Joseph Clark

Dr. Dustin Hannum

1WRT 105E

12 December 2012

"On Holism and Reductionism as Applied to Human Consciousness"

As a 6-year-old on a hot summer's day, there really wasn't much for me to be doing. I had been playing in the sandbox earlier, but that had gotten boring rather quickly. The pool wasn't set up, none of my friends were around, and I was despairing of anything to do. Suddenly, a slight motion caught my eye. I looked down at the ground and saw – an ant. I kneeled down to watch its progress more closely, following it with my eyes as it toiled across my driveway carrying but a single crumb. Soon enough, I noticed another – and another. Much to my surprise, all were following the same path. Despite the fact that they were unable to see each other, each ant seemed to know where the ones before it had gone – and wished to follow the same path. Did they have some sense of desire, some sense of purpose, some sense of consciousness?

I am now far removed from my six-year-old self in many ways. I understand the use of pheromones to communicate, and I understand that an individual ant can safely be said to have little more consciousness than a pebble, if any. However, as I gain newfound knowledge, more and more questions arise. While an ant may not be conscious, the colony considered as a whole certainly appears to exhibit some sense of purpose, some sense of internal logical and goals (Hofstadter, "Gödel" 316). Without formal language, the colony somehow sustains itself, gathering food, reproducing, defending against attack, and, when necessary, moving to a new home (Griffin 59). Considered as a whole, the colony exhibits distinct signs of consciousness.

This consideration of ant colonies serves as a fair analogy for the human brain and neatly delineates the two main schools of thought on the origin of consciousness, reductionism and holism. *Consciousness* itself is rather tricky to define, but an entity possessing consciousness is generally held to exhibit certain characteristics including the concept of self, the ability to have knowledge of the past and future – and thereby the abilities to learn and to plan (Scilipoti). Considered individually, an individual neuron – or ant – cannot be said to have much if any consciousness. Therefore, the principle of *reductionism* holds that we can reduce all functions of the mind to this fundamental level and thereby understand and mechanize the processes of thought (Hofstadter, "Gödel" 312). However, considered as a whole, the mind – the colony – exhibits features that are not immediately evident as the natural sum of their constituent parts. Therefore, the principle of *holism* holds that consciousness somehow emerges from physical phenomena in a way that cannot be expressed in terms of the sum of actions at lower levels (Hofstadter, "Gödel" 312). Both holism and reductionism arise from rich historical backgrounds and offer forth an explanation of the origin of human consciousness. However, with the current state of research, neither appears to be in the ascendant. In fact, in order to gain a full understanding of the nature of human consciousness, the mind must be interpreted both holistically and reductionistically.

Reductionism may seem more immediately logical to the readers of today, situated as we are in a scientific era. Although it had rudimentary proponents in antiquity such as Democritus, Epicurus, and Lucretius, reductionism returned to vogue in the era of Enlightenment (Jaki 19-20). During this period, Newtonian physics sought to explain the world in terms of the random motion of particles, and primitive computing machines were theorized and constructed by the likes of Gottfried von Leibniz and Blaise Pascal (Jaki 22-23). It was only natural, then, to follow

Thomas Hobbes's line of thinking and theorize that the brain could be considered as an elegant, extremely complex machine. This line of thinking only gained in popularity and credibility with the more modern advent of electronic computing machines and their vigorous advancement by the likes of Alan Turing (Hofstadter "Gödel" 594-595). In today's computerized society, the notion of likening a brain to a clearly reducible computer is a natural analogy.

