to be creative. Moreover, we have seen that such compositional semantics often needs to be context sensitive (or polymorphic), both human language and other forms of representation.

It can be stored in various forms, can be modified or extended through various kinds of learning, and can influence processes of reasoning and decision making. Information can also be transmitted in various ways, both intentionally and unintentionally, using bearers of many kinds.

Some items of information allow infinitely many distinct items of information to be derived from them. (E.g. Peano's axioms for arithmetic, in combination with predicate calculus.) Physically finite, even quite small, objects with information processing powers can therefore have infinite information content. (Like brains and computers.)

There is a great deal more that could be said about our current theories about information, but that would take several volumes. Many additional points are in papers in the bibliography, and in other books and journals, as well as in human common sense.

6.2 Life and information

Some of the most important and least well understood parts of a theory about information are concerned with the variety of roles it plays in living things, including roles concerned with reproduction, roles concerned with growth, development, maintenance and repair, roles concerned with perception, reasoning, learning, social interaction, etc. The limitations of our understanding are clearly displayed in the huge gaps between the competences of current robots (in 2009) and the competences of many animals, including human infants and toddlers. For many very narrowly prescribed tasks it is possible to make machines that perform better than humans (e.g. repeatedly assembling items of a certain type from sets of parts arrayed in a particular fashion), but which are easily disrupted by minor variations of the task, the parts, or the starting configuration. Aliens who visited in 1973 and saw what the Edinburgh robot Freddy could do, as described in (Ambler, Barrow, Brown, Burstall, & Popplestone, 1973) and shown in this video `http://groups.inf.ed.ac.uk/vision/ROBOTICS/FREDDY/Freddy_II_original.wmv`, might be surprised on returning 36 years later to find how little progress had been made, compared with ambitions expressed at that time.

Every living thing processes information insofar as it uses (internal or external) sensors to detect states of itself or the environment and uses the results of that detection process either immediately or after further information processing to select from a behavioural repertoire, where the behaviour may be externally visible physical behaviour or new information processing. (Similar points are made in (Reading, 2006) and in Steve Burbeck's web site `http://evolutionofcomputing.org/Multicellular/BiologicalInformationProcessing.html`) In the process of using information an organism also uses up stored energy, so that it also needs to use information to acquire more energy, including the energy required for getting energy.

There are huge variations between different ways in which information is used by organisms, including plants, single celled organisms, and everything else. For example, only a tiny subset of organisms appear to have fully deliberative information processing competence, as defined in Sloman (2006a). As explained in Section 5.10 there can also be major differences between the competences of sub-systems in a single information-user.

6.3 Information processing in virtual machines

A pervasive notion that has been used but not fully explained in this paper is the notion of a virtual machine. Our understanding of requirements for and possible ways of building and using them has gradually expanded through a host of technical advances since the earliest electronic computers were built.

Because possible operations on information are much more complex and far more varied than operations on matter and energy, engineers discovered during the last half-century, as evolution appears to have "discovered" much earlier, that relatively unfettered information processing requires use of a virtual machine rather than a physical machine, like using software rather than cog-wheels to perform mathematical calculations. A short tutorial on virtual machines and some common misconceptions about them can be found in Sloman (2009f). See also Pollock (2008). One of the main reasons for using virtual machines is that they can be rapidly reconfigured to meet changing environments and tasks, whereas rebuilding physical devices as fast and as often is impossible. It is also possible for a physical machine to support types of virtual machine that were never considered by the designer of the physical machine. Similarly, both cultural evolution and individual development can redeploy biological information processing systems in roles for which they did not specifically evolve.

In Sloman (2009f) I suggested that the label "Non-physically-describable-machine" (NPDM) might have been preferable to "virtual machine" (VM) because the key feature is having states and processes whose best description uses concepts that are not definable in terms of the concepts of the physical sciences. Examples are concepts like "winning", "threat", "rule", "pawn", "checkmate", relevant to virtual machines that play chess. These VMs/NPDMs are nothing like the old philosophical notions characterised by Ryle (1949) as referring to "The Ghost in the Machine", for we are not talking about mysterious entities that can continue existing after their physical bodies have been completely destroyed. The crucial point is that the nature of the physical world allows networks of causation to exist that support processes in such virtual machines that not only cause other virtual machine processes to occur but can also influence physical machines, for example when a decision taken by a running chess program causes the display on a computer screen to change (Sloman, 2009e). A crucial step in evolution was the development of causal networks, including sub-systems running in parallel, in virtual machines that could be their own information-users.

