The Mystery of Consciousness
By John R. Searle

There is no general agreement in the interdisciplinary field known as "consciousness research studies" (or "consciousness studies" or "consciousness research;" take your pick) on exactly what the word "consciousness" means. This lack has not prevented a flourishing of such research, especially during the last two decades, any more than the absence of a generally agreed upon definition of the word "life" has hindered the flourishing of the field of biology.

The situation for consciousness research is actually more extreme than that, reminding one of the proverbial story of the four blind men and the elephant. Persons claiming to be talking about the mysteries of consciousness or to have solved them often seem to be talking right past each other about some very different things.

This book contains reviews, originally written for the New York Review of Books, of six significant books or sets of books by major authors in the field. Additionally, it contains summaries of the views of the reviewer, John Searle, a professor of Philosophy at Berkeley and himself a major figure in the field. Together they cover many, though by no means all, of the differing views on the nature of consciousness and why it is a mystery, if indeed it is.

It is my hope that this book may serve as a sort of Cliff's Notes, providing summaries of the essential points in texts without having to read the original book entire. One thing it does offer that a Cliff's Notes can not, is, in two cases, sets of letters heatedly exchanged between the reviewer and the person reviewed following a review's original publication. I have read some but by no means all of these book reviewed books, and do hope that anyone who has read one or more of them will participate actively in our discussion and correct me if at any point my interpretation seems to be wrong.

Some divisions within the conscious studies community and how they manifest here:

A major division in the consciousness studies community exists on this question. If we could learn enough about brain functioning to completely describe and predict the entire chain of events from sensation and prior brain state through behavior and new brain state and do it every time would we then have created a complete description of consciousness. Some argue that if we were able to do this not only would we still not have a complete description of consciousness but possibly we would be no further towards one than we were before. Some claim this final knowledge will always be beyond our understanding.

Surprisingly perhaps, none of those who hold this latter view today do so because they believe in what is now known as "substance dualism," the idea that our physical brains are somehow connected to non-physical minds. In a very broad sense all persons I know of who are currently participating in consciousness studies debates are what were traditionally called "materialists." Why some of them would deny the possibility of completely understanding consciousness through traditional, "materialistic" scientific research and even refuse to be called materialists, sometimes throwing the word at their opponents as an accusation, is a matter much more subtle.

Philosopher Daniel Dennett has described these two groups neutrally as “the A team” and “the B team.” Psychologist Daniel Wegner has described them less neutrally as “the robo-geeks” and “the bad scientists,” using the insult that each group would most likely throw at the other as an identifier. The “robo-geeks” are the ones who believe that such a complete sensation/brain/behavior description, if could it be created, would be a complete description of consciousness. The “bad scientists” are the ones who think such a description would be insufficient and sometimes accuse the robo-geeks of actually denying the existence of consciousness even as they claim to study it.

In this book John Searle himself serves as a nice example of the latter group while accusing Dennett of being a member of the former, one reason for the heartedness of their included exchange of letters. In general, both serve, to me, as exhibitions of a type of thinking about consciousness that attempts to deal with seemingly fundamental issues which would exist as problems regardless of the detailed nature of the brain or the details of most behavior. Issues. The true relationship between subjectivity and objectivity and the possibility of ever doing scientific research on the latter is a common point of contention for these people.

Seemingly at the opposite end of adequacy for experimental research just now are the neuroscientists and clinical neurologists studying just what effect various regions of the brain have on consciousness and how they coordinate their efforts. In this book, Sir Francis Crick, Israel Rosenfield and to some extent Gerald Edelman seem to fit. However, brain scans and pathological dissections are not the only ways to study consciousness, even today. Some researchers believe that a more thorough understanding of the at a more fundamental biological level, elementary nerve nets within brain regions and the sub-cellular functioning of neurons themselves (possibly those of the other kinds of cells which together make up ninety percent of brain tissue as well) is needed. The conjectures of mathematician/physicist Roger Penrose and some of the work of brain scientist Gerald Edelman serve as examples of this kind of research in this book.

Finally, there is the most traditional of experimental study of consciousness, the study of the behavior of intact organisms (frequently college students taking Psychology 101) which was already going on in the laboratories of William James and Wilhelm Wundt well before the end of the nineteenth century. There are, unfortunately, no examples of this type of consciousness study in Searle’s book, but most fortunately the current (December, 2008) issue of the Scientific American contains and excellent example under the title of “Magic and the Brain” which I shall be referring to later.

No papers from the artificial intelligence community are included, yet the influence of that topic is pervasive. At the philosophical end of things the question of whether or not a computer could ever be conscious seems to come up about as often as discussions about objectivity and subjectivity. Artificial Intelligence does not seem to have much influence on brain region studies, though computerized analysis of brain scan data is often central to it. However, below that level computer modeling is making a big contribution with "artificial neural networks" having moved beyond brain research and into a number of applications, some of them quite unexpected. Furthermore, the power of massively multiprocessor computers (not artificially intelligent) is finally on the verge of permitting research on sub-cellular processes by simulating the interactions of individual atoms. Finally "cognitive simulation" of human psychology, originally named in the 1950's, is often derided today as GOFAI, standing for Good Old Fashion AI, but like God it seems to keep hanging around however many times it is declared dead.

