Teaching about Consciousness in Cognitive Science

Author: WILLIAM BANKS

This document relates some of my approaches to teaching consciousness in the context of cognitive science. I cover consciousness in two of my courses. In my upper-division course on perception and cognition I discuss consciousness issues throughout the semester – it is unavoidable -- but I have only one class meeting devoted to consciousness. The other course is a seminar titled Consciousness and Cognition. I have organized this document around the content of the seminar, but all the topics are touched on in the perception course.

Over the past 30 years I have seen a rather large shift in the way consciousness is covered in cognitive psychology classes. Students have always been interested in the question but in the past most accepted the dictum that it was impossible to study it. Now they have no compunctions about pursuing questions about it. They don't simply express a general interest, they ask specific questions that go to the heart of issues. In recent semesters I have had many students ask about the unconscious emotional processing that is discussed in Malcom Gladwell's *Blink*, for example. They want to know how this can be, how unconscious thought can be intelligent and strategic. They ask what consciousness is for if unconscious thought is apparently sophisticated.

For another example, the idea of a largely unconscious dorsal stream of perceptual processing is persistently fascinating, to me as well as to the students (*see Milner & Goodale, 2003* for excellent and accessible coverage of the two streams of visual processing). They are intrigued by the fact that the representation of the world is
Teaching About Consciousness in Cognitive Science

not unitary in the brain. Location and place are processed by the dorsal stream while object identification is processed by the ventral stream. The idea of dissociations between perceptual functions within object recognition is another topic that surprises and intrigues students. Ramachandran has covered such dissociations with dramatic flair in several of his books. Oliver Sachs is often the first name to come to mind in these instances, but Ramachandran has much better theoretical accounts of – and more interesting speculations about -- what neural processes may be responsible for these surprising and revealing phenomena. It is more than a matter of curiosity when a brain-damaged person believes that her paralyzed arm is functional, and then goes on to "prove" it by using the intact hand to move the paralyzed one like a puppet. This disorder tells us much about how the recognition, action, and ownership of our bodies must coordinate for normal functioning.

Teachers of cognitive science need to be prepared to cover questions students have about consciousness. This can be a challenge, one that is worth it because it can be an opening into fascinating territory. We also need to bring findings and phenomena into the classroom. The strategy I recommend is to start by reading several books on the topic for perspective. I've listed a few at the end of this outline. I'd use web resources to answer specific questions and to learn about the controversies. Although most of the information on the web seems to be accurate, it is wise to be skeptical and to consult more than one source. Still, it is difficult to avoid the web as a source. It is so convenient and more current than print sources. The field is growing so rapidly that it is useful to know what has been found in the last few weeks or months.

How is consciousness defined? I am sometimes asked to define consciousness, as if I were an expert in the area who could bring enlightenment to this very difficult research topic by providing a succinct definition. The question often has some urgency behind it because it is thought that we must define it before we can study it. I do not think that a definition is necessary. In most scientific fields the definitions mature as the field progresses. Consider water as an example. I am sure that people have had definitions of water since the beginning of language. The definitions all along have been
adequate to point out the same substance, but the introduction of the idea that it has the chemical formula $\text{H}_2\text{O}$ marked a significant change in the content of the definition. It also marked a scientific accomplishment that took hundreds of years and the work of hundreds of scientists, not to mention a revolution or two in thought. A definition points to a thing and rests on an accepted body of knowledge. To begin research on consciousness we need no more than a common agreement on what we are researching. We can point to it even if we don't have a clue as to what it is. Here are some ways of pointing to consciousness and a few important characteristics, most but not all of which are uncontroversial.

a. Consciousness is personal experience. It is what I enter into when I wake up in the morning and what goes away when I fall asleep (except for dreams).

b. Consciousness is not a thing, but a state.

c. My consciousness is real but not directly observable by anyone else. As John Searle has put it, consciousness is ontologically real but epistemologically not accessible directly.

d. Following point c we call my view of my own consciousness the first-person perspective and everyone else's view the third-person perspective.

e. Consciousness has "me-ness." That is to say, it always has the stamp of self; what I see is mine, even if it is the darkness when I close my eyes.

f. Consciousness is continuous. This is to say there are not obvious gaps in it. This is also termed the unity of consciousness. It is a continuing wonder to me that consciousness has any sort of unity or continuity. The brain is massively parallel in operation and has many separate areas processing different things and in continual competition.

