Cognitive Science and Its Discontents

Cognitive science (CS), put simply, explores the mysteries of the mind pertaining to the way humans think. In other words, it looks to investigate the inner mechanisms and functions of the mind and brain to understand exactly *why* and *how* cognition works (Friedenberg & Silverman, 2006; Anderson, 2015; Gazzaniga, Ivry, & Mangun, 2014). The contributing scientists in this field use interdisciplinary methods and theories to explain the human brain and cognition. The relevant fields to the study include but are not limited to neuroscience, anthropology, evolution, psychology, philosophy, linguistics, artificial intelligence (AI), and robotics (Friedenberg & Silverman, 2006). CS' main method of investigation is the scientific method, where a hypothesis about a phenomenon is proposed, and controlled rigorous experiments are conducted to either prove or disprove the hypothesis (Friedenberg & Silverman, 2006). CS focuses on subjects such as pattern recognition, attention, memory, mental imagery, problem solving, and then some (Friedenberg & Silverman, 2006).

CS is a recent field relative to other forms of sciences such as physics, chemistry, and medicine, taking form between 1950-1970, also known as the period of the cognitive revolution (Anderson, 2015). Before this time, there were no comparable methods to the functional magnetic resonance imaging and electroencephalograms of today for studying brain activity. Brain states and brain activity were inexcessable, and thus were dismissed by some in the study of the mind for being too subjective in nature to study empirically (Skinner, 1953). After the invention of electroencephalograms, magnetic resonance imaging, and positron emission tomography, scientists were finally able to record blood flow, electrical signals, and magnetic signals in the brain while people undertook cognitive tasks, making it possible to draw connections between a brain region and a particulat cognitive activity. This made neuroscience and the study of neurons, neuronal networks, and brain regions feasible (Gazzaniga et al., 2014).

Another influence on CS was from the concept of the modern computer (Friedenberg & Silverman, 2006; Anderson, 2015; Gazzaniga et al., 2014). Computers at the time were described as using formal symbol manipulation on an input to create an output based on an algorithm. A good illustration of this concept is Alan Turing's "Turing Machine", which is a theoretical, mechanical, analogue machine which performs the job of changing symbols on a conveyor belt of slots. Each slot is represented by a symbol; say the symbol can be either a 1 or a 0. The conveyor belt is lined up with a single file sequence of 1s and 0s. The machine hangs over the conveyer, looking down at the symbols. The machine cannot change the symbols randomly, it must strictly follow a set of rules put forth to it, known as an algorithm. This can take the form of rules such as, 'if slot 352 is 1, change it to 0 and move to slot 870; if not, move directly to slot 124'. These rules are followed until the final sequence is complete and all the slots are the appropriate symbol according to the algorithm. By this, Turing concluded that any serial machine which performs formal symbol manipulation will be able to compute anything that follows an algorithm, given enough time (Turing, 1950).

This idea was intriguing to scientists, and it did not take long to translate the concept to humans, and adopt the metaphor of a human as a computer, where the hardware is the brain and the software is the mind. There are a few hypotheses that come from this idea which include the following: *The mind is the software of the brain and thus is implementation-independent* (Fodor, 1981*), and brains formally manipulate symbols* (Harnad, 1994). These are theses which computationalists use when modelling the mind (Searle, 1990).

Keeping this concept in mind, computation will be temporarily set aside in order to explain an architectural theory of the birthing process and evolution of scientific disciplines. In Thomas Kuhn's monograph, *The Structure of Scientific Revolutions* (1970), he outlines stages in which a field of science is conceived, how it grows, and evolves.

Before the conception of unified science, experiments are carried out in what is called the "pre-science" stage. In this birthing process, experiments and theories are based more on common sense rather than underlying, defining principles. It is possible for there to be as many theories about a certain phenomenon as there are experiments on it (he uses the early days of the study of electricity as an example) (Kuhn, 1970). All facts are considered equally relevant to contribute to the body of knowledge. Since there is no unifying theory, many authors must explain their work from-the-ground-up to justify their hypotheses, because facts are not taken for granted. From this, a dominant view of how the world works is assumed, known as a "paradigm".

This bleeds into the stage of what is known as "normal science". This is when scientists agree on underlying principles and theories that guide the direction of research. Much of the time the science taking place is solving the puzzles to which the paradigm draws attention. The paradigm must also possesse the characteristic of not being reducible to another theory that it can explain.