Now on to the individual chapters in the book, chapter by chapter!

1. John Searle and "the Chinese room" *(*The Rediscovery of Mind and other works)

Searle's "Chinese Room" thought experiment may be the most referred to and most criticized such in contemporary consciousness studies. Fellow philosopher Daniel Dennett (quoted elsewhere in this book) may be correct that this is the only major idea that Searle has ever had, but even if this is true it still lifts Searle into the circle of major philosophers currently working on "Philosophy of Mind" issues.

The basic thought experiment is not difficult to understand. Many variations upon it have been presented by both Searle himself and (seemingly inumerical) critics of his over the years, but the basic version is the one presented in this book and is, I think, sufficient for our purposes.

To paraphrase one of Searle's own presentations of it, "Imagine that I (who do not understand Chinese) am locked in a room with boxes of Chinese symbols and rule books describing what I am to do with these symbols (my data base). Questions in Chinese are passed to me (bunches of Chinese symbols), and I look up in the rule books (my program) what I am supposed to do. I perform operations on the symbols in accordance with the rules and from these generate bunches of symbols (answers to the questions) which I pass back to those outside the room."

Again imagine the room says Searle but this time imagine that it contains a person fluent in Chinese who simply reads the passed in questions and, understanding them, simply writes out the answers in Chinese and passes those answers back out of the room. Searle's point is that something very different has happened in the room in each case, in one instance a clerk (or a computer) with no understanding of Chinese has created the output just by manipulating symbols according to rules. In the other case, the person fluent in Chinese and understanding to topics of the questions simply uses his or her understanding too translate and answer the questions, yet the outputs in each case may be identical.

Searle sees this thought experiment as a refutation of the so-called "Turing test," a thought experiment published by that chemist, mathematician and computer scientist in the British philosophical journal Mind in 1955. In Turing's thought experiment judges are allowed to communicate with parties to be tested only by Teletype or the equivalent.

In one version men and women all try to convince the judges that they are really men. In the version, which interested Turing more, real humans and Artificial Intelligence programs all try to convince the judges that they are really human.

Experiments of this kind have actually been carried out numerous times since the publication of Turing's article. See, for example, an article in the on-line magazine Salon few years ago by journalist Tracy Quan on Artificial Intelligence and her participation in such an experiment, link to: archive.salon.com/may97/21st/artificial970515.

In fairness to Turing, he did not claim that his hypothesized test would help in deciding the question of machine "consciousness." Neither he nor any of his contemporaries that I am aware of ever discussed that issue. What he and they were discussing in the mid-twentieth century was the possibility of machine "intelligence," which would seem to imply a strictly behavioral trait, not a subjective one. In the early sixties when Marvin Minsky and others at MIT coined the term "Artificial Intelligence" they defined it as referring to hardware/software systems which could perform acts "which would be described as intelligent if performed by a human." Implicit and intended in that definition was the idea that the *process* by which such acts were performed might be nothing like the processes that would be used by a human. Only the results were to count.

I find it interesting that at the same time that the topic of machine "consciousness" has gained respectability on the intellectual scene the earlier question of machine "intelligence" seems to have disappeared. Once hotly debated, it seems that no one much now wants to argue against the possibility of any sort of strictly behavioral "intelligence" being shown by computer-like hardware and software. For examples of arguments emphatically made before such skepticism disappeared dig up copies of philosopher Hubert Dreyfus books What Computers Can't Do and What Computers Still Can't Do.

Searle has now taken his argument against machine "understanding" much further in this book and elsewhere than he did with his original version of the Chinese room. He argues that, for example, computers can not really do simple arithmetic. What does happen, he says, is that computers are built or programmed to manipulate "symbols," and that it is humans not computers or programs that "understand" that the symbols being typed in or displayed describe numbers and operations to be performed on numbers.

In the customary vocabulary of natural language research, among other areas, the term "syntax" is used to describe grammar and other rules for forming and parsing sentences at the word level (i.e. rules about language). In contrast the term "semantics" is used to describe rules for relating language statements to the thing being described, the "meaning" of sentences in other words. However, Searle now insists repeatedly that since computers can never "understand" anything (by his definition) programs can only do "syntactic" processing and never "semantic" processing of any kind, a highly idiosyncratic restriction on the use of these words.

What does Searle have to say about consciousness then where it does show up, e.g. in brains? He answers repeatedly that brains "cause" consciousness because it is a "natural" product of (some) biological systems, just as digestion is a natural, biological function of a stomach. While the Chinese room argument would seem to counter top-down arguments for consciousness being derivable by writing programs to simulate externally observable behavior, it does not seem to me that it counters the opposite, top-down though experiment. One of creating simulated brains by simulating the interactions of the atoms that make up molecules and so on up.

”Up” in this case would include a simulation not only an entire brain but also as much of the rest of the nervous system, the body and it’s environment as necessary to reach a point where attachment to real world interfaces are possible. Some who have presented this argument have suggested that an appropriate real-world interface might be a humanoid robot with its sensors feeding into the simulated sensory nerves and the simulated motor nerves feeding into the robot body’s effectors.

