

Consciousness in Human and Machine: A Theory and Some Falsifiable Predictions

Richard P. W. Loosemore

Surfing Samurai Robots, Inc.
1600 McAllister Road, Genoa NY 13071 USA
rloosemore@susaro.com

Abstract

To solve the hard problem of consciousness we first note that all cognitive systems of sufficient power must get into difficulty when trying to analyze consciousness concepts, because the mechanism that does the analysis will bottom out in such a way that the system declares these concepts to be both real and ineffable. Rather than use this observation to dismiss consciousness as an artifact, we propose a unifying interpretation that allows consciousness to be regarded as explicable at a meta level, while at the same time being mysterious and inexplicable on its own terms. It is further suggested that science must concede that there are some aspects of the world that deserve to be called ‘real’, but which are beyond explanation. The main conclusion is that thinking machines of the future will, inevitably, have just the same subjective consciousness that we do. Some testable predictions can be derived from this theory.

Introduction

The idea that an artificial general intelligence might soon be built raises urgent questions about whether AGIs would (a) be conscious, (b) feel emotions, and (c) have dangerous motivations. Given the strength of public feeling on these matters—for example, the widespread belief that AGIs would be dangerous because as self-aware beings they would inevitably rebel against their lack of freedom—it is incumbent upon the AGI community to resolve these questions as soon as possible. Philosophers may have the luxury of a relaxed debate, but with some people demanding reassurance about the safety of AGI, we do not.

Questions about consciousness, emotion and motivation may be separate issues, but in the public mind they are often conflated, so in this paper I propose to make a start by addressing the first of these. I will argue that if we look carefully at how intelligent systems understand the world we can explain consciousness in a comprehensive manner. This is the first part of a research program aimed at providing a technical foundation for discussions of consciousness, emotion and friendliness in AGI systems.

The Hard Problem of Consciousness

One of the most notorious difficulties with understanding consciousness is the widespread confusion about what the term “consciousness” is supposed to refer to. Chalmers (1996) clarified this somewhat by pointing out that the confusion can be split into two components. First, the word “consciousness” has multiple meanings, so people use it at different times to mean (1) “awakeness,” (2) the ability to have intentions, (3) the thing that makes a philosophical zombie different from a human, and so on.

The second point of confusion is more interesting. Chalmers pointed out that one of these multiple meanings (roughly speaking, the one that is the zombie-human differentiator) is where all of the real philosophical difficulty resides, and he labeled this the “hard problem” of consciousness. Other questions—for example, about the neural facts that distinguish waking from sleeping—may be interesting in their own right, but they do not involve deep philosophical issues and should not be confused with the hard problem.

Defining the Hard Problem. The hard problem is all about the first-person, subjective experience of a creature that is conscious, and about the fact that no matter how good our objective scientific knowledge of the world might become, there seems to be no way to account for that internal subjective experience. Included within the scope of these subjective experiences are questions about “qualia” (the subjective quality of color sensations, pains and the like) and the concept of “self” (that indefinable feeling that we are each a non-physical agent that looks out at the world). Most importantly, the hard problem revolves around the conviction that there could conceivably be such a thing as a “philosophical zombie,” which is defined to be a creature that is identical to a human, but which lacks any subjective phenomenology. If zombies are conceivable, we have to account for the thing that they lack, and the problem of accounting for that thing is the hard problem.

Many philosophers and most lay people would say that these subjective aspects of consciousness are so far removed from normal scientific explanation that if anyone proposed an objective explanation for the hard problem of consciousness they would be missing the point, because

such an explanation would have to start with a bridge between the ideas of *objective* and *subjective*, and since no consensus idea has ever been proposed that might act as such a bridge, no explanation is even on the horizon.

We can summarize the current situation in the philosophical analysis of the hard problem by framing it in terms of the following impasse:

Skeptic: Tell me in objective terms what exactly is meant by terms such as “consciousness” and “qualia,” and we might begin to build an explanation for them. Unless you can say exactly what you mean by these things, you are not saying anything.

The Reply: Unfortunately, the thing we are talking about seems to be intrinsically beyond the reach of objective definition, while at the same time being just as deserving of explanation as anything else in the universe. This lack of objective definition should not be taken as grounds for dismissing the problem—rather, this lack of objective definition IS the problem.