Of course, the paradigm will not have all answers to all questions. When enough anomalies, or gaps in knowledge which cannot be explained by the paradigm, accumulate to a point where the paradigm is no longer able to update itself to explain, then a crisis period occurs where there is a loss of 'faith' in the underlying principles of the current paradigm. This calls the need for a new theory to be formulated that can explain the anomalies for which the previous paradigm is unable to account. This is what is known as "revolutionary science", where a "paradigm shift" takes place. This is when ground-breaking, Nobel Prize discoveries take place that shake the foundations of the established paradigm until it is no longer honorably acceptable. The new paradigm will be able to predict and explain phenomena more accurately that its predecessor and thus is adopted, leading to a circle back to equilibrium within normal science, where the process continues to cycle.

The Kuhnian theory has been related to many scientific revolutions including those, to name a few, in optics, physics, electricity, and chemistry (Kuhn, 1970). But what if parallels were to be draw between Kuhn's theory and CS? In his monograph Kuhn mentions, "it remains an open question what parts of social science have yet acquired such paradigms at all" (Kuhn, 1970, p. 15). Let's take the bait and fall down the rabbit hole, starting in chronological order with drawing parallels to pre-science.

As mentioned previously, pre-science research is conducted without a common unifying theory or fundamental principle to guide experiments. Research is done based more on common sense, with all investigation and facts equally weighted to contribute to the body of accumulating knowledge (Kuhn, 1970). Is this the pattern seen in CS at present day? There are scientists who claim that computation currently is the favoured paradigm in play (Friedenberg & Silverman, 2006; Anderson, 2015; Gazzaniga et al., 2014;), however to accept their word as fact would be a fallacious *appeal to authority*. To challange opinion and discover truth, one makes thier own investigations and observations of the world. Therefore, in order to review a sampling of current literature in the field, a journal related to CS, *Trends in Cognitive Science*s, was accessed online May 29, 2017. Out of the 10 articles for the June 2017 issue (one was removed for having a misprint), less that half fail to mention the word "compute" (including its variations and only refer to the brain), and the only two that discuss computation substantially point out its drawbacks (Hasselmo & Michael, 2017) or outline the importance of falsifying computational models (Palminteri, Wyart, & Koechlin, 2017).

But this is just a glimpse, what about the big picture? Looking at all articles in the same journal for the year 2017 (issues from January to June), again less than half of the articles mention the word “compute” (including its variations and only refer to the brain) in the main article, and only thirty-two percent mentioned computation to a substantial extent. Does this sound like an underlying principle?

It could be argued that this is a reflection that computationalism is in fact a theory taken for granted, and the lack of acknowledgement is in itself evidence that it is a common belief that does not need to be explained in every article. This is one possibility, however according to Kuhn, a paradigm should direct research efforts (Kuhn, 1970), and the small numbers above are suggestive that research is not substantially focused towards computational models, pointing in the direction of a pre-science. This could also be a reflection of the interdisciplinary nature of cognitive science, which extends to philosophy, linguistics, and evolution, which may not easily be superimposed with computational models. However, it is suggested that such a view would be false, for there are computational theories of mind in philosophy (Friedenberg & Silverman, 2006), models of computation in linguistics (Harnad, 1994), and evolution (Friedenberg & Silverman, 2006), and so on.

There is another reason the topic of computation seldom arises in these journal articles. Some of the articles may discuss related topics such as “Mixed Signals: One Separating Brain Signal from Noise”, “Social Learning in the Medial Prefrontal Cortex”, and “The Neurobiology of Human Attachments”. It is granted that the first of the three is a paper on methodology and by nature would likely not have theory content, and there is sure to be more papers such as this sprinkled throughout the journal. Articles like the latter two mentioned above can have computational theory applied to them, devising hypotheses on how social learning is encoded differently than other types of learning in different brain regions, and perhaps physiological differences (and thus maybe computational differences) in different forms of human attachment. Without the formulation of these hypotheses, however, the above investigations echo the qualities of what Kuhn considers to be the ground work in pre-science, without any particular guidance.

On the topic of unifying principles, remember again the underlying foundations of computationalism. Entertain the idea for a moment that computation could be a paradigm for CS, ignoring the evidence above. Are there underlying principles within computational theory upon which scientists agree? The proposed theses drawn from computational theory are: *the mind is the software of the brain and thus is implementation-independent*, and *brains formally manipulate symbols*. But even these are not agreed upon, which will soon become transparent.

