

My Theory of Consciousness: A Summary

Harry H. Porter III

June 10, 1996

Three Definitions of Consciousness

Consciousness has roughly three meanings in our language and there is always confusion associated with the term “consciousness.” Therefore, I begin by giving several definitions.

First, conscious means awake. A person who is asleep, in a coma, or anesthetized is said to be unconscious, while anyone who can speak or respond to his environment is said to be conscious. In this sense of the word, any dog or mouse experiences consciousness when it is not asleep.

Second, the word conscious is often used to mean thinking the way a human thinks. This sort of consciousness involves a rich variety of mental activities such as talking and communicating with language, listening to music, experiencing anything sensuous, feeling emotions, doing mathematics, solving problems, etc. By this definition of consciousness, only humans can experience full consciousness, although chimpanzees or dolphins certainly experience something similar. One key aspect of this form of consciousness is the ability to think symbolically. It seems this ability to manipulate symbols and abstractions has allowed humans to evolve a very different lifestyle from other animals. Clearly, there is something unusual and interesting about thinking in the way that humans think so naturally.

Finally, there is a third subtle form of consciousness, which interests me.

This third form of consciousness involves being aware of your self and being aware of your own thoughts. This form of consciousness is not experienced during all waking moments, but is experienced only rarely. To distinguish it from the lower levels of consciousness, I will call this the “third level of consciousness.” (Being awake I call the “first level of consciousness” and thinking more-or-less like a human is called the “second level of consciousness.”)

This heightened third level of consciousness occurs when you close your eyes in a quiet room and stop thinking about anything specific. Close your eyes and look inward. What do you see? Perhaps you are distracted by some pressing thought (such as hunger, or the big interview tomorrow, or some other human obsession), but perhaps you begin to experience the sensation of thinking about thinking. Perhaps you become aware of yourself, as separate from your job, your identity, your roles, your past, even your body. This is a taste of what I mean by the third level of consciousness.

Of course, before long your eyes are back open and the pressures and concerns of the real world once again occupy your thoughts. When was I supposed to meet Joe for coffee? Will my application be accepted? Should I see a dentist about this pain in my tooth? The third level of consciousness is a special state and we cannot spend much time in it without risking a collapse in our corporal affairs or a trip to the local mental ward.

Chimpanzees can be coerced into performing crude symbolic manipulations. Their thought patterns probably achieve the second level of consciousness routinely, at least in many respects. Experiments have clearly demonstrated that humans manipulate symbolic information much better than chimpanzees. I believe that only humans are capable of experiencing the third level of consciousness easily and naturally. Perhaps a chimp occasionally closes his eyes and briefly enters the third level of consciousness, crudely glimpsing his own thoughts and inner world in a way approaching what the human brain does so easily and often.

I believe that the third level of consciousness will be accessible to a few appropriately designed computers in the near future. I believe there is no theoretical obstacle preventing computers from becoming conscious. There is no mysterious essence, no vital force to prevent machines from achieving self-awareness and full-blown consciousness. The only two limiting factors are computer processing speed and our ability to appropriately program the computers. Both these barriers will eventually be overcome.

It is speculation to predict when the first computer might begin to achieve the third level of consciousness, but I venture to guess that by 2050, a few computers will begin to experience full consciousness. Today the thought patterns of computers are very different from humans. It seems reasonable to guess that the first fully conscious computers will experience consciousness in a very different form than humans experience it. Since computers process symbolic information in very different ways from humans, their inner experiences will be correspondingly unimaginable to human brains.

Let's next examine various aspects of human thought in an attempt to see what is happening when a human experiences the third level of consciousness.

Mental Models

We form images in our minds of the things we are capable of thinking about and these images are called “mental models.” We form images in our minds of places we've been or things we've seen and these images are useful in thinking about real objects and things. In the acoustic domain, we can create sounds, songs, and sentences in our head. We can play with and manipulate these sounds silently in our minds, almost hearing them as if they were real. We also maintain a complex somatic-sensory model of our body position and its state. We can imagine various physical movements and assume different body positions using only this model without actually moving. In short, for anything we can think about in depth, we build and manipulate mental models—internal data structures, if you will—in our mind's eye.