Searle has countered this argument by saying that such a simulation would be only a simulation. Yet in other places he proclaims himself a “materialist” which presumably means that he is not a “vitalist” who believes that the physical laws governing biological systems are somehow different from those governing non-biological ones. At the same time he has said repeatedly that he believes consciousness might be caused by systems made of materials other than the normal biological ones. Might not one not say then that such a simulation might be simply a consciousness “causing” system whose materials are simulated atoms rather than real ones. Aspects of this argument will come up again when discussing Searle’s review of David Chalmers ideas.

2. **Sir Francis Crick and the “binding” problem** (The Scientific Search for the Soul)

Despite its arresting title, Crick’s book is a very mainstream example of contemporary, experimental “brain science” studies. In this case tracing the flow of information through various specialized brain centers (specifically the visual pathways) and attempting to prove or at least conjecture how the ensemble manages to behave, at least subjectively, as a single entity.

This is one of those books that may have been more important at time of publication for who wrote it than for what it contained. Sir Francis Crick, co-discoverer with Watson of the structure of DNA, was probably, especially to the general public, one of the most recognized and respected scientists of his time. The fact that he had now been devoting himself for some time to the scientific study of the relationship of consciousness to the brain legitimized such studies to a significant degree as something perused by reputable scientists and not by just by somewhat weird people out on the fringe.

The problem Crick was dealing with was one that has long been recognized in consciousness research. In the 1600’s the polymath philosopher Descartes tried to formalize his ideas about the relationship between mind and brain, the famous Cartesian dualism. He conjectured (and presented it only as a conjecture) that the interface where brain and mind connected to each other was the penal gland near the center of the brain. His reason for this conjecture was the knowledge that of all brain features then known only the penal gland was not lobed into left and right halves. To Descartes this was suggestive because he was sure through introspection that consciousness seems entirely unary despite the bilateral structure of most of the body with two arms, legs and eyes etc..

By the time Crick began his work on brains and consciousness much was known about the visual pathways from retinas through the occipital lobe at the back of the head and various mid-brain structures including the thalamus on multiple branching and intersection routes up to and within the cerebral cortex. In fact, more was known then and probably still is now about information flow within the visual system than about any other sense, making it an excellent focus for study.

The "binding problem" as Crick reduced it to a brain function problem was "the problem of how neurons temporarily become active as a unit." Other researchers had already suggested that the solution might involve the synchronized firing of neurons in areas responsive to different features of an object such as shape, color and movement. Crick and his colleague Chris Koch took this idea further and suggested that particular firing rates in the range of thirty-five to seventy-five cycles per second but most often around forty might be "the brain correlate of visual consciousness."

An objection can be raised (and Searle dutifully raises it).that what Crick and Koch seem to have discovered is an important mechanism underlying consciousness rather than consciousness itself. However, as more and more brain research indicates that consciousness resides in a dynamic network of interacting but highly distributed areas and processes it would seem to become increasingly hard to tell one from the other if indeed they can be told apart. On this point, I wish that these reviews of Searles' had included on of Marvin Minsky's book Society of Mind. This is because it would serve both as an example of a major book on consciousness by an eminent Artificial Intelligence researcher and also because its title. That title, it seems to me, so well captures the current view of 'mind" from brain science, a view so different from the unary one which we, like Descartes, seem to know so intimately through introspection.

Since the publication of Crick and Koch's Astonishing Hypothesis research findings have been very good to ideas of "temporal synchronization" as dynamic organizing principles in more and more types of brain activity, where in many cases emerging potential sources for attention competitively recruit other areas for temporary collaboration. For a very elaborate metaphor of how this works from a more recent book see chapter four, "Making Consciousation," in Rita Carter's Exploring Consciousness.

3. **Gerald Edelman, brain maps, robots & much else** (The Remembered Present etc.)

Gerald Edelman is a brain scientist in a broad sense, writing, researching and speculating on everything from molecular embryology through neural networks and on into brain area coordination and theories of conscious functioning including the linguistic and symbolic processes. Furthermore, he has where appropriate resorted to computer modeling including modeling of simulated robots, a whole area of consciousness research not previously in this book and unfortunately not to be discussed again.

Given the breadth of Edelman's interests it was necessary for Searle in reviewing several of his books together to skip around somewhat and I shall do the same. However, a recurrent theme in his work has been processes of self-organization from the neuronal through brain region co-evolution and on into the development of conscious skills such a language and ruminative self-awareness.

Searle, with what I see as unjustified reification of "consciousness" and its "cause," criticizes Edelman for failing to precisely delineate the point at which unconscious processes become conscious ones or explain exactly how or why that transition occurs. However, I believe that the lack of such explanations is precisely the point, that proto-conscious processes evolve from, with and into conscious ones in a coordinated and constantly shifting manner.

To begin at the bottom and be more specific, Edelman believes that brain structures are not genetically programmed to grow gradually but deterministically into some final form. Rather, he thinks, the brain comes equipped with an oversupply of neuronal groups some of which die out while others flourish due to self-organizing interaction with both stimuli from the outside world and each other. This Neural Darwinism (the title of his book on the subject) would be quite in accord with developmental processes of many organs in many different creatures. For example, a butterfly's wings while in the cocoon are initially solid, and the elaborate structure they exhibit on the emergent creature is the result of the dying off of the unneeded parts, a literal cutting or whittling away.

