The Problems with Problem Solving: Reflections on the Rise, Current Status, and Possible Future of a Cognitive Research Paradigm[1]

Stellan Ohlsson[i]

Abstract

The research paradigm invented by Allen Newell and Herbert A. Simon in the late 1950s dominated the study of problem solving for more than three decades. But in the early 1990s, problem solving ceased to drive research on complex cognition. As part of this decline, Newell and Simon's most innovative research practices – especially their method for inducing subjects' strategies from verbal protocols - were abandoned. In this essay, I summarize Newell and Simon's theoretical and methodological innovations and explain why their strategy identification method did not become a standard research tool. I argue that the method lacked a systematic way to aggregate data, and that Newell and Simon's search for general problem solving strategies failed. Paradoxically, the theoretical vision that led them to search elsewhere for general principles led researchers away from studies of complex problem solving. Newell and Simon's main enduring contribution is the theory that people solve problems via heuristic search through a problem space. This theory remains the centerpiece of our understanding of how people solve unfamiliar problems, but it is seriously incomplete. In the early 1970s, Newell and Simon suggested that the field should focus on the question where problem spaces and search strategies come from. I propose a breakdown of this overarching question into five specific research questions. Principled answers to those questions would expand the theory of heuristic search into a more complete theory of human problem solving.

Keywords

action retrieval, cognitive architecture, evaluation function, goal, heuristic search, methodology, problem perception, problem solving, problem space, search strategy, subgoal, think-aloud

[1] This article is a greatly expanded and revised version of a presentation, "Getting to heuristic search", at the New Directions in Human Problem Solving workshop held at Purdue University, November 8-9, 2008.

[i] University of Illinois at Chicago. Please direct correspondence to stellan@uic.edu.

http://dx.doi.org/10.7771/1932-6246.1144

A Theory Yet to be Achieved

If the task of psychology is to explain what it is to be human, then the study of problem solving is essential. The ability to solve unfamiliar problems has played a central role in human history via technological invention as well as in other ways, and it separates us from other animals because we are not merely better at it, we are orders of magnitude better at it.

Within cognitive psychology, the ability to solve unfamiliar problems has served as a prototypical instance of the 'higher' cognitive processes. It would appear impossible to explain problem solving with associationistic theories, but the anti-associationists were on the defensive until the 1950s because they lacked a clearly articulated alternative. This situation changed in the second half of the 1950s when Allen Newell, Herbert A. Simon, and their co-workers and students launched a novel paradigm for the study of problem solving, including an empirical but non-experimental methodology and a new kind of formal theory. The Newell-Simon paradigm was laid out in painstaking detail in their monumental book, *Human problem solving* (Newell & Simon, 1972; henceforth *HPS*). Their paradigm dominated the study of problem solving for almost forty years, from their first article on the topic (Newell, Shaw & Simon, 1958) to the middle of the 1990s. Many expected their paradigm to generate a general theory of how people solve unfamiliar problems.

Today, several of the empirical and conceptual tools that were unique to the Newell-Simon paradigm have either disappeared from our current research practices or been watered down as they spread. Nobody now turns think-aloud protocols into Problem Behavior Graphs, and the term "problem space" is widely used in the loose sense of "context" instead of the specific technical meaning assigned to it by Newell and Simon. The decline of certain specific research practices was part of a wider trend. From the early 1990s and onwards, questions about problem solving ceased to drive research on higher cognition.

The purpose of this essay is to summarize the advances that made the Newell-Simon paradigm a major scientific breakthrough, pinpoint the conceptual and methodological difficulties that brought problem solving research to an impasse, and suggest a possible path forward. To explain why their paradigm constituted a major breakthrough, I begin by sketching the state of problem solving research around 1950.