I think this is a point where we depart from neutral assumptions. The unity discussed in f may be introspectively confirmed, but it may reflect only a post-hoc rationalization. Something like Gazzaniga's interpreter may put experiences in order and bring unity out of chaos. Further, many people, prominent among them being stream-of-consciousness novelists (and William James), are quite aware that many things are
going on in the mind at the same time, and that our conscious stream can shift and jump abruptly.

1. The mind-body problem.

   One way to state the mind-body problem is as the problem of how matter can "light up" and become conscious. Another would be on how soul and body are related. These and all other statements make assumptions that guide thinking. The idea that matter "lights up" takes an implicitly materialistic position, assuming that matter is the substance of the world, and consciousness is generated by physics or biology. This version of the problem leads to the quest for the neural correlate of consciousness (the NCC) and other research on consciousness. I find that putting the problem this way appeals intuitively to my students, who for the most part have a perspective formed by modern science.

   The body-soul formulation, on the other hand, is the heart of the traditional debate on the problem, which started with the dualism of René Descartes (1596-1650), which in turn reflected Christian doctrine. This formulation must deal with the question of how two different sorts of substance – matter and soul – can interact. All the familiar solutions to the mind-body problem, such as Descartes' idea that they connected through the pineal gland in the brain and the doctrine of psycho-physical parallelism (which assumes that soul and body do not interact), derive from this way of putting the problem. The doctrine of epiphenomenalism is another such solution. Thomas Henry Huxley (1825-1895), in a memorable description of epiphenomenalism, likened consciousness to the whistle of a steam locomotive. It is created by the locomotive but cannot influence its action; our consciousnesses are just observers. Ironically, the hardheaded thinkers who declare consciousness unfit for scientific research because it is only an "epiphenomenon" are expressing an implicit belief in a supernatural soul, all the while claiming to be rigorous scientific materialists.

   In my presentations of consciousness in class, I never start off with the mind-body problem. When I have done so students either glazed over or became involved
with questions that lead away from scientific issues. They too often think that the problem must be solved before any progress can be made. I think it's better to let the mind-body related problems to come up in the context of scientific research. The field is infused with questions related to the mind-body problem. All we need to do is show an fMRI of someone looking at a red patch and ask how the subjective experience of red might be related to the imaged events.

There are ways to bring up these issues in a safe way. For example, to provoke thinking I will sometimes ask the class, "What is your brain doing when you are thinking?" This question leads to questions about the difference between brain activity and our knowledge of it, how the mind relates to the brain, and how a thought is represented in the brain.

2. Imaging the working brain.
The methodological techniques are not usually a good topic for a lecture or discussion, but they interest students greatly in context. The idea that we can track processing of a stimulus by monitoring EEG recordings from the surface of the head, or that we can observe rather intricate brain processes with fMRI is very exciting to students. Until very recently we have had to treat the brain as a closed box. Our best evidence about what different parts of the brain do used to come from strokes and brain injuries. Now we are getting some very revealing peeks into the box.

My advice is to be prepared for questions about the different imaging techniques. How do they work? What are the tradeoffs in terms of resolution, speed, and type of neural signal measured?


3. Neural Correlates of Consciousness (NCC).
This topic, the NCC, covers the general question of what activity in the brain is responsible for consciousness. The term "correlate" may be unfortunate. Correlation does not imply causation, and what we really want to know is how consciousness results from (is caused by) neural activity. However, we are so far from answering that question that we settle for correlates of consciousness and hope a causal theory will come along.

The establishment of a correlate of consciousness has some empirical hurdles to pass. For example, if several areas of the brain show activity during processing, which ones are crucial to consciousness? The biggest problem, however, is how consciousness is generated, if it is, by the activity we observe. We have no theories that make this connection.

Christof Koch's research program on this topic is described in Koch (2004). The URL below is for an article written by Koch and Sir Francis Crick that describes their approach and some findings.


The Wikipedia article on the topic is succinct, but the first section on conceptual issues may be confusing to students and is not representative of the way the search for NCC is conducted in research. The first section does, however, illustrate well the intellectual baggage the seemingly innocent term "correlate" carries.

http://en.wikipedia.org/wiki/Neural_correlate

Metzinger (2000) contains articles that give in-depth coverage of the area and is highly recommended.

