Consciousness and Semantic Knowledge:

In having thought about 'information theoretic' ideas of consciousness, certain ideas have struck me in their prescience. The questions asked in contemporary theories of consciousness center on why certain patterns of neuronal firing give rise to the subjective experience of consciousness, while other patterns of firing do not. Several metrics from network theory have been proposed, and I think that these have considerable downstream implications with respect to the nature of semantic information.

What I hope to describe below is a possible connection between the subjective consciousness experience and semantic information.

There is much discussion on consciousness by both philosophers, neuroscientists, psychologists, and a slew of other disciplines. To discuss consciousness, I mean that subjective experience *you* are experiencing at this moment. I mean the way sensory information is being abstracted to a level at which it becomes experienced. Consciousness, as used throughout this essay, is that experience you are having now that you do not have when you are in a deep and dreamless sleep.

The crux of my argument will ultimately lie on motivating information theoretic measures of consciousness, considering the natural implications of such theories, and motivating a testable hypothesis on the nature of semantic knowledge based on these implications

Background:

Modern theories of consciousness seek to provide the empirical rules under which conscious experiences can be expected to arise. Associated questions are about identifying and understanding the scale which conscious processes correspond to, known as the scale problem of consciousness[1]. Neurophysiological evidence suggests that conscious experiences co-vary with information encoded in coarse-grained[2] neural states.

To begin I hope to establish some common ground. First, I would argue that advanced animals have semantic knowledge, that is, they have neural representations of persons, places and things. I mean to distinguish knowledge of concepts from language, which is more analogous to a mapping of semantic knowledge to symbols, which might follow rules, and be used to communicate to other beings. I would argue that dogs have the basic semantic representations for what a ball is, what a food bowl is, what a leader is, and other concepts as well.

Given the difference between these items, it then seems natural to note that, by-and-large, semantic knowledge predates that ability to communicate it.

---

[1] It seems that only information processed at certain levels seems available to conscious awareness. We do not have direct experience of neurons firing, nor do we have experience of macro-level interactions such as interpersonal communication.

[2] For those familiar with this term, I do not mean it here in a strict sense (I would allow for 'black-boxing' of information as well). It is simply stating that when large parts of the visual cortex are active, it is usually accompanied by subjective conscious experience. The simplest example of course graining might be system in which there is averaging of inputs.

Semantic knowledge/concepts predating language seems evident from consideration of the question: "why are people so adept at learning language". While animals do have communication, there is a striking advancement in the complexity and nuance of human spoken/written language compared to all other species. The relatively short period of selective pressure for our language systems begs the question of what led to this? I think viewing semantic information as a "fall-out" of the conscious process, which has been possessed by animals for much longer in evolutionary terms, leads us to a natural answer; the semantic concepts have always been represented in the brain with mechanisms for storage and retrieval! All the brain was lacking was a mapping of semantic concepts to phonological forms!

A word of caution: this argument might require a jump. Consciousness requires retrieval of semantic information

If we look around and notice the color of an object, how is it that our mind even knows the property that is color? The answer is natural, we've had subjective experiences of millions of other objects of which our brain has determined the relevant co-varying attributes. While I think there is little debate that object recognition can trigger semantic retrieval, I mean to more intimately tie together the way "real-time" sensory information is processed and semantic knowledge, even outside the context of explicit object naming or recognition.

I would argue that incoming sensory information needs to be processed in a way that it is meaningfully compared to, and held in relation to, other conscious experiences we've had.

An implication of what I am arguing for is that when we see a car drive by, by the act of experiencing a car driving by, we have recapitulated the semantic information of "car" and "drive". It is important to re-emphasize that I am not saying that we retrieve this knowledge when we say to ourselves "a car is driving by", but that implicitly, every second of our waking experience, sensory information is being abstracted to some level until it is at a level of semantic information. I would argue that if we use the word "car" in the above situation, that is a qualitatively different subjective experience, and semantic representation, than if the same person had thought that was "a blue Honda Civic that drove by". Consciousness, which is fundamentally interested in how the sensory inputs become abstracted, has implications for semantic knowledge.

I think that there is little question that reading about a car, hearing the word car, and hearing a car drive by can all go on to activate the semantic information underlying car. What I mean convey is that the essential semantic information for 'car' is only that information which is associated with a conscious experience (i.e. not the specific mechanism by which is was activated).