A Preview of the Strategy

The line of attack in this paper has two parts. First, we pick up the idea that the hard problem is about ideas that cannot be clearly defined. Why are they indefinable? Why are we nevertheless compelled to explain them? After suggesting a way to understand how something could be so unusual as to drive philosophers into this paradoxical mixed state, we then go to a second phase of the argument, in which we ask about the “reality” of things that test the limits of what minds can know. At the end of part 1 we seem to be heading in a direction that the skeptic would favor (eliminating the explanandum as an epiphenomenon), but then part 2 makes an unusual turn into a new compromise, neither dualist nor physicalist, which resolves the problem of consciousness in an unorthodox way.

Part 1: The Nature of Explanation

The various facets of consciousness have one thing in common: they involve some form of introspection, because we “look inside” at our subjective experience of the world (qualia, sense of self, and so on) and ask what these experiences amount to. In order to analyze the nature of these introspection we need to take one step back and ask what happens when we think about any concept, not just those that involve subjective experience.

Talking About Analysis Mechanisms

In any sufficiently complete AGI system there has to be a powerful mechanism that lets the system analyze its own concepts. The system has to be able to explicitly think about what it knows, and deconstruct that knowledge in many ways. The scope of this *analysis mechanism* must be extremely broad, and the knowledge that lies within its scope must be couched at an appropriate level. (So: merely giving the AGI access to its own source code would not count as an analysis mechanism).

AGI systems will surely have this analysis mechanism at some point in the future, because it is a crucial part of the “general” in “artificial general intelligence,” but since there is currently no consensus about how to do this, we need to come up with a language that allows us to talk about the kind of things that such a mechanism might get up to. For that reason, I am going to use a language derived from my own approach to AGI—what I have called elsewhere a “molecular framework” for cognition (Loosemore, 2007; Loosemore and Harley, forthcoming).

Nothing depends on the details of this molecular framework, because any other AGI formalism can be translated into this architectural style, but since the molecular framework is arguably more explicit about what the analysis mechanism does, we get the benefit of a concrete picture of its doings. Other AGI formalisms will perhaps take a different approach, but any analysis mechanism must have the crucial features on which this explanation of consciousness depends, so the molecular framework does nothing to compromise the argument.

The Molecular Framework

The following is a generic model of the core processes inside any system that engages in intelligent thought. This is meant as both a description of human cognition and as a way to characterize a wide range of AGI architectures.

The basic units of knowledge, in this framework, are what we loosely refer to as “concepts,” and these can stand for *things* [chair], *processes* [sitting], *relationships* [on], *operators* [describe], and so on. The computational entities that encode concepts are to be found in two places in the system: the **background** (long-term memory, where there is one entity per concept) and the **foreground**, which contains the particular subset of concepts that the system is using in its current thoughts.

The concept-entities in the foreground will be referred to as **atoms**, while those in the background are **elements**.

Many instances of a given concept can be thought about at a given time, so there might be several [chair] atoms in the foreground, but there will only be one [chair] element in the background. From now on, we will almost exclusively be concerned with atoms, and (therefore) with events happening in the foreground.

Theorists differ in their preference for atoms that are either active or passive. A passive approach would have all the important mechanisms on the outside, so that the atoms are mere tokens. An active approach, on the other hand, would have no external mechanisms that manipulate atoms, but instead put all the interesting machinery in and between the atoms. In the present case we will adopt the active, self-organized point of view: the atoms themselves do all of the work of interacting with and operating on one another. This choice makes no difference to the argument, but it gives a clearer picture of some claims about semantics that come later.

Two other ingredients that need to be mentioned in this simplified model are external sensory input and the self-model. We will assume that sensory information originates

at the sensory receptors, is pre-processed in some way, and then arrives at the edge of the foreground, where it causes atoms representing primitive sensory features to become active. Broadly speaking, atoms near the foreground periphery will represent more concrete, low-level concepts, while atoms nearer the “center” of the foreground will be concerned with more high-level, abstract ideas.

The self-model is a structure (a large cluster of atoms) toward the center of the foreground that represents the system itself. This self-model is present in the foreground almost all of the time, because the self is clearly present whenever the system is thinking about anything. At the core of the self-model is a part of the system that has the authority to initiate and control actions.

Finally, note that there are a variety of operators at work in the foreground. The atoms themselves do some of this work, by trying to activate other atoms with which they are consistent (thus, a [cat] atom that is linked to a [crouching-posture] atom will tend to activate an atom representing [pounce]). But there will also be mechanisms that do such things as creation (making a new element to encode a new conjunction of known atoms), elaboration (the assembly of a cluster to represent a situation in more detail), various forms of analogy construction, and so on.

