Abstract

This is where the abstract will go. I guess I'll mention a thing or two about the contents, what I plan to discuss and what my analysis shows.

Cognitive Science and Artificial Intelligence: An Interwoven Approach

Supervisor: Francesco Bianchini

Paolo Marzolo

June 18, 2021

1 Introduction

This is Paolo Marzolo's bachelor thesis, written as part of the three-year program in computer science at University of Bologna. The stated objective of this document is to analyze the history of Cognitive Science and Artificial Intelligence and identify how influences among the two disciplines and others led to a partially shared evolution in the overarching research topics throughout their lifespans. Other similarities will be pointed out, and some of the algorithms and concepts contained throughout the sections will be explained in detail, in order to give the reader a more complete understanding.

The structure of the document will be as follows: after this introduction, a brief glossary will introduce some of the terms that will be used in this document with a short definition; this has been included to avoid having "foundational" terms be constrained by a specific philosophy or line of research. Then, the rest of the document will develop parallel to the history of the disciplines. In the final section, a bird's-eye-view will provide additional insight, and **MAYBE** a brief discussion of the roles of symbols will conclude the contents.

2 Terms and Definitions

Before definining our glossary, it is important to understand the reasoning behind why we chose to include it: when discussing researchers' understanding of human thought, it is nearly impossible to avoid using terms that have a strong past history. As an example, "thought" could already be considered too far from a behaviorist point of view. A further example is a recent discussion that took place after a somewhat controversial paper by Nunez was published [9], questioning the multidisciplinarity of Cognitive Science as a discipline (and journal) and declaring "The prospect launched by the cognitive revolution of a unified and coherent interdisciplinary seamless cognitive science did not materialize".

Cognitive Science. As we will see in following sections, saying "definitions of Cognitive Science have evolved throughout the years" would be a massive understatement. ("Thinking can best be understood in terms of representational structures in the mind and computational procedures that operate on those structures"). Its multidisciplinary nature is uncontested from what the International Encyclopedia of Social & Behavioral Sciences [6] reports ' may have been the first published use of the term cognitive science':

> 'The concerted efforts of a number of people from ... linguistics, artificial intelligence, and psychology may be creating a new field: cognitive science'

. Even the "essential original features" identified by Gardner in 1987 [4] (summarized here as (1) necessity to speak about mental representation as a separate layer of analysis from the biological, (2) faith that the computer is central to the understanding of the human mind and (3) de-emphasizing factors such as emotions or cultural factors) would be completely or partially thrown out by contemporary scholars.

In a more recent publication[2], Cognitive Science is characterized as

> The field would be better defined as the study of 'mind as machine' ... More precisely, cognitive science is the interdisciplinary study of mind, informed by theoretical concepts drawn from computer science and control theory.

.

Not only was its definition cloudy and unstable ("cognitive science is ... a perspective, rather than a discipline in any conventional sense" [12]), but as Nunez points out its disciplines have varied wildly in which ones they are and how represented they are in the Cognitive Science enterprise. Because of the reasons outlined here, far removed form the subject of this document, we will avoid using the term "Cognitive Science", and prefer the acronym "DCS".

Descriptive Cognitive Sciences (DCS). As we mentioned, the disciplines which make up Cognitive Science are not only multiple, but subject to interpretation as well. Since the nature of this work is to compare it to the history of Artificial Intelligence, we will from this point on use the acronym "DCS", for Descriptive Cognitive Sciences, as an alternate approach to the Constructive one taken by Artificial Intelligence researchers. This is not to say that a psychologist cannot take a constructive approach to the explanation of consciousness: the only reason we chose this is because we found it to be an intuitive use of the term.

