Newell and Simon's Logic Theorist: Historical Background and Impact on Cognitive Modeling

Article *in* Proceedings of the Human Factors and Ergonomics Society Annual Meeting · October 2006

DOI: 10.1177/154193120605000904

Leo Gugerty
Clemson University

NEWELL AND SIMON'S LOGIC THEORIST: HISTORICAL BACKGROUND AND IMPACT ON COGNITIVE MODELING

Leo Gugerty
Psychology Department, Clemson University, Clemson, SC USA

Fifty years ago, Newell and Simon (1956) invented a "thinking machine" called the Logic Theorist. The Logic Theorist was a computer program that could prove theorems in symbolic logic from Whitehead and Russell's *Principia Mathematica*. This was perhaps the first working program that simulated some aspects of peoples' ability to solve complex problems. The Logic Theorist and other cognitive simulations developed by Newell and Simon in the late 1950s had a large impact on the newly developing field of information-processing (or cognitive) psychology. Many of the novel ideas about mental representation and problem solving instantiated in the Logic Theorist are still a central part of the theory of cognitive psychology, and are still used in modeling the complex tasks studied in human factors psychology. This paper presents some of the theoretical precursors of the Logic Theorist, describes the principles and implementation of the program, and discusses its immediate and long-term impacts.

INTRODUCTION

Fifty years ago, in early 1956, Herbert Simon told a group of students that "over the Christmas holiday, Al Newell and I invented a thinking machine" (Simon, 1996, p. 206). This thinking machine was a program called the Logic Theorist (Newell & Simon, 1956). Its thinking consisted of creating proofs for theorems in propositional logic. In fact, it could prove 38 of the 52 theorems in Chapter 2 of Whitehead and Russell's *Principia Mathematica* (1910). The Logic Theorist was one of the first, and perhaps the first, working program that simulated some aspects of peoples' ability to solve complex problems. The Logic Theorist and other cognitive simulations developed by Newell, Simon and Cliff Shaw in the late 1950s had a large impact on the newly developing field of information-processing (or cognitive) psychology. Many of the novel ideas about mental representation and problem solving instantiated in the Logic Theorist are still a central part of the theory of cognitive psychology, and many of these ideas are still used in modeling the complex tasks studied in human factors psychology. This paper presents some of the theoretical ideas that influenced the development of the Logic Theorist, describes the principles and implementation of the program, and discusses its immediate and long-term impacts.

FORERUNNERS

In his autobiography, Herbert Simon (1996) notes that he and Allen Newell were influenced by theoreticians in mathematical logic and information theory – such as Gödel (1931), Turing (1936) and Shannon (1948) – who showed that complex mathematical ideas and processes could be represented by formal systems of symbols that were manipulated according to well-defined rules. Some of these logicians, especially Turing, claimed that these formal symbol systems could be instantiated in physical machines. In the late 1940s, these ideas led directly to the creation of digital computers. Although digital computers initially were used to implement primarily numerical calculations, Turing (1950) and others felt strongly that computer implementations of formal symbol systems such as the Turing Machine could eventually exhibit a variety of complex thinking processes. In addition to this influence from formal logic, Simon also notes the influence of researchers in cybernetics – such as McCulloch and Pitts (1943) – who developed formal models of mental processes with a close connection to neurophysiology.

A number of researchers have noted that the emergence of information-processing and cognitive psychology was also influenced by a shift in the type of tasks researchers attempted to study and model. In particular, researchers shifted from the very simple tasks of behaviorist psychology to the complex tasks studied in human factors research, i.e., tasks involving problem solving, communication and use of technology. For example, after studying communication and decision-making in submarine crews in 1947, Jerome Bruner (1983, p. 62) commented "develop a sufficiently complex technology and there is no alternative but to develop cognitive principles in order to explain how people can manage it." Another key driving force in cognitive psychology, George Miller, studied how noise affected radio communication during WW II (Waldrop, 2001). Simon (1996) cites human factors work during WW II as influencing his thinking. And in work prior to the Logic Theorist, Simon and Newell studied organizational decision-making and military command & control systems.