4. Attention

This is a large area, with many different applications to consciousness. A common question is the relationship between consciousness and attention. The answer is not simple. The idea that the concepts name the same thing comes from the fact that awareness of objects usually requires attention. However, attention does not always create awareness. In cases of blindsight the patients have a large visual scotoma and no
awareness of images that fall in that area (see Weizkrantz, 1999). Nevertheless, they can report with above chance accuracy some aspects of objects in the area, all the while objecting that they cannot see anything. Attention to the area within the scotoma improves performance but does not confer awareness on objects in the scotoma.

An example of a paradoxical effect of attention is in the attentional blink, in which attention makes images harder to see. The attentional blink takes place when a series of images or pictures is shown very rapidly, at rates up to ten per second. If the participant is asked to report an item cued by an earlier item in the series, the cued item is less likely to be seen than if it is not cued. A demonstration of the attentional blink can be found at http://members.tripod.com/jon_slemmer/shock.html.

This site also demonstrates another attentional effect, change blindness. It is demonstrated with considerable flair at a website created by Daniel Simons, the person who discovered the phenomenon: http://viscog.beckman.uiuc.edu/djs_lab/demos.html. The change blindness effect demonstrates that while we feel we are "seeing" the entire scene as we look over it, that sense of seeing greatly exaggerates what we actually perceive. In one demonstration two seemingly identical scenes alternate with a 1-second gray field in between. The two scenes are not identical but differ on some obvious feature, such as a bicycle that is in one but not the other. Most people have to watch a lot of switches between the two images before they see the difference, and once they see it they are shocked by how large and easily visible the difference is.

While it is natural to think of awareness following attention, the reverse can happen. In so-called bottom-up attention the object is seen first and then attracts attention.

5. Implicit cognition

Many questions fall under this rubric. Is it possible to perceive something without being conscious of it? Do we have emotions, thoughts, desires that we are not aware of? Is there active thought outside of the realm of awareness?
This area has a long history, and there is still controversy about every type of implicit cognition.

Gladwell's *Blink* is a good source of illustrations of situations and experiments in which consciousness lags intelligent action, or in which we are never aware at all of clever things we did automatically. Do these cases imply that we think unconsciously? There are other explanations, but the cases are fascinating in themselves.

Unconscious perception has been researched extensively. Debate on the topic becomes very technical. Some knowledge of the research on the topic is needed to counter the wildly speculative ideas about the topic that abound in popular culture. The Wikipedia entry is a good source of this material, but at best a mediocre source of information on the state of science on the topic. See http://en.wikipedia.org/wiki/Subliminal_message.

To make a classroom demonstration of what seems to be unconscious perception you might try replicating Sidis (1898). He wrote a single letter of the alphabet on a sheet of paper and then showed it at such a distance that the observers could not see anything but a tiny black dot. When forced to guess they were above chance. The demonstration is less effective in a large room where people are at varying distances. It is better to have a "panel" of unconscious perceivers who are all at about the same distance from you. Walk toward them with a single letter on a sheet. When the first person is able to report the letter back up and begin showing more letters, asking them if they can see it, and asking those who can't to give an answer anyway. To make the demo work more reliably, do not have them just guess but give them four choices and ask them to pick one. They will do better than the chance level of 25%.

For some of the research and scientific debate on this topic see articles by Snodgrass, et al. (2004), Holender and Duscherer (2004), and Merikle and Daneman (1996; 1998).

How about awareness during perception? Goodale and Milner (2003) show that much of the machinery of perception is unconscious and unavailable to conscious introspection no matter how hard we try. Does this count as unconscious perception?
Certainly information about the world is being obtained but is never consciously registered.

6. The Self, agency, and the concept of "mine"

The concept of self has had many interpretations and many uses. David Hume (1711-1776) famously rejected the concept. Nevertheless, it refuses to go away. There are good reasons to use a construct of self beyond the intuitive feeling that we have (are?) a self. Some of these reasons are explored in the introduction to a special issue on the self in Consciousness and Cognition 2005, volume 14, issue 4. The 1999 book, *Models of the Self*, edited by Shaun Gallagher and Jonathan Shear, is also an excellent source on the topic. Here are some of the ways the concept has been used:

**Narrative center.** Dennett (1993) seems to be the person to coin this term, and it captures a lot about selfhood that seem essential to our psychological functioning. The self, by this account, would be a central point of organization of autobiographical memories, plans, sentiment, and moral or ethical rules that guide us through life. When we want to know ourselves, this is the memory file we go through. When we resist an unethical or immoral act, we in effect say to ourselves, "That's not the sort of person I am." This aspect of self also organizes our memories. The frame of reference for everyday memory is our autobiographical narrative.

