Yields falsehood when preceded by its quotation

Term paper submitted as a part of HUL281

Harshad Deo, 2010ME10675

Piyush Ahuja, 2008MT50454

Abstract

The central problem in the study of mind, machines and language, the underlying theme of the course if you will, is that of consciousness. Despite widespread ambiguity about the definition of consciousness in the scientific community and among philosophers , there seems to be a commonly shared intuition of what it refers to. In this paper, we attempt to tackle the problem of consciousness in such a way so as to gain valuable insights into the most profound question pertaining to mind, machine and language.

We explore the relationship of consciousness with self awareness through thought experiments like extending the mirror test to machines and chatbots in conversation with each other. In the section on dreams, we use a model to explain observations from dream experiences, taken over a period of two months over a variety of subjects.

A Priori

1. There exists a physical world independent of and governed by rules unprejudiced by the existence of humans and other entities capable of postulating such statements.
2. These entities receive sensory information from this external reality, and are able to influence it through some form of actuators
3. These entities are capable of storing and processing the input and using the results of the computation to generate output.
4. These entities are dependent resources obtained from the physical world for survival and reproduction.

For clarity of subsequent discussion, and in an attempt to avoid stilted terminology, all entities as described above will be referred to as humans and their functional structure described in terms of human anatomy. Therefore, their information-processing module will be referred to as their 'brain', and so on.

Defining Consciousness

"What is an "I", and why are such things found (at least so far) only in association with, as poet Russell Edson once wonderfully phrased it, "teetering bulbs of dread and dream" -- that is, only in association with certain kinds of gooey lumps encased in hard protective shells mounted atop mobile pedestals that roam the world on pairs of slightly fuzzy, jointed stilts?"

- Douglas R. Hofstadter, Gödel, Escher, Bach: An Eternal Golden Braid

In *Animal Consciousness*, Daniel Dennett had remarked that it often passes for good philosophical form to invoke a form of mutual agreement on what is being talked about, even if that which is being talked about can't be explained by those making the agreement. For the sake of clarity and as an illustration of the depth and intractability of the problem of consciousness, the authors describe two paradigms to defining consciousness.

The first paradigm, essentially the one described by Ned Block, draws a distinction between two kinds of consciousness : phenomenal (P-consciousness) and access (A-consciousness). It defines phenomenal consciousness as that referring to 'raw feels' or qualia, sensations, emotions , feelings etc, such as the 'redness' of an apple or the 'pain' of childbirth. The phenomenally conscious aspect of a state is 'what it is like to be' in that state.
A-consciousness is the phenomenon whereby information in our minds is processed and accessed for reasoning, control of behavior,report and inference Thus, we are A-conscious when we introspect, we are A-conscious when we recall something from memory,we are A-conscious when we deceive. Moreover, both P-consciousness and A-consciousness in this light, are not all-or- nothing phenomena - they are more like a spectrum.

The second paradigm describes consciousness as an all or nothing phenomenon, in that it assigns the term 'consciousness' to an abstraction that necessarily arises when a simulating entity attempts to involve itself in the simulation, because the abstraction that arises has the properties of consciousness. For instance, consider the following questions:

- Has this sentence been printed in black ink?
- Is this the second question being asked?
- Are the questions framed as bullet points?
- Are you reading this?
- Are you bored?