This contradicts a number of common mistakes, such as the assumption that information-processing machines have to operate serially, that they have to use only programs installed by a designer, and that they cannot be aware of what they are doing, or why they are doing it, or decide to change their goals. Such mistakes might be overcome if more people studied AI, even if only designing relatively simple agents, as proposed in (Sloman, 2009d).

Although we (or at least software engineers and computer scientists, unlike most philosophers in 2009) understand current virtual machines well enough to create, modify, debug, extend and improve them, the virtual machines that have been produced by biological evolution are another matter: their complexity, their modes of operation, the best ways to describe what they do and how they do it, still defeat scientists, though many subscribe to various personal favourite theories of consciousness, or whatever. Some of them think the known phenomena cannot possibly be explained in terms of information-processing machinery, though in most cases that is because their concept of information-processing is too impoverished – e.g. because based on the notion of a Turing machine, whose relevance to this topic was challenged in Sloman (2002). For example, Turing machines are limited to discrete operations, whereas there is no reason to assume that all information-processing has to be so limited, though it could turn out to be the case that no physical machine could support truly continuous information manipulation. Others take it for granted that brains are information-processing machines, but do not yet understand what information they process or how they do it. For instance, major features of human and animal vision remain unexplained.

6.4 Finally: Is that everything?

It is clear that what I have written so far does not come near exhausting our current notion of information, though it gives an indication of the diversity of phenomena and mechanisms involved. Moreover since most of this was not known a hundred years ago it shows that we are in the process of discovering more about information through scientific and engineering research, though progress has not been as fast as many hoped.

This is just the beginning of an analysis of relationships between information, bearers, users, and contexts. What is written here will probably turn out to be a tiny subset of what needs to be said about information. A hundred years from now the theory may be very much more complex and deep, just as what we know now about information is very much more complex and deep than what we knew 60 years ago, partly because we have begun designing, implementing, testing and using so many new kinds of information-processing machines. The mechanisms produced by evolution remain more subtle and complex, however.

I doubt that anyone has yet produced a clear, complete and definitive list of facts about information that constitute an implicit definition of how we (the current scientific community well-educated in mathematics, logic, psychology, neuroscience, biology, computer science, linguistics, social science, artificial intelligence, physics, cosmology, and philosophy) currently understand and use the word "information". But at least this partial survey indicates how much we have already learnt.

Some physicists seek a "theory of everything", e.g. (J. D. Barrow, 1991; Deutsch, 1997).

However, it does not seem likely that there can be a theory that is recognisable as a *physical* theory from which all the phenomena referred to here would be derivable, even though all the information-processing systems I have referred to, whether natural or artificial, must be *implemented* in physical systems. I suspect that we are in the early stages of understanding how the physical world can support non-physical entities of which simple kinds already exist in running virtual machines in computers, including virtual machines that monitor themselves, and use information about what is happening inside them to take decisions that alter their internal and external behaviours.

My own view has been, for several decades, that as regards information processing our state of knowledge could be compared with Galileo's knowledge of physics. He was making good progress and laying foundations for future developments: including developments he could not possibly imagine.

One of the drivers of progress in science (and philosophy) is improved understanding of what is *not yet known*. I believe the ideas sketched here help us to focus more clearly on aspects of information processing that are not yet understood. Doing that in far more detail with far more specific examples, can help to drive advances that will produce new, deeper, more general explanations. But only time will tell whether this is what Lakatos would call a progressive or a degenerating research programme.

Acknowledgements

I am grateful to Gordana Dodig-Crnkovic and Christophe Menant for useful comments on an earlier draft. This is a much revised version of my attempt to answer a question asking what "semantic information" is, posted on the MINDMECHANISMS discussion forum. My original answer, posted on 20th Sep 2006, is available online at `http://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=ind0609&L=mindmechanisms&T=0&P=1717` as part of a thread with subject "Analysis of conscious functions". Comments and questions by several readers led to major improvements. Many of the points made here were previously also made piecemeal over several years in contributions to the Psyche-D discussion list, now archived at `http://www.archive.org/details/PSYCHE-D`, and in papers and presentations on my web site, listed in the bibliography. Discussions by email and face to face with many colleagues have helped to shape the ideas presented here. It was Max Clowes who first introduced me to computational ways of thinking about philosophical problems.