At the next level of organization Edelman, like Crick, takes on the binding problem, discussed in the last section. . However, where Crick's interest was in the binding of different feature recognition areas within the single sensory modality of vision, Edelman is most interested in the binding across sensory modalities leading progressively to the recognition and classification of objects repeatedly encountered. He does agree, however, that some primitive levels of self/non-self recognition are genetically built in. It would not do to have a baby chew off its fingers in the process of learning that they are part of its unary body.

In attacking the multi-sensory binding problem Edelman emphasizes what he calls "reentry mapping" but most would refer to simply as feedback from sheets of receptor cells providing more abstract representations of the individual's world back to those closer to the sensory sources which provide the more primitive and underlying representations. These "maps" which Edelman refers to are literal, physiological structures in the brain, more than thirty of them in the visual cortex alone. Through 'reentry" entire ensembles evolve together. Extensive work has been done on the feasibility of such process through computer simulations of such networks, and the importance of feedback structures in such networks has been amply demonstrated.

Edelman's group has gone beyond neural net simulations using simple sensory inputs such as black white imaging of pictures into nets which self-organize so as to recognize letters in a variety of typefaces. They have programmed complete simulated robots can learn coordination of simulated hands and eyes to explore their environments, and since this book was written such work has been extended into constructing real robots capable of exploring the real (laboratory) world and at times interacting with humans there.

From these approaches and results Edelman has moved up to conjecturing similar self-organizing hierarchies to explain the gradual acquisition of increasingly self-aware consciousness. Vital within this approach is a conception of memory as something constantly and dynamically organized as part of the evolving complex rather than any sort of separate and passive storehouse, hence his title The Remembered Present. We experience time and sequence directly at higher levels just as we sense motion directly at lower levels (all the way back to the retinal nerves within the eye!) and not as a deduction from a succession of still images. This dynamic view of memory will come up again when discussing the work of neurologist Israel Rosenfield.

.

4. Roger Penrose, Godel's Theorem and Quantum Computing (Shadows of the Mind)

In a nutshell, Penrose argues that the ability of humans to comprehend Godel's theorem in mathematical logic proves that all human thinking is not done using the Newtonian physics that underlies conventional chemistry but must utilize some sub-atomic properties of quantum mechanics. He further argues that this limit would also apply to conventional computers but perhaps not to future "quantum information processing" systems.

Since both Godel's Theorem and quantum computing are, to most people, very big and difficult subjects I feel like saying here, more than at any other point in this precis "go read the books," meaning both Searle's book and Penrose book. Still, here goes…

Kurt Godel developed his theorem in the mid-thirties. It proves that for any system of formal mathematics capable of representing both the addition and multiplication of positive whole numbers there will always be some *true* statements that can never be proven either true or false within the system itself. This only holds for true statements since, theoretically, any false statement can always be proven false by showing one particular instance where it is false. For true statements, by contrast, it must be possible to prove that every single instance is true, and for theorems about sets of natural numbers the number of specific instances can be literally, mathematically infinite. Imagine some theorem which states that for ever number it is true that (something or other) but requires one proof in the case of the number one, a different proof if the number is two and so on -– perhaps you can now get the general idea.

The way in which Godel proved this was to show a way in any such number system of constructing a statement that could never be proven either true or false because it was self-reverently contradictory. As a very simple example of such a statement consider this one, it says, "this statement is false." If it is true then it must be correct when it says that it is false, but if it is false then it must be equally certain that it is true, and so on ad infinitum.

I have a book on Godel’s theorem which I very much wish I could lay my hands on just now to give here as a citation. The reason is that it contains as an appendix a translation from the original German of the complete theorem done by Kurt Godel himself when he was a Fellow at the Princeton Institute of Advanced Studies, in the early sixties. The complete paper is about thirty-five pages long and like much in pure mathematics is actually more tedious than difficult.

What Godel demonstrated in that paper was a way (just one of many) to encode any finite string of characters in a finite alphabet as a number consisting of the product obtained by multiplying together a collection of prime numbers (numbers greater than one with no divisors except themselves and one) with the position of a prime in the sequence of primes starting with the number two representing the position of a character and the number of times that prime occurs representing the value of the character in that position.

Here’s an example: assume we have a two or more letter alphabet with the value of the first letter, call it A, being one and the value of the second letter, call it B, being two. Given those definitions the “Godel number” for the two character string AB can be calculated a follows: the first position is represented by a two, the first of all prime numbers, and the value of the character, A, in that position is one, so we have one two. The second position is represented by three, the second prime number the sequence of all prime numbers, and the value of the character, B, in that position is two, so we are going to have two threes. The Godel number for our string AB is therefore going to be 2 x (3 x 3) or 2 x 9 or 18, and given the number 18 we could work backward and unambiguously discover that the original string consisted of AB.