The Initial Situation

From 1913 and onwards, the behaviorist approach, inspired by Edward Thorndike's work on animal intelligence and grounded in Ivan Pavlov's work on conditioned reflexes, dominated basic research on what we now call cognitive processes (Watson, 1913). The basic principle was that mental connections—associations among stimuli, between stimuli and

responses, or between successive responses—originate in the environment. The cognitive agent, the "organism", expects event B to follow event A, because (and only because) B has followed A in the agent's prior experience. The contingencies of mind are nothing more than internalized versions of the contingencies embedded in the fabric of the world. It follows that when a situation is unfamiliar, that is, has not occurred in prior experience, the agent has no choice but to act randomly, like Thorndike's cats clawing their way out of his problem boxes. This view implies that problem solving, defined as a process of finding a solution to an unfamiliar problem that is more effective than random action, cannot be a real process. The mind is exhaustively described as the ability to internalize and act out the contingencies in the environment. Miller, Galanter, and Pribram (1960, p. 6–7) referred to those who held this view as "the optimists", because they believed that the seemingly complicated phenomena of higher cognition were in fact simple and hence easy to explain.

Arrayed in indignant opposition to this view were the pessimists, especially the Gestalt psychologists, who insisted that the apparent complexity is real (Köhler, 1927). They focused on situations in which humans (and animals) initially fail to solve a problem (thereby demonstrating that the problem is unfamiliar), but nevertheless succeed eventually. Their key observation was that both humans and animals sometimes arrive at solutions without prior trial and error. The ability to construct a brand new solution proved, they argued, that the mind contributes something to the problem solving effort that goes beyond prior experience. They characterized that contribution in terms of the ability to apprehend, in a holistic fashion, coordinated relational structures—*Gestalten* in their native German—and to restructure unhelpful relational structures into more appropriate ones. Their theory was as interesting as it was incomplete: how, exactly, is a bad or inappropriate Gestalt restructured into a better, more appropriate one? (See Ohlsson, 1984a, for a detailed critique.)

The disagreements between the warring schools of psychology were philosophical and intractable. It is not clear that they could ever have been resolved through empirical investigations. In the event, conceptual developments triggered by World War II made them irrelevant. These developments included Claude Shannon's quantitative theory of information; Norbert Weiner's science of cybernetics; and the construction of the first electronic, digital computers. If it was possible to talk about machines as having limited channel capacities, feedback circles, and internal states without being unscientific, then why was it unscientific to talk about people as having a limited short-term memory, goals, and mental states?[2] By 1943, the British psychologist Kenneth J. W. Craik could write, "My hypothesis . . . is that thought models, or parallels, reality—that its essential feature is not

[2] The history of the information technology revolution has now been told in several ways, each with different emphasis and focus (Conway & Siegelman, 2005; Dyson, 2012; Gardner, 1985; Mandler, 2007). Personal testimonies from psychologists who experienced the post-war period include Mandler (2002), Miller (2003), and Newell and Simon (1972).

'the mind', 'the self', 'sense-data', nor propositions but symbolism, and that this symbolism is largely of the same kind as that which is familiar to us in mechanical devices . . ." (Craik, 1943/1967, p. 57) But it took a few years for this new Zeitgeist to take hold and spread.[3] In 1950, anyone who surveyed psychological research with an interest in how people solve unfamiliar problems still had multiple reasons to be dismayed.

There were no satisfactory answers to the deep questions of how problem solving is possible, what kind of process it is, and where new solutions come from. The only tool for formal theoretical work was symbolic logic, but logic is about drawing valid conclusions from true premises, not about how a person decides what to do when he or she does not know what to do. There were no paradigmatic achievements, no convincing explanations of behavior vis-à-vis complex problems. In the psychology of 1950, there were no research methods designed specifically to throw light on problem solving. In spite of their philosophical and theoretical differences, the warring schools did their empirical research in the same way: they ran small-scale experiments in which the participants' behaviors vis-à-vis simple problems was recorded in terms of percent correct or time to solution, both measures taken as indices of an unanalyzed common sense concept of 'difficulty.' In contrast to the situation in other sciences, the invention of new research tools was not an item on the field's agenda.