However, in the Age of Enlightenment – and even in the early part of the 20th century – not terribly much was known about brain structure. Even today, the outlook is not substantially better. A large amount is known about the coarse structure of the brain. The major functions of the cerebrum, cerebellum, and brain stem are taught in elementary school health classes. Various glands and lobes have been identified with specific functions. For instance, the pituitary gland "is known to be the organ which produces a large number of different hormones controlling a wide range of body activities, including growth and sexual development" (Rose 58-59). Then, too, a large amount is known about brain structure at the neural level. Neurons function as interconnected transmitters of electrical impulses, with "inputs" (dendrites) and "outputs" (axons) (Rose 65-66). Once the impulses received by the dendrites exceed a certain threshold, an electrical impulse of "fixed amplitude and shape" is sent down the axons, to be picked up by the dendrites of all the neurons it's connected to – a process that may be repeated up to about 1,000 times per second (Sommerhoff 137). It is worth remarking that, while a fair bit is known about the cellular basis of brain structure as well as large-scale features, not much if anything is known at the intermediary level regarding brain function (Hofstadter, "Gödel" 346). This strict dichotomy of knowledge between small-scale and large-scale structures in the brain mirrors the distinctions between reductionism and holism in relation to the nature of the human mind.[1]

The invariability of neural function does lend a strong argument to the reductionist point of view, particularly with the advent of electronic computers. If each neuron sends out a pre-determined output in response to inputs exceeding a certain pre-determined threshold, why could not such behavior be modeled in a computerized system? Admittedly, it might be complex – the number of neurons in the human cortex is on the order of 10^{10}, with around 10^{14} connections between them (Rose 66) – but why could such a computerized mind not be constructed, at least in theory?

It is precisely this question that fascinated Alan Turing, "one of the true pioneers in the field of computer science" (Hofstadter, "Gödel" 595). While Turing wished to consider the question of whether machines can think, he didn't believe that the question as stated is meaningful due to the wide variety of possible definitions for the words "machine" and "think" (Turing 4). Instead, he chose to re-formulate the question in the form of a game or contest of sorts – one that has become known today as the "Turing test" (Hofstadter, "Gödel" 595). The test is run as follows: An interrogator is alone in a room with a computer terminal. He or she is able to communicate via some sort of instant messaging service with two entities – one human typing on a computer terminal in an unseen location and one computer program. It is the interrogator's job to determine which entity is human and which is computer (Turing 5). Provided one accepts this formulation of the question "Can machines think?", the principle of reductionism of consciousness can now be transferred to the realm of computers and artificial intelligence.

---

[1] For the purposes of this paper, *brain* and *mind* shall continue to be held as distinct terms: *Brain* refers to the physical organ composed of neurons while *mind* refers to the state of thinking consciousness that somehow arises from our brains' structure.

Indeed, a large deal of work in the domain of artificial intelligence has – intentionally or not – gone into precisely this sort of research and development. One of the immediately evident differences between humans and computers – one held in stark relief by the Turing test – is the use of language. With the development of programming languages in the beginning of the second half of the twentieth centuries, one of the first issues to be tackled was that of parsing natural language (Minsky 20). It is one thing for a computer to be programmed with a formal set of symbols held in unchanging relationships. It is another thing entirely to be able to respond appropriately to a natural language – with all of its inherent inconsistencies and ambiguities that make it hard enough for a human to learn. These early programs took such names as "Semantic Informational Retriever", "STUDENT", and "ANALOGY" (Minsky vii, 1) and approached the problem in a variety of ways. These ways included creating a large database of structured information about various words in context, attempting able to solve high school algebra problems and attempting to complete visual analogies (Minsky 14, 33-35). Such efforts have culminated most recently in 2011 when a program known as "Cleverbot" got a sample of 1,334 votes from 100 respondents to rate its responses as 59.3% human-like. As a point of comparison, the respondents who were talking to an actual human only rated responses as 63.3% human-like. This is still a far cry from being a satisfactory resolution of the Turing test – ratings were based on a scale of 0 – 10 and were based on a 4-minute private conversation rather than a direct comparison between a human and computer ("Cleverbot"). Despite these considerations, Cleverbot's success provides a reasonable hope that computers may be able to one day parse natural language as easily as humans – a large argument in the favor of reductionism.