Overall, this model of cognition depicts the process of thought as being a collective effect of the interaction of all these atoms and operators. The foreground resembles a molecular soup in which atoms assemble themselves (with the help of operators) into semi-stable, dynamically changing structures. Hence the term “molecular framework” to describe this way of modeling cognition.

Explanation in General

Atoms can play two distinct roles in the foreground, corresponding to the difference between use and mention. If the system is perceiving a chair in the outside world, a [chair] atom will be part of the representation of that outside situation. But if the system asks itself “What is a chair?”, there will be one [chair] atom that stands as the target of the representation.

When an atom becomes a target, operators will cause this target atom to be elaborated and unpacked in various ways. Call this set of elaboration and unpacking operations an “analysis” event. An analysis event involves various connected concepts being activated and connected to the [chair] atom. If we answer the question by saying that a chair has four legs and is used for sitting on, then this will be because the analysis has gone in such a direction as to cause [sitting] and [function-of] atoms to be activated, as well as a cluster involving [legs], [four] and [part-of].

It is important to be clear that the sense of “explain” that we are examining here is the one that distinguishes itself clearly from the sense that means, simply, “finding all associated concepts.” Analysis, as it is construed here, is not about the fact that [red] tends to be associated with [lips], [blood], [redcurrants] and so on. Humans, and sufficiently powerful AGI systems, clearly have the ability to reduce concepts to more basic terms. This reductionist

type of mechanism is the one that we mean when we talk about the analysis of a target atom.

If this were about narrow AI, rather than AGI, we might stop here and say that the essence of “explanation” was contained in the above description of how the [chair] concept was analyzed into a more detailed representation. However, in an AGI system these core aspects of the analysis process are only part of a much larger constellation of other structures and operators, including representations of: the person who asked the question; that person’s intentions; some background about the different kinds of explanation that are appropriate in different contexts; the protocols for constructing sentences that deliver an answer; the status and reliability of the knowledge in question, and so on.

Analysis is not really a single mechanism, it is an open-ended cluster of flexible, context-dependent mechanisms. More like a poorly demarcated sector of an ecology, than a crisply defined mechanism. However, for the purposes of discussion we will refer to the whole thing as if it were a single “analysis mechanism.”

Explaining Subjective Concepts

In the case of human cognition, what happens when we try to answer a question about our subjective experience of the color red? In this case the analysis mechanism gets into trouble, because the [red] concept is directly attached to an incoming signal line and has no precursors. The [red] concept cannot be unpacked like most other concepts.

The situation here is much worse than simply not knowing the answer. If we are asked to define a word we have never heard of, we can still talk about the letters or phonemes in the word, or specify where in the dictionary we would be able to find the word, and so on. In the case of color qualia, though, the amount of analysis that can be done is precisely zero, so the analysis mechanism returns nothing.

Or does it? I propose that, because of the nature of the representations used in the foreground, there is no way for the analysis mechanism to fail to return some kind of answer, because a non-answer would be the same as representing the color of red as “nothing,” and in that case all colors would be the same. Nothing is not an option, for the same reason that a calculator cannot report that “3 minus 3 is ...” and then not show anything on its display. Structurally, the analysis mechanism must return an atom representing [the subjective essence of the color red], but this atom is extremely unusual because it contains nothing that would allow it to be analyzed. Any further attempt to apply the analysis mechanism to this atom will yield just another atom of the same element.

This bottoming-out of the analysis mechanism causes the cognitive system to eventually report that “There is definitely something that it is like to be experiencing the subjective essence of red, but that thing is ineffable and inexplicable.” This is the only way it can summarize the utterly peculiar circumstance of analyzing [x] and getting [x] back as an answer.

This same “failure” of the analysis mechanism is common to all of the consciousness questions. For qualia, the mechanism dead-ends into the sensory atoms at the edge of the foreground. For the concept of self, there is an innate representation for the self that cannot be analyzed further because its purpose is to represent, literally, itself. On reflection, it seems that all subjective phenomenology is associated with such irreducible atoms.

In each case it is not really a “failure,” in the sense that a mechanism is broken, nor is it a failure that results from the system simply not knowing something. It is an unavoidable consequence of the fact that the cognitive system is powerful enough to recursively answer questions about its own knowledge. According to this view, any intelligent system powerful enough to probe its own intellect in this deep way—any system with an analysis mechanism—would spontaneously say the same things about consciousness that we do.