Mind. Once again, although we take notice of the history of the term, we have to select a few terms to use in our language. Hereafter, we consider the mind as the non-physical correlate of human brains: "the complex of faculties involved in perceiving, remembering, considering, evaluating, and deciding. Mind is in some sense reflected in such occurrences as sensations, perceptions, emotions, memory, desires, various types of reasoning, motives, choices, traits of personality, and the unconscious."[8].

whatever else will come up

3 A History of Influences

As mentioned in the introduction, our approach will follow the historical sequence of events, although some references or explanations may be anachronistic for clarity. In order to give a general view, we split the histories of these disciplines into broad periods: one for (more or less) every substantial shift in approach and views. Generally, every time period will mention two sides of the story: one of them will focus on DCS, and the other on AI and Computer Science.

3.1 Landscape before 1950

Although the official birth of the "Cognitive Science" institutions is in the late 1970s, reasoning about thought has been a staple in philosophical research for centuries. Because of the scope of this document, we will focus on a few important concepts, and use them to set the stage for the first large shift of ideas.

3.1.1 Mathematics and Computer Science

Some of the most relevant contributions to the "reasoning as a process" come from Mathematics and what would later become Theoretical Computer Science. We will outline some of them here, while we trace part of the history of conceiving of thought as computation, first, and computers as devices for computation, second. In this respect, the following step is to be expected: can we use devices for the computation that thoughts "work" with?

Boole's Laws of Thought and Boolean Algebra. To avoid going too deep in mathematical concepts for our purposes, we can think of Boolean algebra as the branch of algebra where the variables can be either true or false (1 and 0), and the main operations on its variables are conjuction (and, $\wedge$), disjunction (or, $\vee$), negation (not, $\neg$). Through these, logical operations can be described. In "An Investigation of the Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities", one of the author's two monographs on algebraic logic, George Boole, then mathematics professor in Ireland, introduces Boole's algebra as an extension to Aristotle's logic. In it, Boole provides Aristotle's algebra with mathematical foundations, and expands it from two-term to any-term. Boole's algebra differs from modern Boolean algebra (in Boole's algebra *uninterpretable* terms exist) and cannot be intepreppted as set operations; still, its introduction marks a step towards the formalization of laws of thought and a possible bridge between mathematical research and thinking processes (even the title of the book it was introduced in gives a very clear direction). Boolean algebra would instead be developed by Boole's successors (Jevons, Peirce, Schroder and Huntington in particular); this work allows boolean algebra to now be defined by the Stanford Encyclopedia as

> the algebra of two-valued logic with only sentential connectives, or equivalently of algebras of sets under union and complementation.

KEY CONCEPTS: laws of thought can be modeled in mathematics, using algebra.

Automata theory. The study of how automatic calculators (more properly, abstract machines **TODO: mi sbaglio?** or automata) can be used to compute and solve problems is a part of theoretical computer science research. The history of Automata Theory is especially interesting, as it will let us meet some important researchers: it features two neurophysiologists, Warren McCulloch and Walter Pitts, and is thus born from the desire of modeling human thought itself. The first model was proposed in 1943 [7], in a seminal paper that also introduced other research themes we will come back to later. A little over twelve years later, two computer scientists, Mealy and Moore, generalized the theory to more powerful machines, "Finite-State machines". The general idea behind them is this: starting from an input and a set of states, a "transition function" maps the current state and an input to an output together with the next state. They do not have any memory, and as such can only "solve" simpler problems: if used to recognize languages, they can only recognize regular ones.

More powerful abstract machines had already been proposed: Turing had introduced "Turing machines" in 1937 [16], as part of his proof of the Entscheidungsproblem. The relationship between automata "expressive power" and language complexity will be explained in later chapters. **TODO: undecidability problem already puts a stop to modeling thought as maths?**

KEY CONCEPTS: Modeling algebra, so thought, is possible through some computational structure