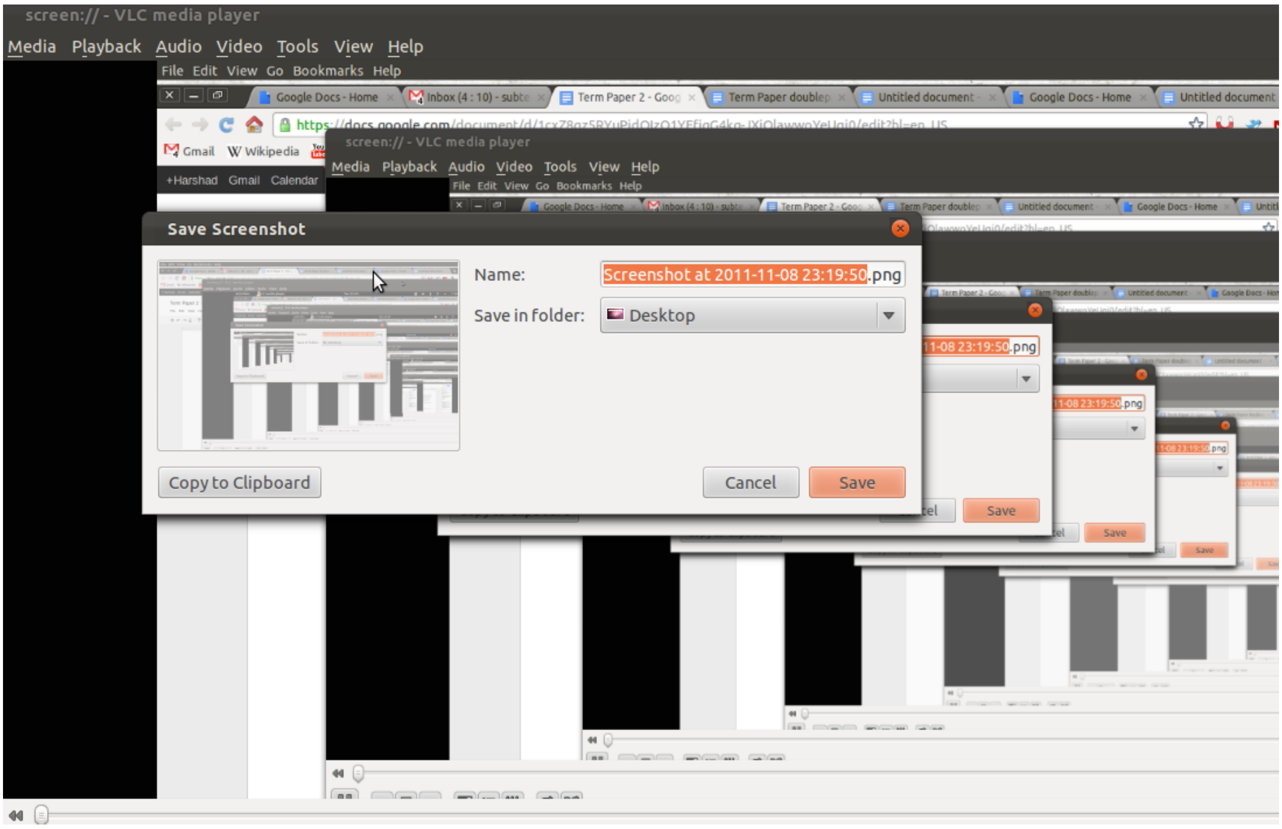
Humans have the ability to receive, process and operate upon information gathered from the environment. They can operate at either the perceptual or conceptual level, i.e. as a set of if-then-else statements or through hypothesis formation. Without going into the details, those that operate on a conceptual level would benefit because they would have a hypothesis on how the world operates, generated through observation, which they would be able to apply to novel situations, since they can simulate the dynamics of the system, and thus be better suited to survive in a dynamic environment.

The essential characteristic of a simulation is that it is less complex than the entity that it is simulating. For instance, a simulation of a deer being chased by a lion would not model the shifting shadows on the ground as the lion chased the deer. As would be the case with a simulating entity, as has been described above, *simulating only its environment*. There's a problem, however, when it tries to simulate it's interaction with the environment, because then it, the simulator, also has to become a part of the simulation, making the simulation almost always necessarily more complex than the simulator.

Returning to the 5 questions posed. The first two are essentially if-then-else questions, that do not necessitate novel hypothesis formation. The third one requires a hypothesis on the questions asked. The forth and fifth however, require, to greater degrees, hypotheses on *both* the questions asked and the individual answering the questions, something that is, as described in the preceding paragraph, not fully simulable, not solvable, not computable.

To prevent the generation of the loops, it becomes necessary to abstract the human, within the simulating system, as a black box. The properties of this black box are functions (more precisely quines) on the meta-representation within the simulating system. So the black box that is 'you' is spontaneously generated when the operation requiring the 'you' is asked. Not otherwise. This 'you' or 'I' is necessarily subjective, and inexplicable, and on probing, always leads to generation of deeper 'you's.

As a crude example, consider the image below, of VLC player displaying the results of screen capture on the screen, causing the generation of an recursive loop. Note dual loops, i.e. in the screenshot and in the screenshot of the screenshot, which shows how, even with a system as simple as that one, the *strange loops* can arise.