Having devised his coding scheme and assigned values to a minimal set of logical operators plus some variables what Godel did next was devise sequences of arithmetic operations which would do the same as the normal character manipulations in symbolic logic or, more precisely, second order prepositional calculus. That was the hard part. With all of this paraphernalia in hand he could then show how to represent a self-contradictory, self-referential statement of the sort exemplified a few paragraphs back and he was done. A way of always being able to create an unprovable and also undisprovable statement in any such number system or one containing it had now been demonstrated.

People have been using Godel’s theorem to try to prove that there are limits on what computers can do that humans are not subject to since at least the 1940’s. They have usually failed because behavior – human or machine – is usually imitated by simulation, not by proving abstract properties. Alan Turing himself disproved some of these early attempts that were already around in his time. Roger Penrose has given the best attempt to use Godel’s theorem in this way that I have ever come across. It has not convinced me, but it is long and subtle enough so that I could not reproduce or critique it without having Penrose’ book at hand to go by, which I do not currently have.

Backing away for a moment and back to the sort of issues that often preoccupy Searle it is never clear to me just what Penrose means by “seeing the truth” of Godel’s theorem. If he means a behavior such a reading a statement or set of statements and then printing out some kind of statement of the result, something that either a human or a machine, conscious or unconscious in the latter case, can do. But just what would those input statements and the output statement look like in detail. I do not know, and if Penrose does I wish he would publish them as examples in some new book.

If Penrose means “understand” in some subjective sense – which I think he does not, but just in case – then we are back to Searle’s argument from his first chapter that no computer can ever understand anything.

Now finally on to the second half of Penrose assertion, that the ability of humans to understand Godel’s theorem proves that they must be able to do quantum information processing of some kind. Quantum information is a very real topic currently being frantically researched by the US National Security Agency among others. This is because it seems potentially capable of decomposing very large numbers into their prime factors, and the inability to do so in any reasonable amount of time is the key (pun unavoidable?) to so-called “trap-door algorithms. These in turn are the basis of cryptographic schemes that allow someone to encrypt a message without a clue as to how to decrypt it and vice versa, and such codes are vital today to many both military and commercial applications,

Quantum information processing is not a candidate to replace your PC or even most scientific super-computers. Their physical set-ups look like arrangements for very expensive physics experiments and basically are. The ways in which they will be used when practical are likewise much more like physics experiments than like computer programming. A major problem in making quantum computing practical is “entangling” (I won’t go into that here. You should be able to sort of guess at the meaning I hope, and for exact definitions there are entire books out there!) enough atoms to represent enough q-bits (again, either guess or read) to do requisite calculations while keeping them separate from all other atoms for the entire (miniscule) time that the computations are going on.

Penrose thinks that water molecules isolated in the cytoskelatical tubules that help neurons keep their shape could provide suitable environments in which this could happen. When I first began to read an article by him on this I suspected that was where he was heading and was pleased with myself when proven right. Currently, many experts are saying that those microtubules would not be suitable for that purpose. More fundamentally, the quantum laws Penrose wants to depend on are not those currently known but new ones which he suspects may come to light from attempts to formulate “Super Gravity” unified field theories, another topic that I am not going to even try to get into here!

5. Daniel Dennett and consciousness rejected or not (Consciousness Explained et al.)

In this chapter of this book Searle states repeatedly and often condescendingly that Dennett does not believe in consciousness, that he rejects the idea of its existence completely. Dennett sees it differently, He states that he does believe in consciousness but that it's not what Searle and many others intuitively think that it is. He states this not only in the books of his that Searle reviewed but also in an exchange of letters between them following Searle's review which, wonderfully and for a wonder, Searle includes here immediately after the copy of his original review.

What Dennett believes and Searle disagrees on seems actually to consist of two different things. One is epistemic, i.e. a question of what can be known accurately and how. Dennett holds to the behaviorist-methodology, which as Searle takes pains to point out was also held by Dennett's philosophical mentor Gilbert Ryle, which claims that reliable science can only be based on things observable in the third person, preferably by several persons who can then compare their observations. This puts him very squarely and proudly on the "A" or "robo-geek" team explained in my introduction, the group that believes that if brain and behavior could be completely explained that would also amount to a complete explanation of consciousness.

In detail and in practice, what this methodological restriction says is that when we study consciousness we never actually do so directly. Instead, it says, we *infer* the consciousness of experimental subjects from their "verbal behavior," brain scans etc. We should always make this clear in research studies wherever there might be any confusion, which in practice seems to happen quite seldom. What this "methodological Puritanism," to use a term coined by one of its critics, excludes from "good science" is any first person report by the experimenter of anything supposedly learned directly from his or her introspective experience. To include such experiences "scientifically" they must be reported and evaluated in exactly the same way that a third person report from any other experimental subject would be.

Most of the time of course, almost all of us including committed behaviorists "intuitively" (as Dennett would probably say) use a mixture of approaches. We infer the subjective states of others by comparing them to real or imagined subjective states of our own. Then we assume, unless the evidence forces us to think otherwise, that others have functioned internally and subjectively pretty much as we know that we would have from our direct and introspective examination of ourselves. Recent neural research on both humans and animals has shown that this process appears to occur at a very basic and often unconscious level in so called "mirror neurons" that fire as if we were performing an act when we see someone else, even of a different species, doing so.