How are these models stored in the human memory? Certainly they are stored very differently from the way data structures are stored in a computer's memory. We know how a computer can use bits, numbers, and pointers to represent quite complex pieces of information. The brain is very different, but it is still capable of holding internal representations that stand symbolically for physical objects and abstract concepts.

Some of the key differences between computer memory and human memory are as follows.

- (1) All computer memory is homogenous; any item of data can be stored anywhere. In the brain, certain types of data are only stored in certain places. Visual images are kept in one place, while aural images are kept elsewhere. As a result we can mentally hum “Happy Birthday” while visualizing an elephant, but it is virtually impossible to mentally hum “Happy Birthday” and Beethoven's Ninth Symphony simultaneously. Try it!
- (2) In the computer, the memory is completely passive and there is a sharp separation between the thread of control and data structures. In the brain, the processing power is distributed and mixed in with the memory. Thus, the stored data structures are active and tend to change and mutate “on their own.”

- (3) The computer memory has virtually no distinction between long- and short-term memory. Sure, data stored on a disk is distinguished from data stored in RAM, but both are pretty much the same. The main differences between disk and RAM (fast, random access times versus slow, non-random seek times) does not appear in the brain. In the brain, long-term memory is distinguished by being constantly available, watching with massive parallelism for patterns to trigger it, while short-term is riddled and permeated by active mental processing.

Logical Thought

There are many different kinds of processing happening in a human mind, many occurring simultaneously. One mode deserving our attention can be called “logical thought” or “mental reasoning.” It consists of a deductive process operating on internal representations, which produces changes to the internal mental state.

For example, a person may sit quietly, reflecting on words that were spoken earlier. As he thinks, various sentences are silently vocalized in his interior monologue. Hearing his own thoughts triggers inferences, deductions, conclusions, and various responses. Some of these responses are selected for further attention and are then “voiced” in his interior monologue. As the person slowly turns his thoughts over in his mind, one thought leads to the next, and his train of thought meanders this way and that. Perhaps he will come to some final conclusion or perhaps the chain of thoughts will simply wander away to some unrelated topic, via a series of small, related steps. For clarity, I have supposed that this train of thoughts is expressed in English sentences, but in actuality, thoughts are expressed in “mentalese,” the representation used by the mind, which is capable of expressing any thought that a human mind is capable of thinking.

By calling such thought logical, I intend to draw attention to the fact that this aspect of thought can be viewed as a formal system operating on existing statements to produce deductions via rules of inference. A first-order predicate logic theorem proving system is another example of such a formal system, although a theorem prover is quite different in the details. It is important to note that the term “logical” has a precise technical meaning (relating to truth vis-à-vis model-theoretic semantics) that does not apply to what the brain does. As we all know, human reasoning has its own logic. The deductions a human draws can certainly be said to have

some form of truth or logic, although in the final analysis, we humans are illogical, irrational thinkers who make mistakes of reasoning all the time, at least when compared to the standards of truth discussed by mathematical logicians.

To elaborate, humans are capable of reasoning about a rich and complex world, doing it semi-reliably, doing it with only partial and fuzzy data, and doing it within fairly strict time constraints. Such a feat of reasoning is far too large for any logically complete, mechanical procedure. It is not simply that today's computers are too slow; the task of complete and rigorously logical reasoning in such a rich and complex domain as the real world, is beyond the means of any reasoning system, human or electronic. The search spaces are hyper-astronomical.

To achieve the level of competence it does, the human approach to logic sacrifices any pretense of rigorous, formal correctness or completeness. As an example, you might ponder a friend's recent actions and conclude that he has a deep insecurity about love. While such a conclusion might be a "logical deduction," only a pedant or fool would remark that such a statement must be either completely true or unquestionably false and that the statement's veracity could be verified mechanically and algorithmically against a set of formal definitions, axioms, and rules of inference. Human reasoning is a different beast from formal systems of mathematical logic.