**Actor and agent.** Searle (2001) points out that there is a formal reason for having the concept of self, the need for an action to have an actor. We don't simply act as the result of pushes and pulls in the social or physical environment. We act as if the self were the agent of the action, the actor. It would be nonsensical to have actions somehow be made by our bodies while we watch, or at least that would be a story quite different from normal volition.

There are pathological cases, such as the alien or anarchic hand in which the hand seems to have a mind of its own. Here the victims do watch their hands doing things they do not intend or control. Sometimes they need to fight the anarchic hand, even protect themselves from it.
Ownership. This concept covers things and actions that are part of the psychological self. The actions are ones we take responsibility for and the objects are, at the most basic level, body parts. We can also feel that things not part of our bodies are part of us. The rubber hand illusion takes place when a person sees a rubber hand being touched in the same way as his or her own hand, which is out of sight. Most people will come to think that the rubber hand is actually part of their body. Ramachandran demonstrated this vividly by threatening the rubber hand with a sharp instrument. The "owner" of the hand showed physiological responses that are elicited when an actual body part is so threatened. It is also possible in clinical cases for patients to deny that an actual body part belongs to them. In one case reported by Ramachandran a woman who was injured in an accident complained to the ambulance attendants that they had left a severed arm from a previous accident in the cot with her. The foreign arm was actually her own. The brain damage she had received in the accident caused her not to recognize her own arm.

7. Volition and Conscious Efficacy

Free will has been a long-standing issue in philosophy. How can the will be free if it is part of a deterministic biological process in the brain? There are a number of philosophical positions on this, which are reviewed well at http://plato.stanford.edu/entries/freewill/. See also http://plato.stanford.edu/entries/compatibilism/ and http://plato.stanford.edu/entries/incompatibilism-arguments/.

The importance of freedom of the will to a cognitive scientist or a psychologist is not so much whether we are truly free or not but why we all believe in it, implicitly in our actions, if not intellectually. Searle has pointed out that such a belief is necessary for action. If we did not believe in our freedom to act as we wished, we could not act. Further, every minute of the day we act without apparent constraint, and our belief in free will is thus continuously confirmed intuitively.
The debate on free will became more concretely scientific when Benjamin Libet in 1983 measured the brain's preparation for a simple action, which begins from _ a second to over a second before it is executed. This becomes a challenge to free will because people report that they made the "free" decision to move only about _ of a second before the action. The result has been taken to imply that free will is an illusion. The debate on the meaning of this finding is ongoing.

The question of conscious efficacy does not depend on any answers to the question of the existence of free will. We could be completely constrained and still the problem exists. How can a conscious thought result in an action. As I put it in one paper, how can an idea move a muscle? The impossibility of answering the question suggests that it is either metaphysically impossible to answer it or that the question is simply the wrong one.

8. Animal consciousness

Are animals conscious? This question touches on a number of issues. The first is how we would tell whether an animal was conscious. What sort of test would we use? The same question comes up in the next topic, machine consciousness. What is our criterion? If a dog yelps in pain, does that mean the dog is feeling pain, or does it mean that it has evolved a mechanism that produces what we call a "yelp" under certain conditions? Descartes rejected the idea of animal consciousness when he declared that they lacked the soul, which only humans have. This argument can be turned the other way as well. A materialist could argue that there is no soul and thus no qualitative difference between human and animal minds. Human brains have much in common with other vertebrate brains, and many of the same mechanisms for perception, action, and other functions. If the human brain creates consciousness, why not the animal brain? If animals are not conscious then what difference in the brain denies them consciousness?

Clearly an issue broached by this question is the nature of consciousness in humans. Does consciousness depend on language or human culture? If consciousness is
the product of evolution, then could we study "earlier" forms of it in animals? Also, if consciousness is the product of evolution, what is its adaptive advantage? That is, what is the function of consciousness? Can epiphenomenalism be refuted if a selective advantage can be shown?