Take the first thesis: *the mind is the software of the brain and thus is implementation independent*. Which is a rearrangement of the argument that brains are not necessary for minds. Searle argues against this notion in his piece *Is the Brain's Mind a Computer Program*. He asserts that the idea of strong AI, that one day AI will be indistinguishable from human intelligence, is impossible because a machine must represent all the relevant casual capacities at least equivalent to that of the brain to have a mind, including the biochemical (1990). This is not a shared opinion if one were to inquire a person who believes in functionalist theory of mind. From the machine-state perspective, minds do not require brains because mental states cannot be reduced to physical states, and therefore a computer could very well possess a "mind"

(Friedenberg & Silverman, 2006). Obviously, there are conflicting opinions on the principle of the mind as implementation-independent.

Moving on to the second thesis of computationalism, that *brains formally manipulate symbols*, in other words, that cognition is equivalent to computation. Searle again criticizes this thesis, asserting that it is impossible to derive semantics from syntax, formal symbol manipulation is exclusively syntactical, and thus brains do not perform formal symbol manipulation (1990). Harnad agrees with Searle on this point, stating that syntax does not allow for a symbol to be grounded, which does not allow semantics, and as soon as symbols are grounded, it is no longer formal symbol manipulation (1994). Paul and Pat Churchland challenge this by arguing that one information processing unit is equivalent to one neuron, and though one unit or one neuron does not extract semantics, the nervous system as a whole arrives at semantics (1990). Thus, although it may take 10^{14} units, since the nervous system has 10^{11} neurons with over 10^{13} connections, the nervous system and thus an equivalent network of units can achieve semantics (Churchland & Churchland, 1990). Again, this is an example of a disagreement on the principles of computationalism.

Computationalism aside, there must be *something* which CS scientists can agree upon. Chater & Brown (2008) suggest two principles which may be eligible to become universal laws in CS,

> the simplicity principle (which states that the cognitive system prefers patterns that provide simpler explanations of available data); and the scale-invariance principle, which states that many cognitive phenomena are independent of the scale of relevant underlying physical variables, such as time, space, luminance, or sound pressure (Chater & Brown, 2008, p. 36).

An example of the first could be the heuristic approach, where the brain uses simple rules and generalizations to make fast responses instead of using rules of logic, which takes more time (Friendenberg & Silverman, 2006). Moreover, the scale-invariance principle is relevant for any non-linear function, such as memory decay (Chater & Brown, 2008). An example of the scale-invariance principle is the ease that people can recognize melody independent of octave (Chater & Brown, 2008). Even in their own paper, however, they acknowledge that CS looks to computer science and its other contributing disciplines as sources for theoretical hypotheses, which (unlike physics) do not have a treasure trove of universal laws. Indeed, it can be hard to think of laws within each discipline that can be translated to all other allied disciplines.

There is one argument from the opposition which does hold some legitimacy. Kuhn states in his monograph, “the formation of specialized journals, the foundation of specialists’ societies, and the claim for a special place in the curriculum have usually been associated with a group’s first reception of a single paradigm” (Kuhn, 1970, p. 19). Indeed, there are journals for CS, such as the journal mentioned in one of the above discussions, institutions for CS in some universities, and the Cognitive Science Society. This is a fair point. To this, it is replied that Kuhn uses Isaac Newton’s “Principia” as an example of work that has the characteristic of a paradigm (Kuhn, 1970). Isaac Newton was alive from 1642-1727, and his work “Principia” was published in 1687. Following the logic of the opposing argument, there should be an establishment of a society or institute, however the Institute of Physics was not established until 1960, after being merged with the Physical society established in 1874 and the institute of Physics established in 1920. If Newton’s “Principia” was accepted as the first paradigm, as Kuhn suggests, the establishment of the Societies and Institutions of Physics seem late to the party. This is the same case with another one of Kuhn’s examples, Antoine Lavoisier’s “Chemistry”. Lavoisier was alive from the years

1743-1794, with his publication the year 1789, and the Royal Institute of Chemistry founded in 1877. Although these two senarious are both examples of societies established much later than the accepted paradigm, the point to be made is that attention to wording is important, as Thomas Kuhn artfully says in the quote mentioned above, “have usually been” (Kuhn, 1970, p.19). This implies that a society is not *always* indicative or necessary for the acceptance of a first scientific paradigm.

To sum up, in the absence of fundamental principles and a unifying theory, CS cannot be in the stage of normal science. The last chapter to analyse, then, is revolutionary science, where there is a shift from an old paradigm to a new paradigm. Turning to the obvious low-hanging fruit, it can be said that since CS has no paradigm to begin with, it is not possible to switch from an ‘old’ to a ‘new’ paradigm. In case the fantasy of computationalism being a paradigm is still floating in the air, it will now be exorcized beyond a shadow of a doubt.