Name: Screenshot at 2011-11-08 23:19:50.png

Save in folder: Desktop

The second paradigm holds that **the phenomenon of consciousness arises out of this abstraction and is therefore a delusion,** which necessarily arises if an entity capable of making hypotheses attempts to make hypotheses on its interaction with the environment.

You are conscious only when an operation is requested on the information in your brain that required the consideration of the relation of that information, and the information processing, to the environment. In other words, **you are conscious only when you ask yourself, implicitly or otherwise, whether you are conscious.**

Therefore, the essential necessary and sufficient conditions under which consciousness *can* arise are:

- The ability to form hypotheses of the world, and associate the hypotheses
- Language, the medium for the formation of hypotheses, and of asking those questions that would indicate consciousness.
- Society/Culture, those external media through which symbol grounding, i.e. acquisition of language, occurs, and those that ask the questions that lead to consciousness.

Presence of these 3 conditions, will guarantee that the existence of consciousness, and the existence of consciousness will guarantee the other three. Further, it can be shown, although not done here because it would be involved and largely tangential to the discussion, that the ability to formulate hypotheses along with either of the other three will guarantee the existence of the other two. Therefore, the authors postulate **an equivalence relation between language, consciousness and culture**.

Obviously, this position has its critics. Some argue that this position mistakes, consciousness with 'meta-consciousness', or consciousness about consciousness. These opponents draw a distinction between the concepts of self-awareness (which is essentially an aspect of A-consciousness in the first paradigm) and the concept of consciousness.

Emotions, Qualia and Subjectivity

This section will attempt to demonstrate how the phenomenon described in the first paradigm of consciousness can arise from the second paradigm, and how the delusion of the first paradigm can be constructed form the ideas presented in the second. At this stage, the authors believe it instructive to define language as any system for abstraction and association of abstractions, with ‘spoken’ language being a transduction of the internal ‘processing’ language.

Humans operating in society and the physical world will generate great entangled webs of hypotheses and abstractions. Need for dynamic response to the environment will necessitate shortcuts or kludges in the hypothesis formation, generation and enaction protocols. Further, some protocols will be reinforced by societal and environmental conditioning, and by the specific neuro-chemical-anatomy of the brain of the human.

These biases will inherently lead to differences in the processing of information and in the circuits that release the neurotransmitters associated with emotions. This will lead to differences in the ‘self’ that will be generated when the self-generating programs are called, i.e. subjectivity, and to the generation of subjective qualia.

Experiment: Chatbots

In the pre-term paper, one of us (Deo), had proposed trying to figure out which sentence constructs necessitate consciousness. What he realized while working on the paper was that it was not sentences, but responses to the sentences that showed whether the entity was conscious. To test this hypothesis, we turned to chatbots.

Chatbots are programmed so as to give statistically likely responses to the statements made by the human, not to form hypotheses about what the human is talking about, and respond accordingly, i.e. that have no meta-representation of information, no tacit knowledge. They are programmed to respond purely syntactically, and thus serve as good subjects to test the hypothesis-formation hypothesis.

The reason chatbot’s can seem to fool humans, as was articulated by Kurzweil in *The Age of Intelligent Machines*, is that human beings, out of desire for conversation, limit the breadth and scope of the conversation, reducing the need for hypothesis formation (since novel situations are not being presented).

So we turned chatbots on themselves, i.e. fed the output of one of the Loebner prize winning chatbot in one window as the input to that in another window, and the results were...to put mildly, hilarious. The conversation degenerated within 10-15 exchanges, into meaningles banter, the most famous example perhaps being cleverbot talking to itself. Without hypothesis formation function and a function that used pre-formed hypothesis and meta-representation to generate responses based on the 'I', such concepts in conversation, such as 'I' or 'me' lost semblance of meaning. And this, once again, demonstrates the foresight of Alan Turing.

Thought Experiment: Mirror-Test with Machines

The mirror test was developed by Gordon Gallup Jr. in 1970, based in part, on observations made by Charles Darwin.
In the mirror test, we expose the subject to a mirror for some prescribed amount of time. If the subject behaves/reacts in such a way which suggests that it can recognise its own image in the mirror and not confuse it with another entity, then it is said to have passed the test. Usually, the a stimuli like a spot is placed on the part of the body of the subject which it can only see through the mirror to ellicit reactions which help in judging whether it has passed the test or not.