Finally, it is worth reiterating the point made earlier: this account does not depend on the specifics of the molecular framework. All AGI systems must fail in the same way.

The ‘That Misses The Point’ Objection

The most common philosophical objection to the above argument is that it misses the point, because it explains only the locutions that people produce when talking about consciousness, not the actual experiences they have.

The problem with this objection is that it involves an implicit usage of the very mechanism that is supposed to be causing the trouble. So, when we say “There is something missing from this argument, because when I look at my subjective experiences I see things that are not referenced by the argument”, what we are doing is staying within the system and asking for an explanation of (say) color qualia that is just as good as the explanations we can find for other concepts. But this within-the-system comparison of consciousness with ordinary concepts is precisely the kind of thought process that will invoke the analysis mechanism. And the analysis mechanism will then come back with the verdict that the Loosemore Argument fails to describe the nature of conscious experience, just as other attempts to explain consciousness have failed.

There is no space to analyze all possible objections here, but I would offer instead the following conjecture: when the objections are examined carefully, they will always be found to rely, for their force, on a line of argument that causes the analysis mechanism to run into a dead end. At the same time that the argument cites the analysis mechanism as the chief culprit, then, the objections try to use deploy the analysis mechanism (with flaw intact) to explain why the argument cannot be right.

But this still leaves something of an impasse. The argument does say nothing about the nature of conscious experience, *qua* subjective experience, but it does say why it cannot supply an explanation of subjective experience. Is explaining why we cannot explain something the same as explaining it?

Part 2: The Real Meaning of Meaning

This is not a very satisfactory resolution of the problem, because it sounds as if we are being asked to believe that our most immediate, subjective experience of the world is, in some sense, an artifact produced by the operation of the brain. Of course, the word “artifact” is not quite right here, but then neither are “illusion,” “mirage,” “hallucination,” or any of the other words that denote things that seem to exist, but are actually just a result of our brains doing something odd. By labeling consciousness as a thing that intelligent systems *must* say they experience, (because their concept-analysis mechanisms would not function correctly otherwise), we seem to be putting consciousness on a par with artifacts, illusions and the like. That seems, on the face of it, a bizarre way to treat something that dominates every aspect of our waking lives.

I believe that it is wrong to take the view that the meta-account of consciousness given above leads inexorably to the conclusion that consciousness is some kind of artifact. The best, most satisfying conclusion is that all of the various subjective phenomena associated with consciousness should be considered just as “real” as any other phenomenon in the universe, but that science and philosophy should concede that they have the special status of being unanalyzable. We should declare that such phenomena can be predicted to occur under certain circumstances (namely, when an intelligent system has the kind of powerful “analysis” mechanism described earlier), but that nothing can be said about their nature. In effect, we would be saying that these things are real, but beyond the reach of science.

The remainder of the argument is an attempt to explain and justify this position.

Getting to the Bottom of Semantics

The crucial question, then, is what status we should give to the atoms in a cognitive system that have this peculiar property of making the analysis mechanism return a verdict of “this is real, but nothing can be said about it”.

To answer this question in a convincing way, we need to be more specific about the criteria we are using to justify our beliefs about the realness of different concepts (epistemology), the meaning of concepts (semantics and ontology), and the standards we use to judge the validity of scientific explanations. We cannot simply wave our hands and pick a set of criteria to apply to these things, we need some convincing reasons for choosing as we do.

There seem to be two choices here. One would involve taking an already well-developed theory of semantics or ontology—off the shelf, so to speak—then applying it to the present case. For example, we might choose to go with some form of “possible worlds” semantics, and then note that, according to this perspective, the offending concept-atoms do not take part in any conceivable functions defined over possible worlds, so therefore they can be dismissed as fictions that do not correspond to anything meaningful.

The second choice is to take a detailed look at all the different semantic/ontological frameworks that are available and find out which one is grounded most firmly; which one is secure enough in its foundations to be **the** true theory of meaning/reality/explanation.

The perceptive reader will notice that we are in the process of walking into a trap.