References

Ambler, A. P., Barrow, H. G., Brown, C. M., Burstall, R. M., & Popplestone, R. J. (1973). A Versatile Computer-Controlled Assembly System. In *Proc. Third Int. Joint Conf. on AI* (pp. 298–307). Stanford, California: IJCAI.

Austin, J. L. (1956). A plea for excuses. In J. O. Urmson & G. J. Warnock (Eds.), *Philosophical Papers* (pp. 175–204). Oxford: Oxford University Press.

Barrow, H., & Tenenbaum, J. (1978). Recovering intrinsic scene characteristics from images. In A. Hanson & E. Riseman (Eds.), *Computer vision systems* (pp. 3–26). New York: Academic Press.

Barrow, J. D. (1991). *Theories of Everything: The Quest for Ultimate Explanation.* New York: OUP.

Bateson, G. (1972). *Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology.* Bungay Suffolk: Chandler Publishing.

Berthoz, A. (2000). *The brain's sense of movement.* London, UK: Harvard University Press.

Bridgman, P. (1927). *The Logic of Modern Physics.* New York: MacMillan.

Brooks, R. A. (1990). Elephants Don't Play Chess. *Robotics and Autonomous Systems*, *6*, 3-15. (http://people.csail.mit.edu/brooks/papers/elephants.pdf)

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

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

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

Cohen, L. (1962). *The diversity of meaning.* London: Methuen & Co Ltd.

Dawkins, R. (1982). *The extended phenotype: The long reach of the gene.* Oxford, New York: Oxford University Press.

Denning, P. J. (2009, June). The profession of IT Beyond computational thinking. *Communications of the ACM*, *52*(6), Pages 28–30. (http://doi.acm.org/10.1145/1516046.1516054)

Deutsch, D. (1997). *The Fabric of Reality.* London: Allen Lane, The Penguin Press.

Dreyfus, H. (2002). Intelligence Without Representation. *Phenomenology and the Cognitive Sciences*, *1*, 367–83. (http://www.class.uh.edu/cogsci/dreyfus.html)

Dyson, G. B. (1997). *Darwin Among The Machines: The Evolution Of Global Intelligence.* Reading, MA: Addison-Wesley.

Floridi, L. (Ed.). (2008). *Philosophy of Computing and Information: 5 Questions.* Copenhagen, Denmark: Automatic Press / VIP. (http://www.amazon.com/Philosophy-Computing-Information-5-Questions/dp/8792130097)

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

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

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

Hare, R. (1952). *The Language of Morals.* Oxford: Oxford University Pres.

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

Jablonka, E., & Lamb, M. J. (2005). *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life.* Cambridge MA: MIT Press.

Kant, I. (1781). *Critique of pure reason.* London: Macmillan. (Translated (1929) by Norman Kemp Smith)

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

Lee, D., & Lishman, J. (1975). Visual proprioceptive control of stance. *Journal of Human Movement Studies*, *1*, 87–95.

Machery, E. (2007). Concept Empiricism: A Methodological Critique. *Cognition*, *104*, 19–46. (http://www.pitt.edu/~machery/papers/machery_2007_conceptmpiricism.pdf)

McCarthy, J. (2008). The well-designed child. *Artificial Intelligence*, *172*(18), 2003-2014. Available from `http://www-formal.stanford.edu/jmc/child.html`

McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of AI. In B. Meltzer & D. Michie (Eds.), *Machine Intelligence 4* (pp. 463–502). Edinburgh, Scotland: Edinburgh University Press. (http://www-formal.stanford.edu/jmc/mcchay69/mcchay69.html)

Minsky, M. L. (1992). Future of AI Technology. *Toshiba Review*, *47*(7). Available from `http://web.media.mit.edu/~minsky/papers/CausalDiversity.html`

M.Minsky. (2005, September). Interior Grounding, Reflection, and Self-Consciousness. In *Brain, Mind and Society, Proceedings of an International Conference on Brain, Mind and Society.* Tohoku University, Japan: Graduate School of Information Sciences, Brain, Mind and Society. (http://web.media.mit.edu/~minsky/papers/Internal Grounding.html)

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

Newell, A. (1980). Physical symbol systems. *Cognitive Science*, *4*, 135–183.