One factor which drives the transfer to a new paradigm is the accumulation of anomalies which the present paradigm fails to sufficiently explain (Kuhn, 1970). Anomalies which cannot be accounted for by computational models include the plasticity and dynamics of neurons and neuronal networks (Churchland & Sejnowski, 1990), and the influence of the body’s perceptual system and the environment (Smith & Thelin, 2003; Clark & Clancy, 2016). Connectionist and dynamic systems theories account for the former, and theories of embodiment account for the latter.

Connectionist models may use artificial network simulations to gain insight on how the neuronal networks of the brain may function. These simulated networks consist of units used to represent neurons, which are networked with one another through weighted connections. Thus, information is not represented as a symbol, but as an idiosyncratic network pattern of activation

(Friedenberg & Silverman, 2006). The advantage of this method is networks often have what are called "hidden layers" which send electronic signals to activate units in the next hidden layer or the output layer. This allows change of the weights in the connections to the units, which opens the door to modulation. The programme AlphaGo, for example, utilizes two network systems which have successfully defeated world champions at the game of Go (Wang et al., 2016). This victory is not possible with a simple formal symbol manipulating programme.

Furthermore, dynamic systems theories and network models acknowledge the fact that the brain is dynamic and self-reorganizes through novel activity (Smith & Thelin, 2003; Churchland & Sejnowski, 1990). This is a fatal flaw for which computationalism cannot account, namely that the nervous system is plastic and neurons change (Churchland & Sejnowski, 1990). It is known that long-term potentiation happens when neurons change their DNA in order to increase or reduce manufacturing of receptor proteins, which makes learning a skill like riding a bike possible (Gazzaniga et al., 2014; Anderson, 2015; Friedenberg & Silverman, 2006). This would appear to be a substantial challenge to the validity of the hardware-software argument mentioned previously.

Moreover, embodiment theories also fill in some holes which computationalism is unable to plug. Embodiment is the idea that the physiological body has an influence on cognition (Wilson & Foglia, 2017). This is apparent when speech seems to be enhanced by bodily gestures, or when "mirror" neurons become active irrespective of physical activity (for example, whether an agent is performing, witnessing, or hearing a person rip a piece of paper) (Wilson & Foglia, 2017; Anderson, 2015). To accept this, one must accept cognition as subject to influence from the body and the perceptual system, a thesis which computation does not and cannot account for.

Finally, Andrew Clark and David Chalmers published an exciting paper with a radical perspective, *The Extended Mind*, which proposes that the mind extends beyond the barriers of the brain and that the environment plays an active role in modulating cognition (2016). They propose that when a human is in interaction with an external entity, it becomes a coupled system that can be seen in itself as one cognitive system. This also dismisses computationalism for the same reasons, namely the dynamical nature and modulation of the hardware and software. The extended mind thesis is also curious because it questions the traditional view that the mind is contained in the brain, and forces the line between mind and external world to be redrawn.

Therefore, if CS was to be in the state of a revolutionary paradigm shift, then connectionist, dynamic systems, embodiment, and the extend mind thesis all seem like possible candidates for the take-over of computationalism. However, recall that Kuhn outlines a paradigm to be a unifying principle or theory that is taken for granted, and the fact that there are at least four theories competing for the spotlight signifies the lack of unity. This travels full circle back to pre-science once again, where fact-gathering and theory generating are based more on common sense than a unifying law or theory. It should be noted that this is not *inflation of conflict*, because although there are points which the experts in CS do not agree upon, the assertion that these experts are incredible or are in question is not being proposed. Additionally, this is not an attack being made on the legitimacy of CS, only that it is in parallel to Kuhn's first stage of science evolution, pre-science.

As a final shot in the dark, the argument could be made that the inability to reproduce results from studies in CS is a reflection of an improper paradigm in place, and the apparent need of a new paradigm. A survey recently published in the journal *Nature* revealed some interesting statistics. Seventy percent of scientists tried and failed to reproduce another experiment's results, and more than half failed to reproduce their own results (Baker, 2016). Disciplines in the survey included chemistry, biology, physics and engineering, medicine, earth and environment, and "other". The inability to replicate results across fields suggests three answers, a) all these sciences currently have insufficient paradigms, b) it is a reflection of mal-practice or insufficient equipment/experiments, or c) there is a different unknown underlying cause. Although it is completely probable that all the sciences have inadequate paradigms, to argue that this is proof CS also has an insufficient paradigm would be fallacious (*appeal to probability)*. Of course, the replication crisis could be due to an interaction between all three factors.