Even in light of this recent success, however, reductionism is far from a proven concept and has a number of vocal detractors. These supporters of holism come have a venerable heritage rooted in historical philosophy and religion. Plato and Aristotle both believed in “the irreducible features of human consciousness” based on their animistic convictions (Jaki 17). Religions in the Judeo-Christian tradition hold that God created humans and imbued each with an immortal soul that cannot possibly be explained mechanistically by the laws of physics. Even the inventors of early computing devices such as Blaise Pascal, Gottfried von Leibniz, Lord Kelvin, and Charles Babbage were quick to note that their devices merely calculated as they had been designed to and were not capable of independent thought (Jaki 22, 25-26, 42-43).

The son of a clergyman and an Anglican himself, one of the first people to respond to the recent rise in reductionist philosophy was the British philosopher J. R. Lucas (Lucas “J.F.”, Hofstadter, “Gödel” 388). His argument is based on formal logic – specifically, Gödel’s incompleteness theorem. While the brilliant particulars of its derivation are beyond the scope of this paper, it states that no formal mathematical system can be both consistent and complete: that is, no strictly rule-based method can produce every true statement. This occurs because once the system is sufficiently advanced to produce simple arithmetical truths, it has become powerful enough to create a statement that in effect means, “This statement cannot be proved within the formal system”. If the system produces this statement then it becomes false, so the system is inconsistent. If the system cannot produce this statement then it becomes true, so the system is incomplete (Lucas “Minds” 43-44). Lucas was well aware of Gödel’s theory and its implications. A computer is, in essence, nothing more than a rule-based formal system. By definition, it will act within the bounds of its programming. Barring a physical malfunction, the input “3 + 3 =” must yield an output of “6” each and every time – and so on and so forth. But then, all computers must be subject to Gödel’s incompleteness theorem. Thus, there is some true statement that each and every computer cannot produce as true – but that we, as humans standing outside the system, can easily perceive as true. Therefore, there must be a fundamental difference between minds and machines (Lucas "Minds" 44-47).

While inviting, Lucas's argument is in fact not quite as sound as it may appear at first blush. C. H. Whitely is quick to point out that – while Gödel's incompleteness theorem does limit the ability of any computer – that does not mean that minds are fundamentally different. Whitely, for instance, presents the statement, "This formula cannot be consistently asserted by Lucas" (Whitely 61). By an argument directly analogous to the one Lucas used above, Lucas cannot possibly be both complete and consistent. Whitely then proceeds to ask, "Does this show that our minds are unlike Lucas's, and in at least one respect superior, since there is at least one truth which we can grasp and state, and he cannot?" (Whitely 61). Clearly this is a ridiculous conclusion to draw, as it applies equally to every human and computer – each has the same "defect", if you will: the same Gödelian roadblock to being both complete and consistent.

Supposing, however, that holism does have some merit – as many do – what form would it take? Simply holding that the mind and consciousness are fundamentally inexplicable is hardly a satisfying solution, and so some positive arguments must be presented. Some, such as the mathematician-philosopher Michael Scriven, argue based upon pure semantics. He believes that, by definition, machines cannot be conscious. Therefore, he has "come to see that a reproduction of a man sufficiently exact to be conscious is too exact to be still a machine" (Scriven 36). This is hardly a satisfying explanation either, though – it simply tries to brush away the issue.

A very different attempt to provide a rational explanation of how holism could explain consciousness comes from the philosophers C.D. Broad and H.H. Price, of Cambridge and Oxford, respectively. Holism holds that mental function cannot possibly be explained in terms of neural activities in our ordinary physical space. Therefore, some other explanation must be provided. Both Broad and Price posit that a mental space exists (separate from ordinary physical space) in which mental events – that is, our perceptions, hypotheticals, daydreams, and such – take place. How these two spaces could relate is even more speculative. Price holds that there are temporal and causal relations between physical space and mental space, but no spatial relations. Broad, in contrast, holds that physical space and various mental spaces form are actually cross-sections of a larger *n*-dimensional space (Smythies 240). While an intriguing idea, particularly in relation to recent developments in theoretical physics that hint at large-dimensional spaces, there has as of yet been no empirical evidence to suggest that mental events take part in their own space only vaguely related to "real" physical space. While perhaps unlikely to our current mode of thinking, this is one of the only explanations as to how the philosophy of holism might be practically applied to explain consciousness.