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

Pollock, J. L. (2008). What Am I? Virtual machines and the mind/body problem. *Philosophy and Phenomenological Research.*, *76*(2), 237–309. (http://philsci-archive.pitt.edu/archive/00003341)

Prinz, J. J. (2005). The Return of Concept Empiricism. In H. Cohen & C. Leferbvre (Eds.), *Categorization and Cognitive Science* (p. Part 6). Amsterdam: Elsevier. ((http://subcortex.com/PrinzConceptualEmpiricismPrinz.pdf))

Reading, A. (2006). The Biological Nature of Meaningful Information. *Biological Theory*, *1*(3), 243–249. Available from `doi:10.1162/biot.2006.1.3.243`

Ritchie, D. (1986). Shannon And WEAVER: Unravelling the Paradox of Information. *Communication Research*, *13*(2), 278–298. (http://crx.sagepub.com/cgi/content/abstract/13/2/278)

Ryle, G. (1949). *The concept of mind.* London: Hutchinson.

Schurz, G. (2009). When Empirical Success Implies Theoretical Reference: A Structural Correspondence Theorem. *Brit. J. for the Philosophy of Science*, *60*(1), 101–133. (doi:10.1093/bjps/axn049)

Searle, J. (1980). Minds brains and programs. *The Behavioral and Brain Sciences*, *3*(3). ((With commentaries and reply by Searle))

Shannon, C. (1948, July and October). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423 and 623–656. Available from `http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html`

Simon, H. A. (1969). *The sciences of the artificial.* Cambridge, MA: MIT Press. ((Second edition 1981))

Sloman, A. (1971). Interactions between philosophy and AI: The role of intuition and non-logical reasoning in intelligence. In *Proc 2nd ijcai* (pp. 209–226). London: William Kaufmann. (http://www.cs.bham.ac.uk/research/cogaff/04.html#200407)

Sloman, A. (1978). *The computer revolution in philosophy.* Hassocks, Sussex: Harvester Press (and Humanities Press). Available from `http://www.cs.bham.ac.uk/research/cogaff/crp`

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

Sloman, A. (1982). Image interpretation: The way ahead? In O. Braddick & A. Sleigh. (Eds.), *Physical and Biological Processing of Images (Proceedings of an international symposium organised by The Rank Prize Funds, London, 1982.)* (pp. 380–401). Berlin: Springer-Verlag. (http://www.cs.bham.ac.uk/research/projects/cogaff/06.html#0604)

Sloman, A. (1985). What enables a machine to understand? In *Proc 9th ijcai* (pp. 995–1001). Los Angeles: IJCAI. (http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#4)

Sloman, A. (1987). Reference without causal links. In J. du Boulay, D.Hogg, & L.Steels (Eds.), *Advances in artificial intelligence - ii* (pp. 369–381). Dordrecht: North Holland. (http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#5)

Sloman, A. (1993). The mind as a control system. In C. Hookway & D. Peterson (Eds.), *Philosophy and the cognitive sciences* (pp. 69–110). Cambridge, UK: Cambridge University Press. (http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#18)

Sloman, A. (2002). The irrelevance of Turing machines to AI. In M. Scheutz (Ed.), *Computationalism: New Directions* (pp. 87–127). Cambridge, MA: MIT Press. (http://www.cs.bham.ac.uk/research/cogaff/00-02.html#77)

Sloman, A. (2003). *The Cognition and Affect Project: Architectures, Architecture-Schemas, And The New Science of Mind.* (Tech. Rep.). Birmingham, UK: School of Computer Science, University of Birmingham. (http://www.cs.bham.ac.uk/research/projects/cogaff/03.html#200307 (Revised August 2008).)