While hearing, seeing, and experiencing a word all have different neural mechanisms and can be distinguished in imaging methods such as fMRI, I'm arguing that the question of "what is the semantic information of concrete and abstract words" bridges into phenomenology. To take another example, while it can be activated in different ways, the underlying information of the meaning of 'red' is most closely identified with the experience 'red', which is intimately tied to consciousness. The question of "what is the information contained by red" is closely tied to "what is the kind of information we

experience". While we know that cone's firing can lead to the experience of red, one can also imagine a red wheelbarrow, with no cone activation. Further, red cones can activate, and we can attend to a red stimulus without conscious experience of the color. Teasing apart the necessary background conditions in order to have the experience of red from the conditions giving rise to experience itself is at the heart of theories of consciousness, and I argue, should be for semantic information as well. The semantic knowledge of a color is not entailed from feed-forward processing of inputs, but requires a 'top-down' activation of primary visual cortices.

The information for the concept of red cannot be distinguished from the experience.

Neural Correlates of Consciousness

For those open to the idea of neural correlates of consciousness, they can likely skip the below section. I hope to address why it might be reasonable to suppose that in some relevant sense, there are some minimum and necessary conditions for the subjective experience to arise.

First, some might propose that consciousness is an emergent property, and is the result of millions of neurons firing and where asking the question of localization is ill-defined or non-sensical. For this, there are strong counter arguments. First, of the 90 billion neurons in the brain, 2/3rds of them lie in the cerebellum, and while there are 4 times as many neurons in the cerebellum as in the neocortex, we know that they play little (if any) role your current conscious experience! Case reports of cerebellar agenesis (with no cerebellum at all) appear to have the usual experience of consciousness! For all those neurons, none of them give rise to the subjective experience of consciousness. A similar argument can be made for the gut, and other autonomic systems as well.

Excluding the cerebellum, one might think that more recently evolved structures like the prefrontal cortex should contribute to the conscious experience, but again, years of lobotomies and other case reports have shown that while you lose emotional affect, the subjective experience of consciousness you and I are currently feeling is as present as ever. This argument continues for the visual cortex, temporal lobes, brainstem and many other structures. While it seems it could be distributed over all these regions, we know that these regions still fire and are active in a dreamless state (or under propofol sedation), so any theory of consciousness would be left asking: 1) Is there any part of the network required to be active for the conscious experience (perhaps integrating information) and 2) what is the specific group dynamic that gives rise to the conscious experience, as having the active proper subcomponents is not sufficient.

I would also make a quick mention that dreams are perfectly valid conscious experience, so it does not appear necessary to have active sensory input to be in a conscious state.

Moving forward with the above idea, if one accepts that the relevant parts of the brain need to be yoked together in the proper way, then it is natural to ask what is the mechanism by which the brain becomes yoked in the proper way? As it turns out, this question is investigated and there is a literature with current evidence for certain posterior hotspots (see Tononi for a review).

I think the above arguments culminate in the ideas behind modern theories of information theoretic models of consciousness. Largely, that consciousness is ascribable to a type of information which arises under certain conditions.

I am not interested in how the abstracted information ultimately *feels* (the phenomenology), but I think there is sense in which the right question to ask is "what is the sort of information that is felt, and what parts of the brain create that information". In seeking an answer to this question, I think that we will inevitably cross paths with understanding semantic knowledge.

A Brief intro to IIT

Much of the above discussion is taken directly from the works of Giulio Tononi. In light of all that has been said, he has proposed a measure for the quantity of consciousness (φ) and a way to describe the qualia of an experience. The details of the theory are beyond the reach of this essay, but I hope to make some approachable presentation based on his writing with an emphasis on how I am proposing it relates to semantic information.

It is worth noting here that in recent years there have been competing information theoretic theories of consciousness which have strengths and drawbacks. However, information theoretic ideas share a similar frame-work and for the purpose of understanding semantic information, I will largely stick to the most well-known theory which is IIT[3].

Implications of Information Theory Consciousness:

1) All semantic information is experiential

IIT is centered on the idea that it is integrated information that is ultimately felt. As such, it is necessary that for something to be an experience, it cannot be a simple feed-forward process. Based off ideas of IIT, the kind of information that our brain is adept at handling is integrated information. In imagining that there is a continuum of processing, if we want to generate the proper integrated information (access a concept), we necessarily need to readily generate the kind of information that is available after the feed-forward (FF) process. This asks give rise to the investigation of questions like "what is the minimum activation necessary of visual cortex to experience color". I do not know the relevant literature in this area.