In the early 1950s, Simon and Newell began collaborating and set themselves the goal of developing a formal symbolic system that could execute a complex thought process. They first focused on the task of chess and then moved on to geometry proofs, but later dropped both of these visual tasks because of difficulties in formalizing the perceptual processes involved. In the fall of 1955, Simon and Newell switched to modeling a less visual task – proving theorems in propositional logic.

DEVELOPING THE LOGIC THEORIST

The major insight that helped Newell and Simon understand how people generated logic proofs was to focus on peoples' heuristics. While an undergraduate at Stanford, Newell had learned about the importance of heuristics in problem solving from the mathematician George Polya. Simon and Newell discovered potential heuristics by noticing and recording their own mental processes while working on proofs. By December of 1955, they had implemented some promising heuristics in a fairly complete version of the Logic Theorist and hand-simulated the operation of this program. In January 1956, they performed a detailed hand-simulation with their family members and students acting out the various "methods" of the program. In conjunction with this work on the Logic Theorist program, Newell and Shaw worked on developing a list-processing language (IPL) that could implement the program on a computer. In August 1956, the first Logic Theorist proof was run on a JOHNNIAC computer (named after John von Neumann) using IPL. In September 1956, the first published description of the Logic Theorist was presented at the Second Symposium on Information Theory at MIT (Newell & Simon, 1956). Since the Logic Theorist was then capable of proving a number of the theorems in Whitehead and Russell's *Principia Mathematica*, Simon informed Russell of this fact, no doubt savoring the irony that their program developed based on work in symbolic logic could now generate and prove important theorems in symbolic logic.

DESCRIPTION OF THE LOGIC THEORIST

The basic principles underlying the Logic Theorist are:

- Thinking is seen as processing (i.e., transforming) symbols in short-term and long-term memories. These symbols were abstract and amodal, that is, not connected to sensory information.
- Symbols are seen as carrying information, partly by representing things and events in the world, but mainly by affecting other information processes so as to guide the organism's behavior. In Newell and Simon's words, "symbols function as information entirely by virtue of their making the information processes act differentially" (1956, p. 62).
- Symbols represent knowledge hierarchically. For example, the representation of a logic expression in the Logic Theorist was hierarchical, with elements and sub-elements. Also the processes used by the Logic Theorist were hierarchical, in that processes would set sub-goals that initiated new processes.
- Complex problems are solved by the use of heuristics that are fairly efficient but do not guarantee solution.
- The Logic Theorist works backwards from the theorem to be proved by using the heuristics to make valid inferences until it has reached an axiom.

The claim here is not that Newell and Simon initiated any of these principles, but that they integrated and applied them to develop a working system that could solve complex problems..

Knowledge Representation

A logic expression (e.g., $\sim P \rightarrow (Q \vee \sim P)$; read as "not P implies Q or not P") is represented in the Logic Theorist as a hierarchy of elements and sub-elements. The main connective (here $\rightarrow$) is the main element. Other elements include the left ($\sim P$) and right sub-elements. In this expression the right sub-element is a sub-expression, $(Q \vee \sim P)$, which has its own main and sub-elements. Each element (E) in an expression contains up to 11 attributes, including:

- the number of negation signs before E,
- the connective in E (if any),
- the name of the variable or sub-expression in E (if any),
- E's position in the expression, and
- the location of the expression containing E in storage memory.

The Logic Theorist contains two kinds of memories, "working memories" for temporary storage and "storage memory" for longer-term storage. A single working memory holds a single element and its attributes while solving a single problem. Usually one to three working memories are used. Storage memory is used for storage of expressions (e.g., axioms and proved theorems) across problems, and for temporary storage of expressions and sets of elements while solving a single problem. Storage memory consists of lists, with each list containing a whole expression or a set of elements. Each list has a location label that is used to index the list from working memories.