Sloman, A. (2006a, May). *Requirements for a Fully Deliberative Architecture (Or component of an architecture)* (Research Note No. COSY-DP-0604). Birmingham, UK: School of Computer Science, University of Birmingham. Available from `http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0604`

Sloman, A. (2006b). *Spatial prepositions as higher order functions: And implications of Grice's theory for evolution of language.* (Research Note No. COSY-DP-0605). Birmingham, UK: School of Computer Science, University of Birmingham. Available from `http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0605`

Sloman, A. (2007a). Diversity of Developmental Trajectories in Natural and Artificial Intelligence. In C. T. Morrison & T. T. Oates (Eds.), *Computational Approaches to Representation Change during Learning and Development. AAAI Fall Symposium 2007, Technical Report FS-07-03* (pp. 70–79). Menlo Park, CA: AAAI Press. (http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0704)

Sloman, A. (2007b). Why Some Machines May Need Qualia and How They Can Have Them: Including a Demanding New Turing Test for Robot Philosophers. In A. Chella & R. Manzotti (Eds.), *AI and Consciousness: Theoretical Foundations and Current Approaches AAAI Fall Symposium 2007, Technical Report FS-07-01* (pp. 9–16). Menlo Park, CA: AAAI Press. (http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0705)

Sloman, A. (2007c). *Why symbol-grounding is both impossible and unnecessary, and why* *theory-tethering is more powerful anyway.* (Research Note No. COSY-PR-0705). Birmingham, UK. (http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#models)

Sloman, A. (2008a). Architectural and representational requirements for seeing processes, proto-affordances and affordances. In A. G. Cohn, D. C. Hogg, R. Möller, & B. Neumann (Eds.), *Logic and probability for scene interpretation.* Dagstuhl, Germany: Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany. Available from `http://drops.dagstuhl.de/opus/volltexte/2008/1656`

Sloman, A. (2008b). *Evolution of minds and languages. What evolved first and develops first in children: Languages for communicating, or languages for thinking (Generalised Languages: GLs)?* (Research Note No. COSY-PR-0702). Birmingham, UK. (http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0702)

Sloman, A. (2008c). Virtual Machines in Philosophy, Engineering & Biology [Extended abstract]. In N. McCarthy & D. Goldberg (Eds.), *Proceedings Workshop on Philosophy & Engineering WPE-2008* (pp. 31–33). London: Royal Academy of Engineering. (http://www.cs.bham.ac.uk/research/projects/cogaff/08.html#803)

Sloman, A. (2009a). Architecture-Based Motivation vs Reward-Based Motivation. *Newsletter on Philosophy and Computers*, *09*(1), 10–13. Available from `http://www.apaonline.org/publications/newsletters/v09n1_Computers_06.aspx`

Sloman, A. (2009b). *Ontologies for baby animals and robots. From "baby stuff" to the world of adult science: Developmental AI from a Kantian viewpoint.* University of Birmingham, School of Computer Science. Available from `http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#brown` (Online tutorial presentation)

Sloman, A. (2009c). Some Requirements for Human-like Robots: Why the recent over-emphasis on embodiment has held up progress. In B. Sendhoff, E. Koerner, O. Sporns, H. Ritter, & K. Doya (Eds.), *Creating Brain-like Intelligence* (pp. 248–277). Berlin: Springer-Verlag. Available from `http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0804`

Sloman, A. (2009d). Teaching AI and Philosophy at School? *Newsletter on Philosophy and Computers*, *09*(1), 42–48. Available from `http://www.apaonline.org/publications/newsletters/v09n1_Computers_17.aspx`

Sloman, A. (2009e, Sep). *Virtual Machines and the Metaphysics of Science.* Nottingham. (http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#mos09, PDF presentation for AHRC Metaphysics of Science Conference)

Sloman, A. (2009f). What Cognitive Scientists Need to Know about Virtual Machines. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1210–1215). Austin, TX: Cognitive Science Society. Available from `http://www.cs.bham.ac.uk/research/projects/cogaff/09.html#901`

Trehub, A. (1991). *The Cognitive Brain.* Cambridge, MA: MIT Press. (http://www.people.umass.edu/trehub/)

Webb, B. (2006). Transformation, encoding and representation. *Current Biology*, *16*(6), R184–R185. Available from `doi:10.1088/1741-2560/3/3/R01`