In addition, there were few attempts to study problem solving in complex, real-world situations. The supposed plan, pushed most strongly by the behaviorist school, was to research simple situations until they were thoroughly understood, and then gradually build up, hypothesis by experimentally verified hypothesis, a body of theory that could account for complex behaviors. In the half century after Watson's 1913 article, this research strategy never moved out of the starting blocks. The move up the complexity gradient remained a promissory note. In fact, the experimental situations studied in the 1950s were sometimes simpler than those that had interested the first-generation behaviorists. The Gestalters likewise studied simple laboratory tasks, nowadays often referred to as insight problems. As a consequence, what little was known about how professional problem solvers (chess players, diplomats, engineers, managers, physicians, scientists, etc.) go about solving unfamiliar problems came from introspective reports from the thinkers themselves. The famous testimonies of Albert Einstein (Hadamard, 1949/1954, pp. 43-44) and Henri Poincare (1908/1952) illustrate this genre. Almost all such testimonies were gathered in the two collections by Ghiselin (1952) and Hadamard (1949/1954).

The Research Strategy

When seen against the backdrop of the discontents in the previous two paragraphs, Newell and Simon's work appears as a spectacular advance. Inspired by the developments

[3] The impact of some wartime achievements was delayed because they were classified.

of information technology in general and the emerging digital computer technology in particular, Newell and Simon proposed a new and radical answer to the question of what counts as an explanation of a person's behavior vis-à-vis an unfamiliar problem. For the behaviorists this was not a serious question because they assumed that all cognitive functions were amenable to breakdown into sets of pair-wise associations. Hence, they had no explanatory standards that were specific to problem solving. For the Gestalters, the ultimate explanation for restructuring, and hence problem solving, was to be found in the self-organizing properties of the cortex (Köhler, 1924). Neither approach led to clear criteria for what counts as a psychological explanation.

While Craik (1943/1967) had proposed re-enactment as a substantive hypothesis about how thinking works, Newell and Simon proposed re-enactment as a methodological principle. To explain an observed behavior is to specify an information processing device that can reproduce that behavior: *"At [the information processing] level of theorizing, an explanation of an observed behavior . . . is provided by a program of primitive information processes that generates this behavior."* (Newell, Shaw, & Simon, 1958, p. 151) The radical aspect of this meta-theoretic principle is that an explanation for how people solve problems must itself be capable of solving problems. "*All information processing theories of cognition have this property: they actually perform the tasks whose performance they explain . . . they provide a rigorous test of the sufficiency of the hypothesized processes to perform the tasks of interest*." (Simon, 1992, p. 153). This *sufficiency criterion* for what counts as an explanation is rigorous: if a program doesn't run, then it is not, by this criterion, explanatory. If it runs then its behavior should match the observed human behavior, down to some desired or agreed-upon level of detail. The adoption of successful computer simulation as an explanatory standard raised the expectations pertaining to clarity and completeness far beyond prior conceptions of what an explanation in psychology should, or could, accomplish (HPS, pp. 13).

Newell and Simon did not merely state a philosophical position on the nature of scientific explanation. They followed through by programming computer systems that solved problems. Their key hypothesis was that people can solve unfamiliar problems because they can choose tentatively among alternative actions, anticipate the outcomes of the chosen actions, evaluate the outcomes, and back up and vary their approach when the evaluation is unfavorable. They called this type of process *heuristic search*. In everyday life, we use the word "search" to refer to a type of behavior, and we normally think of search as taking place in the *task environment*, i.e., the actual or physical situation of problem solver and the situations he or she can produce by acting on the current situation. Driving through an unfamiliar neighborhood in search of a given destination is an example. Newell and Simon turned search into a theoretical concept by proposing that people search

intra-mentally, 'in the head' as we say. The effects of actions are anticipated by *imagining* them being carried out, a process sometimes called *look-ahead*. The choice of actions to perform in the environment is guided by the evaluations of alternative imagined outcomes. For example, a chess player will choose among the available moves on the basis of the comparative strengths of the board configurations that result from making those moves.