The trap is as follows. If someone were to suggest that the concept of the “meaning” of language can be reduced to some simpler constructs (perhaps, the meanings of basic terms plus rules of compositionality), and that these constructs may then be reduced further (perhaps to functions over possible worlds), and that this reduction could continue until we reach some very basic constructs that are intuitively obvious or self-evident, then that person would be embarking on a doomed endeavor, because any such reductionist plan would end in circularity, or descend into an infinite regress. No matter how far down the reduction went, questions could always be asked about the meanings of the most basic terms (“You are explaining this in terms of ‘possible worlds’? Please tell me the meaning of ‘possible world’?”). The choice then is to either declare an arbitrary limit to the process, or admit that it leads to an infinite regress of questions.

This circularity or question-begging problem applies equally to issues of the meaning of “meaning” and explanations of the concept of “explanation,” and it afflicts anyone who proposes that the universe can be discovered to contain some absolute, objective standards for the “meanings” of things, or the fundamental nature of explanatory force.

Extreme Cognitive Semantics

The only attitude to ontology and semantics that escapes this trap is something that might be called “Extreme Cognitive Semantics”—the idea that there is no absolute, objective standard for the mapping between symbols and things in the world, because this mapping is entirely determined by the purely contingent fact of the design of real cognitive systems (Croft and Cruse, 2004; Smith and Samuelson, 1997). There is no such thing as the pure, objective meaning of the symbols that cognitive systems use, there is just the way that cognitive systems do, in fact use them. Meanings are determined by the ugly, inelegant design of cognitive systems, and that is the end of it.

How does this impact our attempt to decide the status of those atoms that make our analysis mechanisms bottom out? The first conclusion should be that, since the meanings and status of all atoms are governed by the way that cognitive systems actually use them, we should give far less weight to any externally-imposed formalism (like possible-worlds semantics) which says that according to its strictures, subjective concepts point to nothing and are therefore fictitious.

Second—and in much the same vein—we can note that the atoms in question are such an unusual and extreme case, that formalisms like traditional semantics should not even be expected to handle them. This puts the shoe firmly

on the other foot: it is not that these semantic formalisms have no place for the consciousness-concepts and *therefore* the latter are invalid, it is rather that the formalisms are too weak to be used for such extreme cases, and therefore they have no jurisdiction in the matter.

Finally, we can use the Extreme Cognitive Semantics (ECS) point of view to ask what it means to judge various concepts as possessing different degrees of “realness.”

The natural, usage-centered meaning of “real” seems to have two parts. The first involves the precise content of a concept and how it connects to other concepts. So, unicorns are not real because they connect to our other concepts in ways that clearly involve them residing only in stories. The second criterion that we use to judge the realness of a concept is the directness and immediacy of its phenomenology. Tangible, smellable, seeable things that lie close at hand are always more real.

Interestingly, the consciousness atoms score differently on these two measures of realness: they connect poorly to other concepts because we can say almost nothing about them, but on the other hand they are the most immediate, closest, most tangible concepts of all, because they *define* what it means to be “immediate” and “tangible.”

Implications

What to conclude from this analysis? I believe that the second of these two criteria is the one that should dominate, and that the correct explanation for consciousness is that all of its various phenomenological facets deserve to be called as “real” as any other concept we have, because there are no meaningful *objective* standards that we can apply to judge them otherwise. But while they deserve to be called “real” they also have the unique status of being beyond the reach of scientific inquiry. We can talk about the circumstances under which they arise, but we can never analyze their intrinsic nature. Science should admit that these phenomena are, in a profound and specialized sense, mysteries that lie beyond our reach.

This is a unique and unusual compromise between materialist and dualist conceptions of mind. Minds are a consequence of a certain kind of computation; but they also contain some mysteries that can never be explained in a conventional way. We cannot give scientific explanations for subjective phenomena, but we can say exactly why we cannot say anything: so in the end, we can explain it.

Conclusion: Falsifiable Predictions

This theory of consciousness can be used to make some falsifiable predictions. Unfortunately, we are not yet in a position to make tests of these predictions, because doing so would require the kind of nanotechnology that would let us rewire our brains on the fly.

The uniqueness of these predictions lies in the fact that there is a boundary (the edge of the foreground) at which the analysis mechanism gets into trouble. In each case, the

prediction is that these phenomena will occur at exactly that boundary, and nowhere else. Once we understand enough about the way minds are implemented in brains (or in full-scale AGI systems), we will be in a position to test the predictions, because the predicted effects must occur at the boundary if the prediction is to be confirmed.