On the other hand, it can be argued that the inability to reproduce results in CS could be the outcome of the absence of a paradigm because it is in the stage of pre-science. It is nearly impossible to account and predict the mechanics of an automobile in the absence of knowledge of how a combustion engine works. As the reader may have observed, it is indeed possible to drive an automobile without the knowledge of the operating principles of an internal combustion engine. However, this parallels the mentality of behaviorism, namely that it is possible to predict and control behavior in the absence of investigation of internal states (Skinner, 1953). This is not the goal of CS, however, because CS asks the questions *why* and *how*, and is interested in the knowledge of the mechanisms of the mind, not purely controlling the mind.

Therefore, CS cannot be in a state of scientific revolution, for if there hypothetically was a paradigm in place, there is no consensus on a new paradigm to switch to. The replication crisis could be interpreted as the indication of an insufficient paradigm and the need for a change, but it can also be interpreted as the absence of any paradigm. If CS were to be in the state of normal science, universal laws, underlying principles, and/or a paradigm would need to be agreed upon by the scientific community as granted, which is not the case. Some authorities claim computation is the main theory while studying cognition (Friedenberg & Silverman, 2006; Anderson, 2015; Gazzaniga et al., 2014), but lack of directed research and the dispute of fundamental principles within computationalism contradict their claim. The glass slipper that fits Cinderella, then, must be pre-science; with numerous theories and investigations contradicting one another and no agreed universal laws to follow, research is conducted in a pick-and-choose fashion. This is not to say that the research being done and the contributing knowledge produced by CS is by no means without merit. Perhaps a rephrasing of the label cognitive science to cognitive *sciences* would be a more appropriate reflection of its embryonic and divided nature.

References

Anderson, J. R. (2015). *Cognitive psychology and its implications* (7th ed.). New York, NY: Worth Publishers

Baker, M. (2016). Is there a reproducibility crisis? *Nature*, 533(5), 452-454

Chater, N., & Brown, G. (2008). From universal laws of cognition to specific cognitive models. *Cognitive Science: A Multidisciplinary Journal*, 32(1), 36-67. doi:10.1080/03640210701801941

Churchland, P., & Churchland, P. S. (1994). Could a machine think? *Thinking Computers and Virtual Persons*, 157-171. doi:10.1016/b978-0-12-215495-9.50012-0

Churchland, P. S., & Sejnowski, T. J. (1990). Neural representation and neural computation. *Philosophical Perspectives*, 4, 343-382. doi:10.2307/2214198

Clark, Andy; Chalmers, David (2016). *The extended mind*. InterAction: The Journal of Solution Focus in Organizations, 8(1), 48-64.

Fodor, J. A. (1981). The mind-body problem. *Scientific American*, 244(1), 114-123. doi:10.1038/scientificamerican0181-114

Friedenberg, J., & Silverman, G. (2016). *Cognitive science: an introduction to the study of the mind*. Thousand Oaks, CA: SAGE Publications, Inc.

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (2015). *Cognitive neuroscience: The biology of the mind* (4th ed.). New York, NY: W. W. Norton & Company, Inc.

Harnad, S. (1994). Computation is just interpretable symbol manipulation; cognition isn't. *Minds and Machines*, 4(4), 379-390. doi:10.1007/bf00974165

Hasselmo, Michael E. (2017). Avoiding catastrophic forgetting. *Trends in Cognitive Sciences*, 21(6), 407-408.

Kuhn, Thomas S. (1970) *The structure of scientific revolutions* (2nd ed., Vol. 2). Chicago, IL: The University of Chicago Press.

Palminteri, S., Wyart, V., & Koechlin, E. (2017). The importance of falsification in computational cognitive modeling. *Trends in Cognitive Sciences*, 21(6), 425-433.

Searle, J. R. (1990). Is the brain's mind a computer program? *Scientific American*, (1), 26-31.

Skinner, B. F. (1953). *Science and human behavior*. New York: Free Press, London.

Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences*, 7(8), 343-348. doi:10.1016/s1364-6613(03)00156-6

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, LIX (236), 433-460. doi:10.1093/mind/lix.236.433

Wang, F., Zhang, J. J., Zheng, X., Wang, X., Yuan, Y., Dai, X., . . . Yang, L. (2016). Where does AlphaGo go: from church-turing thesis to AlphaGo thesis and beyond. *IEEE/CAA Journal of Automatica Sinica*, 3(2), 113-120. doi:10.1109/jas.2016.7471613

Wilson, R. A., Foglia, L. (2017) Embodied cognition. *The Stanford Encyclopedia of Philosophy* (Edward N. Zalta ed.). Retrieved from https://plato.stanford.edu/archives/spr2017/entries/embodied-cognition/.