Information Processes

The lowest-level unit in the Logic Theorist's information processes is an "instruction." An example instruction is: "find the right sub-element of the expression in working-memory 1 and put this sub-element in working-memory 2". Instructions can also shift control of processing to other instructions, using a branching technique similar to "goto" statements in a computer program.

The next-highest level in the Logic Theorist's information processes is "elementary processes." Each elementary process is a sequentially-executed list of instructions and their associated control flow that achieves a specific goal. Elementary processes are similar to "routines" in a computer program or methods in a GOMS model (Card, Moran & Newell, 1983).

The next-highest level in the Logic Theorist's information processes is "methods." Each method is a sequentially-executed list of elementary processes, along with associated control flow. There are four main methods in the Logic Theorist, each instantiating a heuristic for proving logic theorems. The methods are:

- Substitution – this method seeks to transform one logic expression (e.g., the theorem to be proved) into another (e.g., an axiom) via a series of logically-valid substitutions of variables and replacements of connectives
- Detachment – this method implements the logical inference rule of modus ponens, that is, if the goal is to prove theorem B and the method can prove the theorems $A \rightarrow B$ and A, then B is a proven theorem. If the goal is to prove theorem B, the detachment method first attempts to finds a proved theorem in storage memory of the form $A \rightarrow B$ where the right side either matches B or can be made to match B by substitution. If successful, a sub-goal is set to prove theorem A. If A is not in the list of proved theorems, the detachment method attempts to prove A by substitution.
- Chaining forward – this method implements the transitive rule: if $A \rightarrow B$ and $B \rightarrow C$, then $A \rightarrow C$. If the goal is to prove $A \rightarrow C$, this method first searches for a theorem of the form $A \rightarrow B$ (or which can be transformed into $A \rightarrow B$ by substitution). If successful, the method then attempts to prove $B \rightarrow C$ by substitution.
- Chaining backward – in a similar manner, this method attempts to prove $A \rightarrow C$ by first proving $B \rightarrow C$, and then $A \rightarrow B$.

The highest-level information process in the Logic Theorist is the executive control method. This method applies the substitution, detachment, forward chaining, and backward chaining methods, in turn, to each proposed theorem.

Newell, Shaw and Simon (1958) saw the Logic Theorist as an example of a program composed from primitive information processes that could generate (or perform) a complex behavior. They also pointed out that information processing programs such as the Logic Theorist offer explanations of the cognitive control structures and processes underlying complex human behavior.

Performance

In one test, the Logic Theorist was started with the axioms of propositional logic in its storage memory and then presented with 52 theorems to prove from Chapter 2 of *Principia Mathematica*. The theorems were presented in the same order as in the book. Upon proving a theorem, the Logic Theorist added it to its storage memory for use in later proofs. Given these constraints, the Logic was able to prove 73% of the 52 theorems. Using a computer that took 30 ms per elementary information process, half of the theorems were proved in less than 1 minute, and most in less than 5 minutes.

EVALUATION OF THE LOGIC THEORIST

In the rest of this paper, I will evaluate the Logic Theorist by considering whether it is an artificial intelligence (AI) program or a cognitive simulation, and by assessing its immediate and long-term impacts on theory and models in cognitive psychology.

AI program or cognitive simulation?

In an article in the *Psychological Review* in 1958, Newell, Shaw and Simon pointed out that the elementary information processes in the Logic Theorist were not modeled after human thinking, and that the model was not shaped by fitting to quantitative human data. Also, the branching control structure and the list-based knowledge representation of the Logic Theorist were later determined to be psychologically implausible. These considerations support the conclusion that the

Logic Theorist does not simulate human cognitive processes, and therefore, given its intelligent behavior, is an AI program.