The Mirror-Test is often used to detect self-awareness amongst animals. However, the results are arguable, as there might be many instances where an animal may fail the test inspite of being self-aware. This might happen due to biological reasons like visual impairments, or certain behavioral reasons.For instance, in many species, it is seen as a signal of threat to look at another in the eye - this might hinder activities that lead to self recognition in an image.

One might ask: Can we extend the mirror test to machines as a test of a machine's degree of consciousness (or meta-consciousness) similar to how Turing Test is used to test a machine's ability to exhibit intelligence? The question helps us probe the nature of consciousness, and whether it can be seen as a purely mechanistic phenomenon.

The mirror test does demonstrate the capacity or ability of the subject to model the system efficiently. The subject is supposed to have a notion of 'self' as separate from the surroundings, and exhibit actions which suggest the existence of such a notion. But it still evades the so called ‘hard problem of consciousness’, or whether the machine or the subject possesses Phenomenal consciousness.

On Dreams

Since we are investigating the different aspects of consciousness, we cannot possibly overlook the state of dreams. The experimental observations include dream-like states need not only refer to internal representations and simulations that occur during sleep. We include mental states similar in nature, often experienced as part of the hypnopompic and the hypnogogic states.

Experiment: Dream Diary

A diary where dream experiences, as well as the experiences of the Hypnopompic and Hypnogogic state, are noted down.The subjects include a variety of people over a period of two months, some of whom were woken up (abruptly or gently) from their blissful sleep , for the purpose of the experiment.

Observations :

1. There is little or no control over what one dreams about. If one intends to dream on a particular subject, the likeliness of having that dream decreases drastically.
2. There is a state where one becomes aware of being in a dream. But that state is rare and difficult to sustain, and it is virtually impossible to interfere with the dream or control it.
3. In some instances, people seem to experience/see stuff in dreams that they havent experienced/seen in real life. Examples : a stanger, an unfamiliar song
4. Sometimes theres a slow transition from the conscious to unconscious or sleep state, where normal or real thoughts which one has control over slowly turn into absurd uncontrollable thoughts characteristic of dreams towards the state of sleep.
5. External inputs while we are sleeping tends to be transformed into various forms in our dreams. So, an alarm ringing in reality might be seen as a doorbell ringing in your dream, a pinprick on the body might be experienced as a spear piercing your body, or if someone

throws water on the subject , he might experience it as a tsunami in his dream.

An Explanation

We will now present an explanation for the above dream observations, partly through the first paradigm on consciousness.

Let us first look at a thought experiment proposed by Ned Block, which will prove to be helpful in understanding our arguments. (Note that Ned Block in his paper, though noting that dreams are a form of consciousness, doesn't provide any explanation for their occurrence or any description - “On my account, dreams are a form of consciousness,...though they are of less intensity than full blown waking alertness”. He takes no position on dreams).

A blindsighted person is ‘blind’ to stimuli in certain regions called blind field. If a stimulus is flashed in the blind field, the person fails to see it, or reports not having seen anything. This usually happens due to damage to some primary area of the visual cortex. But the interesting thing is that he might be able to ‘guess’ correctly some of the feature of the stimuli from a set of alternatives. For example, he might be able to guess verticality (from horizontal line/vertical line), or he might be able to guess colour of the object.So, if presented with an apple in the blind field, a blind sighted person would see nothing, but he might be able to guess from a set of alternatives (like apple, orange etc) that what he saw was an apple.

Ned Block imagines a ‘super-blindsighter’ who is able to prompt himself to guess at his will, instead of choosing an answer from a small set of alternatives.So the person is able to guess that he saw an apple, without actually having seen it, or lacking P-consciousness information of it. As Ned Block says, “Visual information from his blind field simply pops into his thoughts in the way that solutions to problems we've been worrying about pop into our thoughts, or in the way some people just know the time or which way is North without having any perceptual experience of it.” Theoretically, such a person contrasts what is it like to ‘see’ an apple (P-Consciousness) with having visual information about such an apple without having actually seen it. (Lacking P-Conscious experience about it).

When we are asleep, most of our P-Conscious (and sensory parts) are dormant or inactive.