Automatic Thought and Optimized Reasoning

Many actions we perform are executed with a reflex-like inevitability and thoughtlessness. We perceive a stimulus, such as recognizing a friend's face, and we automatically take an action, such as asking "How's it going?" without considering in any great detail any other possible actions. As our brains mature, our neural circuits are made constantly more efficient (through learning) allowing us to take appropriate actions automatically and quickly, without spending extra time pondering alternative courses of behavior. Of course, there are micro-decisions being made constantly (When you spot your friend's face, think of all the possible behaviors that were quickly eliminated!), but these decisions are constantly being moved closer to the lower level of neural firing decisions. They become more automatic, unconscious, unconsidered, and reflex-like. I call this process "optimization."

Likewise, certain internal patterns of thought are optimized. For a small child to reach out and grasp an object, careful attention and thorough concentration are required. In the mind of an

adult eating dinner, he may only notice that the food needs more salt, and without even slowing his involvement in an ongoing conversation, his hand will reach out for the salt shaker and sprinkle some on his food. At the conscious level, his mind doesn't need to attend to the action in the slightest. The connection between the desire for salt and the appropriate behavior has been optimized. The adult can always consciously interrupt the optimized behavior and he can introspect or control his hand with close attention if the need should arise, but the optimized behavior is like a pre-programmed subroutine: ready to use. Of course, such optimizations form slowly and gradually over a lifetime.

Many of our abstract thoughts are also being optimized. The optimization process happens automatically and we really cannot stop or control it. As a result of this automaticity, a person may constantly fall into the same mental trap. For example, when a person hears someone ask him "Why did you do it that way?" he may automatically fall into thinking that his personality is being questioned, and that he is being criticized, and concluding that he is disliked. This mental chain may be right or wrong, helpful or hurtful, but that is not important here. The point is that during his childhood, this mental chain was not locked in place, but over time, as a result of many repeated activations of the initiating thought, with many repetitions of the same reasoning patterns, and with many repetitions of the same conclusion, the chain of thought has become automatic. After the adult brain has made the optimization, it might take years of therapy with skilled psychologists to break such fixed mental chains involving abstract thought patterns.

Generally, optimizations are useful and good. They speed our ability to think clearly. For example, when we see a dangerous person approaching, we react quickly to avoid him without having to "learn our lesson" first. When we see an arithmetic problem, we automatically know whether to add, subtract, or multiply to get the correct result. These optimizations allow us to think in larger chunks. The granularity of our thinking becomes bigger. We can reason through much more complex problems because we don't need to spend hours on each little step, the way a child must. When a child learns a new pattern of thinking, we traditionally call it a skill. All that has happened is that the child has spent time focusing on some mental chain and has managed to commit it to memory, optimizing the thought pattern into a fixed, automatic chain of reasoning.

The Role of Consciousness

To recap so far, we have seen that the mind manipulates mental representations using a process of deduction or reasoning. In the mind, one thought or representation triggers another and this is automatic at the level of the neurons. Mental deduction occurs as the product of many individual acts of pattern recognition, followed by triggering of further mental representations. As a human ages and learns, the associations his brain makes are refined and optimized. Useless, incorrect, and extraneous associations become inactive and useful associations become strengthened.

There is a problem with the brain's system of deduction. It is neither complete nor correct, but simply a system of associations. As a result, cycles of feedback can occur. X triggers Y, Y triggers Z, and Z triggers X again, starting the cycle over. These cycles can be small, where X, Y, and Z are single neurons, but nature seems to have taken care to prevent this. The brain seems to be designed and wired in such a way that neuron-scale cycles do not incapacitate it. In some cases, X, Y, and Z are large clusters of neurons and the resulting cyclic feedback is called a seizure. Feedback loops at this scale also occur rarely and can be interrupted by either drugs that dampen the cyclic behavior or surgical removal of some of the connections in the cycle. But there is another type of feedback problem that can arise. Here, X, Y, and Z are individual thoughts or mental representations. Some thought (X) triggers another thought (Y), which in turn triggers a third thought (Z). The last thought then triggers the first thought (X) causing a feedback loop. Obviously, the chain might be much longer and may have variations and side-tracks, but the idea is that there is a potential for feedback loops at the thought level.