Dennett does not deny that we use such methods or that they are frequently effective and have stayed with us through evolution precisely because they so often are. He simply denies that they constitute a valid "scientific" method of drawing scientifically accurate conclusions. Searle, in contrast, seems to think that subjective experience provides the most certain evidence that we have about consciousness. This is not because it always provides us with an accurate representation of the outside world but because we can not deny the reality of our own thoughts and sensations as things that we have subjectively experienced.

This is emphatically not a new issue in the study of consciousness. Descartes is still remembered a third of a millennium later for having said, “I think therefore I am.” Unfortunately (and to me frustratingly) this is misremembered as being the heart of his philosophy. In fact it was simply the start of in example in chapter four of his Discourse on Method of how his “method for reaching reliable conclusions” (which *was* the heart of his philosophy) could be applied to the problem of metaphysics.

In discovering this, as something that could not be doubted (the first step in his “method” he was specifically contrasting its certainty with all other seemingly true but potentially fallacious beliefs. An example was the belief that he had a body and that there was an outside world rather than, for instance, being a spirit, alone in the universe who was having a dream. (Proving that there did exist in the universe something besides himself was the next step in his example.) In other words, he was concluding only that *something* existed, not at all the nature of that something. This leads us nicely to the second of the two points about consciousness on which Dennett and Searle disagree so violently.

The first point of their disagreement was/is epistemic, how to gain accurate (be a modern and say “scientific”) knowledge about consciousness and how to make sure that it is accurate. The second point of contention between them is ontological, i.e. is consciousness “real” and if so what does it mean to say that it is real. They do at least agree that these are the two points in contention and use the same words for them!

The reality of consciousness and what that means is a slippery issue now just as it has been for centuries for people trying to understand Descartes. When it comes to Dennett’s views on the matter I’m inclined to say don’t trust me too much; go and read his books, preferably several of them. Fortunately, in this little book we also have the exchange of letters between the two men to go by. These are two men, both considered to be currently eminent philosophers in the area of philosophy of mind, really seem to genuinely despise each other, perhaps somewhat because they both are considered so eminent. Forgetting the relevance to the topic of consciousness, these letter might be read for fun simply as examples of how bright intellectuals can sometimes go at each other in public like small boys in a schoolyard and apparently both enjoy it.

What Searle believes about consciousness seems to be easier to understand, he believes it to be something unary and fundamental, not to be doubted by anyone except as an example of “intellectual pathology,” which is something he explicitly accuses Dennett of at one point in this published review. I always get a bit edgy when I come across someone emphatically asserting that something can not possibly be doubted by any sane person. I lean toward the modern methodological doctrine of falsability, that something which can not even potentially be proven to be false can probably not be shown to be true in the usual sense either but is perhaps an artifact of our language or some such. I do however quite agree with Searle that those facts seem intuitively, unquestionably true.

Perhaps surprisingly, Dennett agrees with that, he simply holds that such “intuitions” (he uses that word) however hard to shake are in fact wrong.

What does Dennett believe then about consciousness? For one thing he believes what he calls his “multiple drafts” theory about it. If I understand this (and I am not at all sure that I do) it takes the known fact of neurologically distributed consciousness processing in the and conjectures that it applies to subjective states as well. This is not the sort of conscious/unconscious distinction of Freud as I understand it; it seems to have more similarity to the known separate minds within split brain patients, the verbal on the dominant side and the silent but visually and manipulatively acute one on the other. Dennett himself describes the separate conscious states he is talking about as being like successive drafts of an article, hence the name. How it is that we have the illusion of a single stream of consciousness with no perception of these concurrent drafts either as they are going on or after one is selected and others discarded (if indeed that is what he thinks happens) I do not understand. However, it does seem to rule out their being like different personalities in a person with multiple personality disorder since in those cases some of the personalities are persistent and very much aware of each other even though only one may be in control at a time.

The final point about Dennett’s ideas that arouses Searle’s wrath is his rejection of the possibility of “philosophers’ zombies and alleged acceptance of the possibility of “strong AI.” The philosophers’ zombie is a common creature in consciousness researcher thought experiments. Supposedly he or she would be like us in *all* ways including showing emotions and talking about consciousness, yet they would in fact be totally unconscious and reactive or robotic. Some who have considered them accept the idea that they could in theory exist, though I know of no one who has ever suggested that they exist in practice; others have argued that without consciousness such behavior would simply be impossible. Dennett argues that any such being, exhibiting such behavior would inevitably *be* conscious, which seems at the least to fly in the face of what Searle’s “Chinese room” thought experiment, discussed earlier, is supposed to prove.

Searle also claims that Dennett is a believer in “strong AI.” Strong AI claims that it is possible to build computer hardware/software systems that are truly intelligent or conscious. “Weak AI,” claims that such systems, however they might appear would only be “simulating” intelligence or consciousness. It is logically possible to split the difference of this, to claim for example that computer systems could be genuinely “intelligent” (strong AI) but never do more than “simulate” consciousness (weak AI). Searle however, believes that neither consciousness nor genuine intelligence can ever be manifested, at least by the sort of computer systems we now know. His reasons for those beliefs I have also tried to explicate in my earlier discussion of Searle and his Chinese room.