Thus we receive very little or no Qualia inputs about the world around us. It is our contention that dreams arise when some of the A-conscious parts are partially active, but have no or little(partial) P-conscious experiences to operate on. The parts which organize or process information in our brain , the language areas etc , then operate on memory, or past phenomenal experiences and Qualia inputs and sensations. This, almost in all cases, results in a lack of synchronisation and incoherence arises. This is what we perceive as a dream. Now, if we look back at the though experiment, we might reach an understanding of our dream experiences. What we experience when dreaming, is not very different from what the super-blindsighter experiences when he visualises an apple or any such stimuli. We are not actually seeing or perceiving things in our dream. This is why people often speculate that dreams lack colour, or sound. When a subject perceives person in a dream, he lacks any sort of P-conscious experience of the person, yet he has some information which leads him to the conclusion that it was infact, that person. In this light, when we recall dreams, it is similar to the blindsighter recalling why he gave a particular guess or answer.

We can now explain most of our observations. When we become aware of a dream, we risk activating all our P-conscious parts and A-conscious parts, and thus losing the state of the dream, where only some of A-connscious parts are active. Waking someone up (by sprinkling water or giving a jolt), results in the sudden activation of the P-conscious parts.

The slow transition from wakefulness to sleep, where a dreamlike state arises is essentially the incoherence arising when the P-conscious parts become slowly inactive.
The transformation of external stimuli (like an alarmbell ringing) can be seen as an input which has been disrupted or not processed properly due to partial activation of P-conscious parts. The mismatch arising results in the inputs being irregularly processed by our partially active A-conscious parts which lead to the experiences in the dream.
Finally, in the state of dreamless sleep, we are neither P-conscious or A-conscious.

This incoherence, or mismatch - the paradox which arises out of partial A-conscious without Qualia inputs essentially leads to the absurd experiences characteristic of dreams. Ofcourse, similar mismatch between different forms of consciousness might also be present due to other reasons - like psychological disorders, or intake of drugs or other such substances.This explains why people involved with hallucinogenics often report having experienced dream-like states. Some forms of psychological disorders might also lead to such states, but these can deviate from dream-like experiences depending on the particular disorder.

Criticisms/Limitations

Probing the nature of consciousness through dreams (like we did), can be criticised, and we wish to make the limitations of our experiment clear:
1. The observations, since they are made in a semi-conscious state, can be erroneous.
2. The nature of the experiment necessitates that the observations are made in retrospect, instead of actively while experiencing the dreams.
3. We are relying on the *memory of dreams,* which is itself very fragile and thus very unreliable.

Concluding remarks to the paper:

The problem of consciousness has befuddled scientists and philosophers for millenia, and as such a single paper is insufficient to probe into even the most perfunctory elements of consciousness. So what we've tried to do here is through presenting differences in the very basic definitions of consciousness, illustrate the depth and gravitas of the problem being considered. We expect that by the end of this reading, the reader would have some idea as to where he places himself on the Penrose Continuum, and if not, atleast he would be asking himself that question.

Also, this paper, and the accompanying presentation, was an experiment in how term papers should be written and presentations be given, as a departure from the necessarily dry, boring and information-heavy paradigms currently operational.

References

As stated above, the authors, given the scope of the subject they chose to tackle, tried to avoid using direct references from specific texts, which is why there are no numbered citations. Instead they wrote the paper based on their respective interpretations of the topic, built post-reading numerous texts, some of them incompletely, on the topic being discussed.

1. Godel, Escher, Bach: An Eternal Golden Braid, Douglas Hofstadter
2. Consciousness: A Very Brief Introduction, Susan Blackmore
3. Consciousness Explained; Daniel Dennett
4. Animal Consciousness, Daniel Dennett
5. http://kogler.wordpress.com/2008/06/27/toddlers-and-the-mirror-test/
6. http://www.pnas.org/content/98/23/12874.full
7. http://en.wikipedia.org/wiki/Theory_of_cognitive_development
8. http://en.wikipedia.org/wiki/Cognition
9. http://www.conscioustthoughts.net/definition.php
10. http://www.integrativepsychiatry.net/neurotransmitter.html
11. http://www.ulm.edu/~palmer/ConsciousnessandtheSymbolicUniverse.htm
12. http://en.wikipedia.org/wiki/Self-awareness
13. http://en.wikipedia.org/wiki/Metacognition
14. On a confusion about a function of consciousness, Ned Block
15. The age of intelligent machines, Ray Kurzweil