What would such a feedback loop feel like? Well, to a person who only experiences the second level of consciousness, the fact that his thoughts are repeating cyclically will be invisible. Remember that he is just thinking, not thinking about his own thinking.

Fortunately, humans have mechanisms to prevent some of the simpler feedback loops. First, there is a lot of chaos and noise (especially in a young, unoptimized brain) and feedback looping simply becomes a statistical improbability. The second time around the loop, the flow of activation is likely to veer off down a different path. Second, neurons get fatigued and desensitized. After participating in any tight loop, you will simply “get tired of thinking about X.” We've all gotten trapped in mental loops and kept thinking obsessively about some topic until we have mentally exhausted ourselves on the subject. I think a common type of this looping involves a melody stuck in your mind. Each portion of the melody triggers the next portion, so the melody plays itself out. But at the end, the whole melody has sufficiently triggered the first few bars to start it over again. As another example, we have all experienced

times when we went over and over some thought sequence. Each time we feel we get nowhere, yet we begin going over the scenario one more time.

These examples point out yet another mechanism we have for detecting loops: we notice them. This is important. To notice that your thoughts are in a cyclic loop requires a kind of meta-level reasoning ability. Once we notice that our thoughts are repeating, we can take action (like going for a coffee break, for example) to break the cycle and get some fresh air into our minds.

Control of cyclic feedback looping is only one example of something more general. To be efficient and skilled at reasoning, we need a mechanism that can monitor our mental activity and make changes when necessary.

Another important control task is determining how much time and energy to invest in thinking about some particular problem. This control mechanism is very important. If a hungry lion is approaching, we need to make a very quick decision about which tree to climb and then commit ourselves to executing that action quickly. If a hungry carpenter ant is approaching, we know we may take longer to decide on a plan of action.

The human system of deduction is set up so that it can yield some results quickly, but it will yield better results if it is allowed to think longer and more carefully. The real problem is knowing how much time to spend thinking about a given problem. Sometimes, it is better to make a quick decision, sometimes it is better to think longer about something before acting. This is a control problem and can only be achieved by meta-level reasoning.

This meta-level reasoning is what constitutes the third-level of consciousness. When we reason on the third level, we are looking at our own thoughts and asking about how they relate to our ultimate goals. Often at times of increased consciousness, questions about the meaning of life, mortality, and so on arise. This is natural. At the meta-level of controlling your thoughts, these are the some of the factors that must be considered. The fact that we can occasionally reflect on our selves and see our own thoughts for what they are is a byproduct of the more important function of higher-level consciousness. This role is to evaluate our progress toward our largest goals and to organize, channel, and direct our more mundane thoughts so as to achieve these goals.

Emotions

Emotions are set off automatically and are in many ways beyond our control. They are global parameters and affect the way the logical portions of our minds reason and think. (By parameters, I mean that each emotion is like a number that all parts of the brain can perceive simultaneously.) The most basic and fundamental emotions are fear, lust, pain, hunger, anger, the maternal instincts, and so on. These feelings are operationally important for any species more intelligent than an insect and are probably important for insects as well.

I imagine that insects may experience something that we might call fear, but when examined closely, the “emotion” of an insect is substantially different from a human emotion. For example, a fly might sense a shadow moving toward it and perceive it as a threat, triggering “fear.” The details about what aspects of the shadow will trigger the “fear” response are fixed and can be measured. The object must be at least so large, moving at least so fast, and at least so close. If the stimulus meets the criteria, fear always results. The resulting behavior is then triggered unconditionally and the behavior is always the same. A fly will always respond by taking to the air and following a random trajectory. Once a certain time has elapsed, the “fear” state will dissipate and dissolve, and the fly will then land. After the “emotion” has subsided, I suspect that nothing we could call a memory of the event would remain. Perhaps over time, an insect could become habituated to a particular fearful stimulus, but this would not be a memory, in the normal sense. With an insect's mind, a stimulus will invoke a mental state and this will lead to a pre-ordained response. The only interesting questions are what are the exact parameters of the stimulus and what happens when the insect is exposed to conditions that trigger different responses in different degrees and the insect must make a “decision” with regard to a novel situation.