2) There is no natural distinction between concrete and abstract words

In IIT, it seems natural to define the meaning of a word along the lines of "information that is contained in and above the individual sensory input that goes into that word". For this reason, the fundamental nature and way abstract words are stored should be the same as for concrete. The reason for this is because what is important about a semantic concept (or what is subjectively felt) is information that exists after any strictly feed-forward processing. For visual, auditory, tactile input,

[3] Other theories include: Information Closure Theory of Consciousness (which I do reference later), Pennartz's neurorepresenational theory, Intermediate Level Theory of Consciousness, Predictive Processing, Sensorimotor Contingency, and Global Workspace Theory.

processing which we cannot be conscious of cannot contribute to our semantic representation. Thus, words that have little associated feed-forward processing (i.e. abstract words) are on the same ground as concrete words.

This idea seems plausible as it is well known that most of our vocabulary is made up of abstract words. In fact, most of the words in this exact sentence are abstract! We know that there is no computational difficulty in dealing with abstract words over concrete. Further, this would then imply that there is no reason to think that abstract concepts should be later to evolve as conscious processes make no distinction between concrete and abstract semantic information.

3) All words require widespread topological distribution over the neural correlates of consciousness

This is implicit in the above discussion. If there are parts of the brain that need to be "yoked together" to give rise to a conscious experience, then I am implying that semantic meaning is necessarily distributed over this entire network. I don't mean to say that the information can't be activated from other 'top-down' or 'bottom-up' processes, just that the core information of the semantic concept doesn't exist in those mechanisms.

4) Words should have minimal mutual information with the underlying sensory information

This is more deeply connected to a notion of informational closure. In a different but related theory of consciousness, there is a notion of information closure. The argument here is that for something to enter the conscious experience it must have an internal representation that is compatible with and predictive of the outside sensory information. This means the current state of the system should be highly predictive of the next state of the environment. Here, rather than integrated information being felt, it is information that a system uses to model the environment that is felt. In this sense, since reflexive behaviors simply respond to the environment, but don't predict the environment, they do not give rise to subjective experiences. This theory then predicts that conscious processes will have minimal mutual information with the environment as a process which is able to predict itself entails most of the information in the current state being dependent on the previous state, rather than environmental input.

5) Nouns, verbs, and other parts of speech have the same neural substrate

As said above, I am proposing that the underlying information that we mean by semantic is that which can be 'experienced'. As such, verbs have meaning, but prepositions and syntax have no sense in which they are 'felt', so while they are involved in language, they have no semantic content and cannot be involved in semantic processing.

Minor Implications:

a) As stated previously, IIT predicts that strictly FF NN's will not contain any significant semantic information. In this way, perceptron-based modelling of word meaning, along with any non-recurrent architectures in the brain (i.e. cerebellum) should not contain semantic information.

b) When there is sensory input, but we are on “auto-pilot”, the incoming sensory information is not being abstracted to semantic representations. Much of the function is being done as a reaction to the environment in a way that is modelled by a FF neural network. Accessing semantic representations will always require top-down activation as integrated information doesn’t arise from only FF processes.

Closing:

I think that to answer the core question: why is language so natural for us? We must consider that proper abstraction of our sensory inputs has had an intimate role in life-forms for millions of generations, and that understanding what information is “felt” is understanding the basis of semantic knowledge. Somehow, it seems that whatever consciousness is, it has been an emergent property of life, so it would seem critically interwoven with processes built on top of it.

In light of the above discussion, I think the reason the study of conscious experience in the brain has been so elusive and unseen in fMRI studies is the difficulty in obtaining contrast from tasks. Except under propofol sedation, we are always conscious in the scanner and thus creating semantic information. While we can observe and study the control processes involved in creating the right conditions to generate the proper inputs to the parts of the brain that give rise to the subjective conscious experience, the study of the information which “we are conscious of” will require novel paradigms, and even then perhaps, might be outside the reach of fMRI. It might well be that the information of consciousness is widely distributed over the cortex, but understanding the nature of the information that is felt is what I mean to hone in on.

I hope to have motivated in the above discussion, a possible avenue the lab might let me explore the for the study of semantic representation.

I am also assuming that there is no such thing as a context independent representation of a word. Or that any one representation captures the full meaning of a word. The best we can do is have a continuous vector which has some probability of mapping onto a word stored in our language system, but there is no reason to think that the experience itself should fit nicely into a word