Cybernetics. Although in recent years the term "cybernetic" has been used to mean futuristic/sci-fi technology, Cybernetics is a transdisciplinary discipline that studies regulatory systems. The core of the discipline are feedback loops (or circular causality), where the result of action is taken as input for (choosing) future actions. Cybernetics isn't bound to any particular application, so its applications include biology, sociology, computer science, robotics and many others. Its flexible approach led to many different definitions: two early ones are the one used in Macy cybernetics conferences, "the study of circular causal and feedback mechanisms in biological and social systems"[14], and the definition by Norbert Wiener, considered the originator of cybernetics, "the scientific study of control and communication in the animal and the machine"[19]. Although the word itself was used by Plato to signify the governance of people, our interest resides in contemporary cybernetics, born in the 1940s. Before the aforementioned paper by McCulloch and Pitts, the study of feedback was considered by Anokhin in 1935 [1] (physiologist). In the same year as the McCulloch-Pitts paper was published, Wiener, together with Rosenblueth and Bigelow, published "Behavior, Purpose and Teleology" [11]: these three researchers, together with McCulloch, Turing, Grey Walter and Ross Ashby, would go on to establish the discipline of cybernetics. Wiener coined the term to denote "teleological mechanisms".

An important addition to the field would be the Von Neumann cellular automata: these are yet another model of computation part of automata theory. A cellular automata is a grid of cells (of any dimensions, but for clarity, consider a 2-dimensional one first), of which each has a finite number of states it can be in; the cellular automata evolves by moving from generation zero ($t = 0$) to the next generation ($t = 1$) following mathematical rules: the state of every cell is determined by its past state and the surrounding cells. Without going into the specific rules Von Neumann determined, this is relevant to us because it introduces the concept of self replication, soon adopted by cybernetics as a core concept. Another important contribution from cybernetics is the creation of Artificial Neural Networks, introduced in the same McCulloch-Pitts paper we mentioned earlier.

KEY CONCEPTS: Study of feedback is subject-agnostic, self replication, artificial neural networks.

Information theory and technical advances. As we have seen, theoretical advances were many and varied, but the technical advances were what drove the ability to put those in practice. Among those, we have to mention the move from electromechanical devices to vacuum tube-based computers, which gave birth to a device for controlling the connections between telephone exchanges, thanks to Flowers, in 1934. The record for the first general-purpose stored-program (as in, controlled by wires, the opposite of a stored-program computer) went to Konrad Zuse, with the Z3 machine. This machine also used a binary system, but it was not a universal computer. In 1944, the Bletchley Park cryptanalysts started using Colossus. The first Turing-complete (i.e. with the same computing ability as the Turing machine) computer was completed in 1945. It used over 18.000 vacuum tubes. The first stored-program computer, built as a testbed for new technology and design, was the Manchester Baby, ran in June 1948[3].

As part of the advances of this period, we must mention the birth and development of Information Theory. Information Theory encompasses the study of quantification, storage and communication of information, in digital form. After being introduced by Nyquist and Ralph [10], the field was firmly established by Shannon's "A Mathematical Theory of Communication" in 1948. Without going into mathematical details, its main influences include the bit as a unit of informationa and the necessity of redundancy of a source when using unreliable communication channels.

Lastly, we note that neuroscience had new tools at his disposal: electrophysiological techniques, such as brain stimulation, single cell recording and EEG recording [6] were instrumental to the research into localization studies (like deficits derived from brain lesions) approached by Geschwind in the 1950s.

3.1.2 DCS

The DCS landscape around 1950 was strongly rooted in Behaviorism, with hints of the revolution that was soon to come. Some of the larger influences from the Computer Science side, such as the McCullough Pitts artificial neural network we mentioned, would in fact be ignored and re-discovered at the following shift.

Behaviorism. Behaviorism emerged as the dominant school in Western psychology as a reaction to depth psychology and other forms of psychology that did not fit well with scientific experimental verification. That is not to say it was unprecedented: Thorndike presented the law of effect (using consequences to strengthen or weaken behaviour) in 1898. Still, behaviorism was introduced as "methodological behaviorism" by a 1924 publication by John Watson [18], and then further expanded by many researchers, of which we must mention B. F. Skinner.