In humans, the processing of emotions is more complex. Like an insect, a human emotion will be triggered automatically as a result of pattern recognition circuitry. A child who finds himself alone and in the dark will invariably become afraid. What makes it interesting is that the human is capable of reasoning and semi-logical thinking; his response is rarely automatic. Instead, the emotion serves to channel his thought. It exerts a pressure on the pattern of his logic. It distorts and changes the chains of reasoning that would otherwise occur. I call this effect “emotional pressure on logical thought.” The emotion can be modeled as a dial being turned up, based on the perception of some combination of patterns being recognized. When a person becomes afraid, it causes his fear index to move up, for example, to 0.8 on a scale from 0.0 (no fear) to 1.0 (sheer terror).

Once a person becomes afraid to any degree, his subsequent actions, decisions, and perceptions will be influenced by the presence of a high level on the fear-o-meter. A noise that would previously be ignored might now be interpreted as a monster under the bed. Or perhaps a low level of fear might cause an adult to get up and turn on the porch light, just in case, whereas normally he wouldn't have wasted the energy. The response might range from breaking into a dead run toward the nearest tree to sending a memo in order to preempt a strategic attack in an office meeting. With emotions we talk about levels of response and appropriateness of behavior. When the degree of influence of emotion over logic is too low or too great, the person is sent into therapy to try to adjust behaviors.

The other basic emotions (such as lust, pain, hunger, anger, and the maternal instincts) can all be modeled similarly. I suggest some emotions can be closely modeled by a single number on a dial. For example, a person's degree of sexual appetite (lust) might be reducible to a single number. Granted, there may be a little more information associated with each emotion than this simple model suggests. With hunger, for example, a human may crave water or salt or ice cream. His behavior will be different in each case. The seasoned sybarite might say that lust, too, has too many variations to be modeled by a single number.

Physical pain seems to lie beyond this single parameter model of emotion since there is quite a bit of information accompanying each instance of pain. For example, a specific pain might consist of the information "my big toe hurts from a hot temperature." Human responses tend to be automatic, fast, inflexible, and highly dependent on the particular type of pain. Interestingly, the response to painful stimuli is processed symbolically and, as a person matures, various optimizations take over this symbolic processing. Perhaps you have experienced a minor pain, such as a hang-nail, in a disassociated third level of consciousness. You know your finger hurts but it doesn't interrupt or interfere with your thought patterns. Nevertheless, pain is like the other emotions since, in certain cases, it can influence your entire way of thinking. Pain can twist and distort the normally logical reasoning processes. Anyone who has become emotionally depressed after a long bout of low-level pain knows that pain functions as an emotion.

The Evolution of Self-Consciousness

The third level of consciousness involves our ability to see and think about our own thoughts. We observe ourselves think. We also alter the flow of our own thoughts. More precisely, one

section of the brain observes another section, makes a decision, and then exerts control to influence and alter the path of thinking that the second part of the brain then follows. We “change our minds.”

Why do we have the ability to observe and reason about our own thoughts? I think one piece of the answer is that the ability to control our thoughts is useful. We can see ourselves get stuck in a mental rut and we can force ourselves into a new way of thinking. We can make decisions about goals and objectives and then shape our thoughts in ways that help achieve those goals. For example, you may decide that there is nothing you can do about some problem in your life; your mind naturally comes back to work on that problem over and over, but finally you decide to stop wasting your energy thinking about it. When, once again, your mind falls back into thinking about the problem, you (actually, another part of your brain) forces your thoughts to go in a different direction.

This ability to observe, reason about, and control our own thoughts imparts survival value and it is natural to suppose that it was selected for. Humans are thinkers and we achieve our survival by thinking about things. An ability to use your mental resources more efficiently would clearly be of great survival value.

But in order to reason about your own thoughts, a certain amount of underlying reasoning ability has to be present in the brain. Humans are good at thinking about certain things and bad at thinking about other things. For example, humans are good at finding their ways through forests and bad at finding their ways through mathematical equations.