It would be helpful to the holist cause if it could be shown that human consciousness were somehow unique and discontinuous from what had come before it. If this were the case, it could then be argued that consciousness is an emergent phenomenon as it occurs only in humans and not in other similar structures such as the brains of other animals. Unfortunately, this is not the case. Although it is difficult if not impossible to observe directly, animals have been shown to exhibit various signs of consciousness discussed above: to wit, the concept of self, the ability to have knowledge of the past and future, and the abilities to learn and to plan. Researchers have seen that certain animals (including great apes) are able to recognize that an image in a mirror as themselves (Griffin 30-31). Therefore, these apes must hold some concept of self in their minds. Animals as diverse as ants, octopuses, vultures, sea otters, and chimpanzees have been observed to clearly use tools – that is, to plan future actions (Griffin 109). Other experiments with colored objects show that even pigeons can learn to peck in in a specific color order (Griffin 92). Even language shows a somewhat continuous progression, from the symbolic dances of honeybees to mutable birdsong to the ability of great apes to communicate with sign language (Griffin 42, 154, 70).

It would then be logical to conclude that the possession of consciousness is not truly a black-and-white issue. Rather, it is an issue of degree. Just as it would be counterfactual to say that a chimpanzee possesses no consciousness, so would it be counterfactual to say that it has as much consciousness as an adult human. Then, there must be shades of grey involved – degrees of consciousness roughly progressing from unconscious beings like ferns or mosquitos to those octopuses and ants and sea otters and great apes that appear to have some degree of consciousness all the way to humans whom we assume to possess full consciousness. This makes the holist argument very difficult to maintain. If consciousness is some sort of emergent property unique to the human mind, then why is it found in other species and in varying degrees? Indeed, this spectrum of consciousness appears to closely tie consciousness with physical brain structure: a reductionist's dream.

However, we must not be too hasty in our judgment, for that is not to say that reductionism has the final say on the issue of consciousness. It too is not without flaws. Much of the discussion regarding reductionism, as we have seen, revolves around the question of artificial intelligence, which in turn revolves around the Turing test. How reliable is the Turing test, though, in establishing consciousness? At first glance, it may appear to be a valid substitute – but it really is rather arbitrary in the absence of an adequately reliable definition of consciousness, as the philosopher Keith Gunderson points out. Suppose we construct another imitation game as follows: An interrogator is alone in a room with a small hole in the wall. He or she is able to stick his or her foot in this hole and thereby gain information about two entities – one human who will step on his or toe and one box filled with rocks and attached to an electronic sensor. It is the interrogator's job to determine which entity is human and which is computerized rock-box (Gunderson 62-63). I find it highly unlikely that anyone would accept that the rock-box were conscious even if it were able to fool the interrogator a significant percentage of the time. But, without a rigorous definition of consciousness, why then should we accept the originally proposed Turing test? The ability to converse is certainly a more important aspect of consciousness than toe-stepping – and would provide evidence for the perception of self and sense of past and future – but it is certainly not the sole criterion. At this time, there is no set of criteria that would provide a rigorous definition of consciousness. Then, as affairs stand, it is not possible to conclusively demonstrate whether or not the reductionists have proven their case with a computer model of the brain – whether or not they have truly created a mind.