I have never come across a passage in any of Dennett’s works that I have read or scanned which seemed to explicitly either accept or reject strong AI as the hypothesis relates to either intelligence or consciousness. However, I suspect that Searle is right and that Dennett would or does accept it in both cases. What is certain is that Searle claims to be sure that Dennett accepts the strong AI possibility as it relates to consciousness. He seems to take this as a final proof in his argument that Dennett does not really believe in consciousness at all, however paradoxical that may seem, because he claims to believe that Dennett has a severe case of "intellectual pathology" as mentioned above.

6. David Chalmers, "the hard problem" and panpsychism (The Conscious Mind)

Given that Searle so dislikes Daniel Dennett, a leader of the "A team robo-geeks," one might expect him to like David Chalmers, a leader if the "B team bad scientists," but not so. This chapter on Chalmers is, I think, the worst in this entire book, saved somewhat by the inclusion of a response by Chalmers to the review at its end. For an excellent summary of Chalmers views see the transcript of his dialog with Susan Blackmore in a book we discussed in this group a few months ago, Conversations on Consciousness.

In this chapter Searle manages to drastically misstate not only Chalmers views but also some rather basic facts about both behaviorism and functionalism as twentieth century intellectual movements. Specifically and to start with, I have never come across anything by or about any behaviorist denying the existence of consciousness, though I have come across that charge many times. What behaviorists did consistently deny, rightly or wrongly, was the legitimacy of subjective experience *as such* as valid raw data for scientific investigations, a point just elaborated on with regard to Daniel Dennett.

Functionalism, which is still around, abstracts analogous properties from different situations and tries to explain those similarities as alternate ways of achieving analogous functional results. Searle mentions electric and mechanical clocks in passing and showing how similar sub-functions in such physically different devices are achieved and why, for example, explaining why both need of some actuator with a very steady beat, whether it be a pendulum or an alternating current from a wall socket. Functionalism demonstrates correlations not causation between examined systems. Mechanical clocks with their pendulums did not cause electric clocks that plug into AC outlets, rather both needed parts of some kind to accomplish an analogous function.

Chalmers is most noted for having coined the phrase "the hard problem" to contrast the difficulty of explaining why we have consciousness at all, rather than being consciousless philosophers zombies of the sort just previously discussed, with the "easy problems." Relating conscious perceptions and intentions to external stimuli and behaviors are the (relatively) "easy problems" by his and most others' standards. Curiously, Searle never even mentions "the hard problem" when discussing Chalmers book, and I wonder why.

In his response to Searle, printed in this book, Chalmers describes himself as being an "agnostic" about most aspects of the mind-body problem, and a very thoroughgoing one he is. To return to the philosophers' zombies who behave like us but without any speck of consciousness Chalmers concedes that we can imagine them but asks whether they are possible in this physical universe and if not would they be possible in some other universe with different physical laws. The belief that some assemblages of matter can exhibit "mental" properties as well as physical ones is what is now called "property dualism," and as a way of trying to simultaneously hold on to both materialism and consciousness has been around for some time. Bertrand Russell proposed such a "dual aspect" theory in explicit analogy to the wave/particle dualism thus then being elucidated in quantum physics as far back as the 1920's.

In investigating just how far down such dualism might be able to extend, beyond the level of lower animals Chalmers asked whether inanimate objects might also have a conscious aspect of some sort. In the book reviewed by Searle as well as in other places he has tried to imagine just what the phenomenological life of a thermostat might be like, very simple he concluded. Speculating still further, he asked whether consciousness might be a fundamental property of this universe like space and time and coextensive with them. This is where the panpsychism comes in: if everything including empty space might have a conscious aspect of some sort wouldn't that imply a universal consciousness and why would we not be aware of it?

Such a consciousness for most of the universe would be very uninteresting he felt, but when entangled with "informationaly complex" entities such as ourselves or even lower animals it might manifest as the sort of consciousness that we would recognize as such. Just what defines informational complexity and why consciousness would become entangled with it does not seem clear, and Searle is quite right about that. However, Chalmers does not present this as a conjecture, merely as a logical possibility that seems hard to exclude simply on the basis of logic and challenges us to discover just how it might be rejected. In sum, Chalmers does not affirm so-called "substance dualism" (minds and brains connected but composed of different substances) or mentalism (the idea that everything is really mind with matter being merely an illusion). However, in his demand for good reasons to exclude such possibilities he examines them more closely than anyone else I know of doing consciousness studies today from a secular perspective.

7. Israel Rosenfield, the Self & Body Image (The Strange, the Familiar and Forgotten)

Rosenfield, a clinical neurologist, uses different source materials from others reviewed by Searle, clinical case histories of the results of various forms of neural damage, and uses them to address a somewhat different topic, not consciousness as such but the normal, seemingly continuous sense of self. This is a topic with a long history in consciousness studies, Psychologist William James was writing about it more than a century ago and with the return of interest in consciousness studies within the last couple of decades a number of researchers have grappled with it, usually with difficulty. Some have considered it an achievement to prove to their own satisfaction that there is no such thing as the self, often relying on common experiences reported in meditative states as evidence. Daniel Dennett has suggested that while the self seems real it is in fact an illusion, a sort of virtual, sequential machine that runs atop the multiprocessing brain.