In proto-humans, the question is why did the ability to do meta-reasoning arise? Why did humans evolve the ability to think about thought? It is very different than thinking about gathering berries and the other mundane tasks of survival faced by early humans. Many other mammals get on well enough, without any apparent ability to reason about reasoning. Why did it evolve in humans?

The answer lies in the fact that humans are a social species. Each individual is different, and as humans evolved, each individual began to have complex inner mental states. An ability to understand what the other guy is thinking about would have been a great advantage to an early human. For example, being able to see a coconut fall on someone else's head and understand that such an event would hurt and to then predict the ensuing behavior of the wounded individual would be beneficial. Perhaps proto-humans had a tendency to go into a rage of violence after

getting bonked on the head with a coconut—any other human who could see that coming would be at an advantage. Obviously, there are many more subtle and realistic examples of how the ability to understand and reason about the mental state of friends and relatives would have been beneficial to an individual.

I suggest the ability to reason about other humans' mental states helped to drive the pressure of natural selection to evolve an ability to think about thinking. The ability to think about thinking evolved to allow humans to think about other individuals' thoughts. Only after that ability began to arise, did it begin to be applied to the self. Social pressures led to the ability to reason about reasoning; later it allowed an individual to think about his own thoughts. I suggest our ability to think about our thoughts is merely a side-effect, an unintended consequence of our ability to forecast the actions of other human brains.

We probably reason about our own thoughts so much better than we reason about other people's thoughts since we have a lot more raw input information about our own thoughts. We don't really know what goes on inside other people's minds, but knowing just a little can make a big difference in survival with a little careful consideration. On the other hand, that same machinery can easily be applied to our own thoughts. When we think about our own thoughts, our brains can take off. Once we evolved this ability to think about our own thoughts, there may have then arisen a survival value to the ability to exert some control over how we used our own mental resources. The human who could force himself to stop worrying about the inevitable and unchangeable realities of life, for example, could get on with thinking about something directly related to his survival. Such an individual would have been at an advantage in the contest for survival.

We evolved the ability to do meta-reasoning in order to excel in our role in a social society; this ability then allowed the human brain to exert some control over itself. This opened up a new path for the pressure of natural selection to follow and the action of evolution improved the ability of humans to think about and control their own mental efforts. The key to the survival of the human species has been its ability to perform symbolic reasoning. Such reasoning can easily get out of hand. As computer-based reasoning systems demonstrate so clearly, it is easy to waste huge computational resources on very small symbolic reasoning tasks. The ability to control and channel how mental resources are spent would be of immense survival value to evolving humans. This second phase in the evolution of consciousness has gotten us to where we are today, along with a greatly evolved ability to do symbolic reasoning in general and the

development of culture. This evolution of the mind has given us the ability to enter the third level of consciousness often and easily.

Where will this new level of consciousness take us? Of course it is impossible to predict, but I would guess that it will lead humans to develop computers that are capable of similar levels of consciousness. Ultimately, this may lead to a new species of evolving, conscious entities based not on the human genome, but constructed out of synthetic materials according to plans that can be adapted and evolved much more readily than DNA.

Perhaps our ability to enter sustained periods of third level consciousness will not be useful to humans from an evolutionary point of view. Perhaps our ability to do deep introspection, our desire to spend years in psychotherapy will not be helpful. Perhaps the Zen monk who spends his life with his eyes closed examining his own consciousness will not be raising successful offspring. On the other hand, perhaps the party-goer, who drinks to suppress his consciousness and ends up conceiving children in spite of semi-conscious mental edicts telling him not to have sex, will be the individual who propagates his genes into future generations. But I doubt it. Whatever evolutionary forces have led humans to develop full-blown consciousness probably still apply and will force the continued development of human consciousness. As far as machine consciousness goes, I believe the evolution of consciousness will soon take off and continue in exciting ways we cannot predict. There is no doubt that the Third Millennium will be far different from the last thousand years. Synthetic consciousness will play an important role in a future that may no longer belong to humans alone.