On the other hand, the higher-level information processes in the Logic Theorist – the methods instantiating the four heuristics – were explicitly modeled after the introspective protocols of Simon and Newell themselves. Newell and Simon explicitly claim that heuristics are a good way to model the quick but error-prone nature of human problem solving, and they used heuristics to model other kinds of problems solving (e.g., chess) around this time. In their 1958 *Psychological Review* article, Newell et al. point out a number of other similarities in how people and the Logic Theorist solve logic problems – e.g., both generate sub-goals, and both learn from previously solved problems. These considerations suggest that in terms of higher-level information processes such as heuristics, subgoaling, and learning, the Logic Theorist was a simulation of human cognition.

Immediate Impact of the Logic Theorist

The completed Logic Theorist was initially presented to the research world on September 11, 1956 at a star-studded conference, the Second Symposium on Information Theory at MIT. In addition to Newell's presentation, Noam Chomsky presented his ideas on transformational grammar, George Miller discussed limitations in short-term memory, and John Swets applied signal detection theory to perceptual recognition. Miller has called this day the "moment of conception" of cognitive science (2003, p. 142).

Other evidence for the impact of the Logic Theorist on other researchers is contained in Miller, Galanter and Pribram's book, *Plans and the Structure of Behavior* (1960), which was itself a seminal early work in cognitive psychology. This book outlines a theory of how people use plans – structured knowledge – to guide behavior, and it sketches out a formal, computer-program-like mechanism for plans based on test-operate-test-exit units. Thus, in a sense *Plans* was a generalization of the ideas that Newell, Simon and Shaw had actually implemented in the Logic Theorist. The following excerpts from *Plans* demonstrate the strong influence of the Logic Theorist on Miller et al.'s ideas.

- "The first intensive effort to meet this need … to simulate the human chess player or logician … was the work of Newell, Shaw and Simon (1956), who have advanced the business of psychological simulation further than anyone else" (p. 55)
- Referring to the use of a formalized program to solve complex problems, Miller et al. praise Newell and Simon's "demonstration that what so many have long described has finally come to pass."

Miller et al. agreed with the emphasis of Newell and Simon on heuristics as a general method for modeling human problem solving, and they describe two other heuristics used by Newell and Simon – means-ends analysis and simplification (constraint relaxation) – in work that led up their General Problem Solver.

Finally, Miller et al. anticipated and gave answers to some of the common criticisms of cognitive simulations – criticisms that are still relevant today. The first criticism is that cognitive simulations are too complex and have too many parameters to qualify as a valid, general model of behavior. Miller et al. reply that "if the description is valid … the fact that it is complicated can't be helped. No benign and parsimonious deity has issued us an insurance policy against complexity." While later cognitive modelers have agreed that parsimony is not the most important criterion for evaluating models of complex thinking (Anderson, 1983), they have also tried to reduce the number of free parameters in their models by using consistent parameter estimates across models based on empirical research in cognitive psychology (Card et al., 1983; Kieras, Wood & Meyer, 1997).

The second criticism of cognitive simulations anticipated by Miller et al. is the homunculus problem, i.e., that cognitive simulations may need to posit a smart but unexplained mental process to interpret mental representations and make decisions. Miller et al.'s response to this is that cognitive simulations solve complex problems without a homunculus, using only decision-making processes that are explicit and evident in their rules and heuristics. The third criticism focuses on how cognitive simulations are to be validated. Miller et al. suggest Newell and Simon's main validation technique, verbal and behavioral protocols, as one way of validating simulations. Later, proponents of cognitive simulation developed other kinds of data to validate models against, including human response times, error rates, and eye movements.

Long-Term Impact of the Logic Theorist

In a review of the construct of mental representation in cognitive psychology, Markman and Dietrich (2000) describe the classical view of representation in the same way that Newell and Simon did for the Logic Theorist – i.e., that information processing consists of transforming amodal mental symbols so as to guide the organism's action. Newell and Simon were key figures in developing the classical view of representation, which is still followed in a number of cognitive modeling systems, including

GOMS (Card et al., 1983), ACT-R (Anderson & Lebiere, 1998), SOAR (Newell, 1990), and EPIC (Meyer & Kieras, 1997).