Behaviorism, more than a way to impose empirical constraints on studying psychology, is a doctrine of how to do behavioral science itself. The Stanford

Encyclopedia identifies three claims as the roots of behaviorism (as a doctrine):

- Psychology is the science of behavior. Psychology is not the science of the inner mind – as something other or different from behavior.
- Behavior can be described and explained without making ultimate reference to mental events or to internal psychological processes. The sources of behavior are external (in the environment), not internal (in the mind, in the head).
- In the course of theory development in psychology, if, somehow, mental terms or concepts are deployed in describing or explaining behavior, then either (a) these terms or concepts should be eliminated and replaced by behavioral terms or (b) they can and should be translated or paraphrased into behavioral concepts.

These fundamental truths identify three of the various flavours behaviorism is studied in. Skinner, mentioned above, was the first to suggest that covert behavior, such as cognition and emotions, is governed by the same controlling variables as observable behavior: although focused on the third "truth", his philosophy combines all three mentioned pillars, and is described as *radical behaviorism* by skinner himself [13].

One can easily see how the philosophy itself forced the practitioners into a state of absolute experimental dependency, which constrained the concept explored to the scientific realm. At the same time, its complete rejection of mental processes (or at least their relevance to scientific study) is the complete opposite of the assumptions that were made on the "CS" side of comprehension. Other behaviorists, though, were less radical: Clark Hull was willing to put drive inbetween stimulus and response, but only to create a corresponding theory that explained it in terms of behavior[5]; Edward Tolman, instead, proposed rats navigate a maze following a mental map[15].

Cognitive signs. Just like Tolman, other cognitive-leaning psychologists proposed ideas that did not fit with the behavioral narrative. Among them, we mention some relevant ones. Gestalt psychology: refusing the behavioristic assumption that conscious experience could be considered by reducing it to the sum of it parts, proposed the principle of totality; it also proposed the principle of psychophysical isomorphism, which meant the cerebral activity was correlated to conscious activity [17]. Vygotsky and Luria pioneered "cultural-historical psychology", which noted the role of culture and language in the development of higher pschological functions; Luria, alone, also published research on individuals' thought processes as his doctoral dissertation.

Lastly, we mention Miller, who was just a trainee at Stevens's Psychoacoustic Laboratory at Harvard: he will soon become relevant, as part of the 1956 cognitive revolution.

In our exploration of the state of disciplines around 1950, it is clear that Computer Science was firmly en route to a first attempt at thought modeling though mathematical "symbols": if, as they suspected, thought was to be considered a use of (or better yet, possible to model with) algebra, then once physical computers were capable enough they would be capable of thought. On the other side of the fence, DCS was still firmly rooted in behaviorism: in their view, the entire discussion would be based on false premises which were in turn based on wrongful research; the roots of human behavior were to be found in human behavior itself, and assuming otherwise was not only useless but unscientific, as it would lead to unprovable theories and impossible experiments. At the same time, cognitive suggestions were starting to appear, challenging the general (or at the very least American) current view.

3.2 1956: A Pivotal Year

3.3 1960-1970: Great Promise

3.4 1975-1985: Ashes and Embers

3.5 1987-1993: Bodies as the Key to Minds

3.6 1993-2000: Agents and Cooperation

3.7 2000-now: Hybrid Systems: New Perspectives

3.7.1 gianandrea

4 Perception shifts

Should I merge these two?

4.1 Symbolism and Connectionism

4.2 Symbols and Subsymbols: Collect or Extract

5 Conclusion

References

[1] P. K. Anokhin. "Problems of Centre and Periphery in the Physiology of Nervous Activity". In: *Gorki, Gozizdat* (1935).

[2] Margaret A. Boden. *Mind as Machine: A History of Cognitive Science.* en. Clarendon Press, June 2008. ISBN: 978-0-19-954316-8.