It has shown that both holism and reductionism have inherent flaws in their explanatory power. Neither philosophy appears to provide an infallible explanation of human consciousness. How, then, ought we to consider this thorny question? Ought we to simply dismiss both points of view outright? This is surely extreme, as both modes of thought offer useful insight. Reductionism directly relates to the physical structure of the brain, which we are continually researching and learning more about through neuroscience. Holism directly relates to our intuitive sense of the world and how the brain "should" work from a psychological perspective. Fortunately, we are not forced to choose between the two. Reductionism is a perfectly valid explanation of neural firing processes while holism is a perfectly valid explanation of large-scale structures in the brain. But, where then does consciousness lie?

A definitive answer will not appear for many years yet, if ever, but it seems important to give due consideration to both holism and reductionism in any such answer. Just as ants form teams and signals to accomplish specific tasks, so do neurons function in concert with other neurons (Hofstadter, 319). The ants' purpose cannot be understood by considering each ant individually – but nor can the achievement be understood without considering each individual ant. Likewise, neural function cannot be understood by considering each neuron individually – but nor can it be understood without doing so. If one such reductionistic collection of neurons could be understood on a holistic level as representing a concept of self – why, that is the very stuff of consciousness itself (Hofstadter, "I" 86)!

While holism and reductionism both offer valid and useful information regarding brain function on their own, neither provides a truly consistent and complete concept of consciousness. Rather, each expresses but one facet of a fundamentally complex and multi-faceted idea. In order to fully come to terms with the nature of human consciousness, we must embrace both holism and reductionism. Only by accepting multiple levels of meaningful information can we truly begin to wrap our head around how we truly wrap our head around things.

Works Cited

"Cleverbot - Turing Test at Techniche 2011." *Cleverbot - Turing Test at Techniche 2011*. Rollo Carpenter, n.d. Web. 02 Dec. 2012.

Griffin, Donald R. *The Question of Animal Awareness: Evolutionary Continuity of Mental Experience*. New York: The Rockefeller University Press, 1981. Print.

Gunderson, Keith. “The Imitation Game.” *Minds and Machines*. Ed. Alan Ross Anderson. Edgewood Cliffs, NJ: Prentice-Hall, Inc., 1964. 60-71. Print.

Hofstadter, Douglas R. *Gödel, Escher, Bach: An Eternal Golden Braid*. New York: Basic Books, Inc., 1999. Print.

- - -. *I Am A Strange Loop*. New York: Basic Books, Inc., 2007. Print.

Jaki, Stanley L. *Brain, Mind and Computers*. New York: Herder and Herder, 1969. Print.

Lucas, J. R. *J. R. Lucas Homepage*. N.p., 2002. Web. 06 Dec. 2012.

- - -. “Minds, Machines and Gödel.” *Minds and Machines*. Ed. Alan Ross Anderson. Edgewood Cliffs, NJ: Prentice-Hall, Inc., 1964. 43-59. Print.

Minsky, Marvin, ed. *Semantic Information Processing*. Cambridge, MA: The MIT Press, 1968. Print.

Rose, Steven. *The Conscious Brain*. New York: Paragon House, 1989. Print.

Scilipoti, Turi. Personal interview. 29 Sep. 2012.

Scriven, Michael. “The Mechanical Concept of Mind.” *Minds and Machines*. Ed. Alan Ross Anderson. Edgewood Cliffs, NJ: Prentice-Hall, Inc., 1964. 31-42. Print.

Smythies, J. R. “Aspects of Consciousness.” *Beyond Reductionism: New Perspectives in the Life Sciences*. Ed. Koestler, Arthur, and J. R. Smythies. New York: The Macmillan Company, 1969. 233-245. Print.

Sommerhoff, Gerd. *Logic of the Living Brain*. London: John Wiley & Sons Ltd., 1974. Print.

Turing, Alan M. “Computing Machinery and Intelligence.” *Minds and Machines*. Ed. Alan Ross Anderson. Edgewood Cliffs, NJ: Prentice-Hall, Inc., 1964. 4-30. Print.

Whitely, C. H. “Minds, Machines and Gödel: A Reply to Mr. Lucas.” *Philosophy* 37.139 (1962): 61-62. *JSTOR*. Web. 19 Nov. 2012.