For recent scientific work on animal consciousness, the journal Consciousness and Cognition had a special issue on the topic. This is Volume 14, issue 1, 2005. A strong case for animal consciousness is made in Griffin (2001), where many classic examples of behavior that imply consciousness are discussed.

9. Machine Consciousness

It is sometimes speculated that a computer simulation of consciousness would be itself conscious. For a number of reasons, this speculation is mistaken. First, a simulation of anything is just a simulation, it is not the real thing. As the philosopher John Searle has pointed out, a program that simulates the action of the digestive system does not digest anything. Just try stuffing food into the computer and see what happens! The simulation itself is a process that takes lines of code representing the input to the digestive system and produces lines of code that represent secretion of enzymes and stomach acid, extraction of nutrients, and so forth. A human operator must interpret these lines of code, which are arbitrary symbols that are designed as part of the interface to the algorithm and have meaning only in that context. The program itself is a system for manipulating 1’s and 0’s in a manner that obeys mathematical formulas that express the model of digestion. If you look at the process from the beginning to the end, you will see input 1’s and 0’s, 1’s and 0’s getting moved around by the computer in accord with the algorithm, and output 1’s and 0’s. This is so far from actual digestion only a person who knows the codes would have an idea that it was digestion and not, say, an internal combustion engine being modeled. To suggest that this simulation can be conscious would be to misunderstand the nature of simulation. The philosopher John Searle is responsible for stating the problem this way. It suggests that the idea that a
simulation can itself be conscious is the result of a misunderstanding of the nature of computation.

There is a further, possibly technical problem with the computer simulation of consciousness, and this is that we do not know how to simulate it. We would need to know how the behavior of a creature that is conscious would differ from the behavior of a creature that is not. Of course if we knew this (and cognitive psychologists are working on it), we would be able to make a good simulation, but it would be no more conscious that a bad simulation. A complete model of consciousness, if we had it, might at least make it possible to construct a conscious machine, just as complete knowledge of digestion might allow engineers to make an artificial stomach.

One could well ask if in principle it is impossible for a non-biological construction, whether a computer or robot, to be conscious. There is no reason we know of to preclude it. Daniel Dennett offered a thought experiment that supports the possibility. Here is a version of it. Suppose that we gradually replaced every part of a person’s brain, every neuron, every synapse, with a transistor or microchip that behaved exactly like that part and was connected to every other part exactly as in the original brain. The result would be a functional replica of conscious person’s brain and it should be itself as conscious as the brain it replicates.

Computer science is not particularly invested in trying to make conscious machines. The object is to get things done and ideas about consciousness may be used. The artificial intelligence approach to consciousness may have practical as well as theoretical benefits. One example is the software that uses an approach based on theories about consciousness. If consciousness confers an adaptive advantage on living possessors of consciousness, then some of its supposed functions may be used to make better computer programs. Such implementations serve as an indirect test of how consciousness may be functional in living organisms. Stan Franklin and Art Graesser (1999) report the results of their development of "conscious" software. Is the software conscious? This seems like a secondary and unnecessary (and probably unanswerable)
question. If ideas about consciousness may be tested out in this manner, we may learn a lot about consciousness.

There's another problem: how can we tell if a machine is conscious? Let’s assume we had made a machine we thought might be conscious. How would we know that it was conscious? Since we cannot in principle have first-person access to the experience of the machine, we have to infer consciousness from third-person observations. With a person, the way we would find out would be simply to ask questions such as, Are you conscious? Do you understand what it means to be conscious? With machine, this process could be maddeningly difficult. The machine could be programmed to give all the right answers to questions about first-person experiences. We cannot infer consciousness from nonverbal behavior, either. If the machine is designed to simulate conscious behavior there is no way of testing for the type of response characteristic of a conscious organism. It will always perform as though it were conscious because it is programmed that way, not because it is actually conscious. In short, there may be no way to tell whether a machine is conscious.

References and sources

**Some useful websites:**

Stanford Encyclopedia of Philosophy:
http://plato.stanford.edu/

Science and Consciousness Review
http://sci-con.org/

Mind Science Foundation
http://www.mindscience.org/index.cfm

Association for the Scientific Study of Consciousness
http://www.assc.caltech.edu/index.htm

David Chalmers' online papers on consciousness
http://consc.net/online.html
References