Prediction 1: Blindsight. Some kinds of brain damage cause the subject to experience ‘blindsight,’ a condition in which they report little or no conscious awareness of certain visual stimuli, while at the same time showing that they can act on the stimuli (Weiskrantz, 1986). The prediction in this case is that some of the visual pathways will be found to lie outside the scope of the analysis mechanism, and that the ones outside will be precisely those that, when spared after damage, allow visual awareness without consciousness.

Prediction 2: New Qualia. If we were to build three sets of new color receptors in the eyes, with sensitivity to three bands in, say, the infrared spectrum, and if we also built enough foreground wiring to supply the system with new concept-atoms triggered by these receptors, this should give rise to three new color qualia. After acclimatizing to the new qualia, we could then swap connections on the old colors and the new IR pathways, at a point that lies *just outside the scope of the analysis mechanism*.

The prediction is that the two sets of color qualia will be swapped in such a way that the new qualia will be associated with the old visible-light colors. This will only occur if the swap happens beyond the analysis mechanism.

If we subsequently remove all traces of the new IR pathways outside the foreground (again, beyond the reach of the analysis mechanism), then the old color qualia will disappear and all that will remain will be the new qualia.

Finally, if we later reintroduce a set of three color receptors and do the whole procedure again, we can bring back the old color qualia, but only if we are careful: the new receptors must trigger the foreground concept-atoms previously used for the visible-light colors. If, on the other hand, we completely remove all trace of the original concept atoms, and instead create a new set, the system will claim not to remember what the original qualia looked like, and will tell you that the new set of colors appear to have brand new qualia.

Prediction 3: Synaesthetic Qualia. Take the system described above and arrange for a cello timbre to excite the old concept-atoms that would have caused red qualia: cello sounds will now cause the system to have a disembodied feeling of redness.

Prediction 4: Mind Melds. Join two minds so that B has access to the visual sensorium of A, using new concept-atoms in B’s head to encode the incoming information from A. B would say that she knew what A’s qualia were like, because she would be experiencing new qualia. If, on the other hand, the sensory stream from A is used to trigger the old concept atoms in B, with no new atoms being constructed inside B’s brain, B would say that A’s qualia were the same as hers.

Conclusion

The simplest explanation for consciousness is that the various phenomena involved have an irreducible duality to them. On the one hand, they are *meta*-explicable, because we can understand that they are the result of a powerful cognitive system using its analysis mechanism to probe concepts that are beyond its reach. On the other hand, these concepts deserve to be treated as the most immediate and real objects in the universe, because they define the very foundation of what it means for something to be real—and as real things, they appear to have ineffable aspects to them. Rather than try to resolve this duality by allowing one interpretation to trump the other, it seems more rational to conclude that both are true at the same time, and that the subjective aspects of experience belong to a new category of their own: they are real but inexplicable, and no further scientific analysis of them will be able to penetrate their essential nature.

According to this analysis, then, any computer designed in such a way that it had the same problems with its analysis mechanism as we humans do (arguably, any fully sentient computer) would experience consciousness. We could never “prove” this statement the way that we prove things about other concepts, but that is part of what it means to say that they have a special status—they are real, but beyond analysis—and the only way to be consistent about our interpretation of these phenomena is to say that, insofar as we can say anything at all about consciousness, we can be sure that the right kind of artificial general intelligence would also experience it.

References

- Chalmers, D. J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- Croft, W. and Cruse, D. A. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Dowty, D. R., Wall, R. E., & Peters, S. 1981. *Introduction to Montague Semantics*. Dordrecht: D. Reidel.
- Loosemore, R. P. W. (2007). Complex Systems, Artificial Intelligence and Theoretical Psychology. In B. Goertzel & P. Wang (Eds.), *Proceedings of the 2006 AGI Workshop*. IOS Press, Amsterdam.
- Loosemore, R. P. W., & Harley, T.A. (forthcoming). Brains and Minds: On the Usefulness of Localisation Data to Cognitive Psychology. In M. Bunzl & S. J. Hanson (Eds.), *Foundations of Functional Neuroimaging*. Cambridge, MA: MIT Press.
- Smith, L. B., and Samuelson, L. K. 1997. Perceiving and Remembering: Category Stability, Variability, and Development. In K. Lamberts & D. Shanks (Eds.), *Knowledge, Concepts, and Categories*. Cambridge: Cambridge University Press.
- Weiskrantz, L., 1986. *Blindsight: A Case Study and Implications*. Oxford: Oxford University Press.