Rosenfield sees a coherent self as being a product of dynamic memory built around changes and stabilities in ones ongoing, physical body image. This claim of body image as central to sense of self and even more is not new.. Hubert Dreyfus in his What Computers (Still) Can't Do books, previously mentioned, centered much of his argument on the assertion of Nazi philosopher Martin Heidigger that not only consciousness but intelligence requires the existence of a body. Likewise, a number of AI researchers have felt that providing real or simulated robot bodies to their computer programs was essential if those programs were to be able to learn/evolve in any deep sense.

Developmental studies of infants both before and after birth have usually reached similar conclusions, and psycho-linguist George Lakoff has tried in a rather large book to demonstrate that even seemingly very abstract branches of mathematics rely on, often unconscious, body image metaphors for more than is customarily realized. The negative evidence, if any, regarding the importance of body image and coherent sense of self to intelligence and consciousness seems to come from recent studies of memory. The "episodic" memory, which is what one thinks of when thinking of cases of amnesia appears to be only one of several types which seem to be associated in differing ways with differing brain regions. Others commonly recognized today are the semantic, the procedural and the emotional as well as the working memory, which does play such an important part in Rosenfield's examples. Few seem to consider in the case of the proverbial amnesiac of Hollywood films how it is that the person who has forgotten even his or her name is nonetheless able to use their native language and tie their shoes without difficulty.

Conclusion/addendum, findings from modern psychologists' consciousness studies

While Searle's reviews unfortunately do not contain any examples of conscious research carried out by psychologists, we are fortunate that there is an excellent example of such research in an article entitled "Magic and the Brain" in the current (December 2008) issue of Scientific American. Furthermore, it seems to dovetail nicely with the matter that Searle brings up in his conclusion as an important topic for future conscious research, the phenomenon of blindsight. This is a neurological condition in which an individual can not consciously see out of all or part of his or her visual field yet can reliably answer questions about things which are going on in that "blind" area such as "how many fingers am I holding up?" The blindsighted person will insist that he or she can not see what is going on in that area, but when asked to "guess" will usually "guess" correctly and be very much surprises at how often those guesses turn out to be correct.

The neural areas and pathways whose damage produces the blindsight phenomenon are now pretty well understood, but a relevant question I want to bring up that Searle seems to have missed. Who if anyone in this person's brain is conscious of what is going on in that reportedly blind area? Not the self doing the verbal reporting obviously, but does that mean that there was no consciousness in the blindsighted person's brain aware of what was going on in that area?

Remember the results of the tests that can be performed on persons whose interhemispheric connections have been cut. Only one side is able to speak but both sides seem to be able to think and to feel emotions. Remember too the questions raised by both Daniel Dennett and David Chalmers about just how many consciousnesses may be simultaneously active in even a normal brain. Remember finally that in the case of the split brain patients the speaking side usually has and believes reasonable sounding but factually quite incorrect interpretations of behaviors initiated by the other side, other half.

In the Scientific American article an experiment on perfectly normal people, that seems nonetheless to remind one of those split brain experiments. In the experiment described in the current article each subject was shown a pair of photographs and asked which one he or she found more attractive. The photographs were then covered momentarily and switched by the experimenter, using slight of hand. Each subject was then asked to explain why he or she found the person in the selected photograph (really the person in the photograph *not* selected) to be more attractive. Only about one forth of the subjects realized that the photographs had been switched, even though they had been out of sight only momentarily. Of those who did not catch on to the switching almost all were able to give detailed, introspective reports of why they had just made the selection that they did, even though they had, in fact, just made exactly the opposite selection.

Four other types of “cognitive illusions are also described. Change blindness: a viewer misses changes made to a scene during a brief interruption. I have seen some extraordinary examples of this in books on the subject where two photographs were displayed side by side. Repeatedly, I was unable to spot what, after the fact, were indeed obvious changes from one to the other even though the only “interruption” was the movement of my eyes from one side of the page to the other.

Inattentional blindness: a person does not perceive items that are plainly in view. Example: a person in a gorilla suit wanders across a scene of a group playing basketball, stopping briefly in the middle to pound its chest, but goes completely unnoticed. Yes, this is a real experiment that has been done repeatedly with variations.

Choice blindness: the experiment with the photographs, just described, is an example of this one, and finally illusory correlation: a stage magician waves a wand and a rabbit appears.

One I find even more dramatic was an experiment in which an experimenter with a clipboard pretended to be taking a survey. Part way through, two men, also part of the experiment, carrying a large board passed between the supposed interviewer and the person being interviewed. Out of sight, the initial interviewer swung up to hang behind the board and a different interviewer, differently dressed swung down and, once the board had passed, continued the interview as if nothing had happened. Incredibly, when asked about the incident just afterward more than half the subjects had not noticed or could not remember that the switch of clipboard wielding interviewers had occurred!

What I take from all these examples as well as the blindsight and split brain results is that Dennett is right, though perhaps not in the sense he meant it, that consciousness isn’t what you think it is. Also, that the difference between consciousness and unconsciousness is not nearly so clear cut as Searle assumes.