[3] *Computer Resurrection Issue 20.* https://web.archive.org/web/20120109142655/http://www.cs.man.ac.u Jan. 2012.

[4] Howard E. Gardner. *The Mind's New Science: A History of the Cognitive Revolution.* June 1987.

[5] C. L. Hull. “Goal Attraction and Directing Ideas Conceived as Habit Phenomena”. In: *Psychological Review* 38.6 (1931), pp. 487–506. ISSN: 1939-1471(Electronic),0033-295X(Print). DOI: 10.1037/h0071442.

[6] “International Encyclopedia of Social & Behavioral Sciences - 1st Edition”. In: ().

[7] Warren S. McCulloch and Walter Pitts. “A Logical Calculus of the Ideas Immanent in Nervous Activity”. en. In: *The bulletin of mathematical biophysics* 5.4 (Dec. 1943), pp. 115–133. ISSN: 1522-9602. DOI: 10.1007/BF02478259.

[8] *Mind*. en. https://www.britannica.com/topic/mind.

[9] Rafael Núñez et al. “What Happened to Cognitive Science?” en. In: *Nature Human Behaviour* 3.8 (Aug. 2019), pp. 782–791. ISSN: 2397-3374. DOI: 10.1038/s41562-019-0626-2.

[10] H. Nyquist. “Certain Topics in Telegraph Transmission Theory”. In: *Transactions of the American Institute of Electrical Engineers* 47.2 (Apr. 1928), pp. 617–644. ISSN: 2330-9431. DOI: 10.1109/T-AIEE.1928.5055024.

[11] Arturo Rosenblueth, Norbert Wiener, and Julian Bigelow. “Behavior, Purpose and Teleology”. In: *Philosophy of Science* 10.1 (Jan. 1943), pp. 18–24. ISSN: 0031-8248. DOI: 10.1086/286788.

[12] Noel Sheehy and Antony J. Chapman. *Cognitive Science*. Sept. 1995.

[13] B. F. Skinner. *About Behaviorism*. About Behaviorism. Oxford, England: Alfred A. Knopf, 1974, p. 256. ISBN: 978-0-394-49201-8.

[14] M. D. Steer. “Cybernetics: Circular Causal and Feedback Mechanisms in Biological and Social Systems. Transactions of the Seventh Conference, March 23-24, 1950, New York. Heinz von Foerster, Ed. New York: Josiah Macy, Jr. Foundation, 1951. 251 Pp. $3.50”. en. In: *Science* 115.2978 (Jan. 1952), pp. 100–100. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.115.2978.100.

[15] Edward C. Tolman. “Cognitive Maps in Rats and Men”. In: *Psychological Review* 55.4 (1948), pp. 189–208. ISSN: 1939-1471(Electronic),0033-295X(Print). DOI: 10.1037/h0061626.

[16] A. M. Turing. “On Computable Numbers, with an Application to the Entscheidungsproblem”. en. In: *Proceedings of the London Mathematical Society* s2-42.1 (1937), pp. 230–265. ISSN: 1460-244X. DOI: 10.1112/plms/s2-42.1.230.

[17] Johan Wagemans et al. “A Century of Gestalt Psychology in Visual Perception I. Perceptual Grouping and Figure-Ground Organization”. In: *Psychological bulletin* 138.6 (Nov. 2012), pp. 1172–1217. ISSN: 0033-2909. DOI: 10.1037/a0029333.

[18] J. B. Watson. “The Unverbalized in Human Behavior”. In: *Psychological Review* 31.4 (1924), pp. 273–280. ISSN: 1939-1471(Electronic),0033-295X(Print). DOI: 10.1037/h0071569.

[19] Norbert Wiener. “Cybernetics : Control and Communication in the Animal and the Machine –2nd. Ed”. In: *212 p. Cambridge, Mass.: The MIT Press, 1961. Includes Index*. CUMINCAD, 1961.