BEYOND THE LOGIC THEORIST

In the 1960s, Newell and Simon continued their work on information processing programs for complex problem solving. This work was published in their book *Human Problem Solving* in 1972. In the early 1970's, Newell initiated the use of production systems as an alternative to the branching control structure of the Logic Theorist (Newell, 1973). The modular nature of productions is now felt by many to be a better description of human procedural knowledge (e.g., Anderson & Lebiere, 1998) and production systems are widely used in cognitive modeling architectures.

In addition to the switch from branching control structures to production systems, cognitive modelers have also updated a number of the other techniques used in the Logic Theorist. In the 1990's, some cognitive modelers integrated sub-symbolic, connectionist processes into classical symbol-processing models. For example, ACT-R now conditions retrieval of information from declarative memory on the flow of activation in a memory network with varying strengths of associations among nodes. Also in the 1990's, when creating the EPIC modeling architecture, Meyer and Kieras (1997) integrated perceptual and motor processes into the heretofore purely cognitive architectures.

Other changes in cognitive modeling architectures are still in progress. These include shifting from amodal symbols to symbols that include sensori-motor representations (e.g., Barsalou, 1999), and integrating emotional and stress responses into cognitive models.

However, Newell and Simon's demonstration in the Logic Theorist that an information processing program could manipulate symbols so as to perform complex problem-solving tasks is still reflected in current cognitive modeling. Also, their use of heuristics as the core of these information processing programs is still very influential.

REFERENCES

Anderson, J. (1983). *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Quin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111,* 1036-1060

Anderson. J. & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Barsalou, L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577-660.

Bruner, (1983). *In Search of Mind: Essays in Autobiography*. New York, NY: Harper & Row.

Card, S., Moran, T. & Newell, A. (1983). *The Psychology of Human Computer Interaction*. Mahwah, NJ: Erlbaum.

Gödel, K. (1931). On formally undecidable propositions of Principia Mathematica and related systems. I Monatshefte für Mathematik und Physik, 38, 173-198.

Kieras, D., Wood, S. & Meyer, D. (1997). Predictive engineering models based on the EPIC architecture for a multimodal high-performance human-computer interaction task. *ACM Transactions on Computer-Human Interaction, 4*(3). p. 230- 275.

Markman, A. & Dietrich, E. (2000). Extending the classical view of representation. *TRENDS in Cognitive Sciences, 4*(12), 470-475.

McCulloch, W. & Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5*, 115-130.

Meyer, D. & Kieras, D. (1997). A computational theory of executive cognitive processes and multiple-task performance: I. Basic mechanisms. *Psychological Review, 104*(1), 3-65.

Miller, G. A (2003). The cognitive revolution: A historical perspective. *TRENDS in Cognitive Sciences, 7*(3), 141-144.

Miller, G., Galanter, E. & Pribram, K. (1960). *Plans and the Structure of Behavior*. New York, NY: Henry Holt and Company.

Newell, A. (1973). Production systems: Models of control structures. In W. Chase (Ed.), *Visual Information Processing*, Oxford, England: Academic.

Newell, A. (1990).*Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Newell, A. & Simon, H. (1956). The logic theory machine: A complex information processing system. *IRE Transactions on Information Theory, 2*, 61-79.

Newell, A., Shaw, J. C. & Simon. H. A. (1958) Elements a theory of human problem solving. *Psychological Review, 65*(3), 151-166.

Shannon, C. E. (1948). A mathematical theory of communication, *Bell System Technical Journal,* 27 (July and October), pp. 379-423 and 623-656.

Simon H. (1996). *Models of My Life*. Cambridge, MA: MIT Press.

Turing, A. (1936). On Computable Numbers, with an application to the Entscheidungsproblem, *Proceedings of the London Mathematical Society*, 2, 230-265.

Turing, A. (1950). Computing machinery and intelligence, *Mind, 59*(236), 433-460.

Waldrop, M. M. (2001). *The Dream Machine*. New York, NY: Penguin Group.

Whitehead, A. & Russell, B. (1910). *Principia Mathematica*. Cambridge, UK: Cambridge University Press.