Once again, Newell and Simon did not merely announce an abstract theoretical principle, but proceeded to articulate and formalize it. A heuristic search process takes place in a *problem space*. The latter is a generative representation of the set of possible solutions that a problem solver might consider for a given problem. The specification has three components: the mental *representation* of the problem itself, the *goal* to be accomplished, and the set of *actions* that a problem solver will consider in the course of solving the problem. (Because mental search can make use of hypothetical actions that cannot be performed in the physical task environment, mental representations of actions are sometimes called *operators*, a more inclusive concept than action.) Each application of an operator takes the problem solver from one *problem state* to another. The choice of operator is guided by a *search strategy*. A strategy is a collection of *heuristics*. The latter can be either *preferences* (action selection heuristics, often displayed as *condition-action rules*) or *evaluations* (state evaluation heuristics, also known as *evaluation functions*).[4] The main difference is that preferences are used to choose one action over another before their outcomes are known, while evaluations are applied to the outcomes to decide whether to continue or to back up. When a particular strategy is applied to a given problem space, it generates a *solution path*, a sequence of (real or imagined) actions. If the path ends with the goal, it constitutes a solution to the problem; if it not, the path represents a failure to solve. A given problem can typically be solved in multiple different problem spaces (but perhaps not with equal ease), and a given problem space can typically be searched by different strategies, each of which generates a different trajectory from the initial state to the goal state.

To explain why a person solved a problem in the particular way he or she did—that is, why he or she traversed the particular trajectory he or she was observed traversing—the psychologist should identify his or her problem space, articulate his or her strategy, and demonstrate in a rigorous fashion that *that* strategy, applied to *that* problem space, generates a solution path that corresponds to the observed behavior (at some level of detail). For this purpose, treatment-control group experiments and the traditional measures of percent correct and mean solution time are minimally useful. Newell and Simon turned instead to *trace data*, particularly think-aloud protocols. The purpose was to increase the temporal resolution of the empirical record so that they could follow the problem solver's

[4] Newell and Simon did not use the terms "preferences" and "evaluations" in this way, but those terms succinctly captures the difference between the two types of heuristics.

thoughts step by step and identify his or her path through the problem space. The strategy was then to be induced from the path, or regularities therein.

The particular strategy-identification method that Newell and Simon's proposed proceeds in four steps. First, the subject's problem space is identified by inspecting the trace for clues (HPS, pp. 166-172; Ericsson & Simon, 1984, pp. 263–312). What concepts does the subject use to think about the problem situation and the goal, and what actions does he or she mention as possible next moves? This step can draw upon task analysis, concurrent verbal protocols, retrospective interviews, eye movements, and yet other sources. The output of the first step is the particular problem space the subject constructed in response to the problem (which may or may not be an appropriate or useful space).

In the second step, the trace (recorded action sequence, think-aloud protocol, etc.) is translated into a path through that problem space (HPS, pp. 172-191; Ericsson & Simon, 1984, Chap. 7). This is done by interpreting each successive utterance as expressing an output from one of the problem space operators. The output of this second step is the particular path the subject traversed. This is an inferred description of the person's stream of thoughts, decisions, and actions. This step relies heavily on the temporal sequence information in the trace.

The third step aims to explain the inferred path by identifying the heuristics that guided the subject's choices in each successive state of the problem. Like the previous two steps, this step proceeds bottom up (HPS, pp. 191-230; Ericsson & Simon, 1984, Chap. 7): to identify the heuristic that controls the application of operator X, inspect all problem states along the path in which X was executed. Identify the unique features of those states. Hypothesize that the subject believes that X is the right thing to do when those features are present. Proceed similarly vis-à-vis the remaining operators Y, Z, etc. The output of this step is a collection of situation-action rules (heuristics) that select the same operators as the subject in each successive problem state. The rules constitute the desired explanation: the subject behaved the way he or she did because he or she possessed the corresponding heuristics.

There are multiple methodological issues: because the empirical record is incomplete, some internal actions might not be expressed in the trace but have to be interpolated; the situation features that trigger some operator might be impossible to identify; an operator might be triggered by more than one rule; more than one rule might apply in a particular state; and so on. The quality of the explanation is assessed by the ratio of rules to problem states, the proportion of steps that the heuristics cover, and in other ways (HPS, pp. 191-230; Ericsson & Simon, 1984, Chap. 7; Ohlsson, 1990a).

The fourth and final step in the Newell-Simon strategy identification method draws on their re-enactment concept of explanation. While the first three steps proceed inductively, the fourth step proceeds top-down. The purpose is to verify that the induced

strategy is sufficient to solve the relevant problem, and that it does in fact generate the observed behavior. This is done by turning the collection of heuristics into a computer program, deriving its behavior by running it, and comparing its behavior to the behavior of the experimental subject. The moment of truth, Newell and Simon liked to say, is when the explanation-cum-program, so painstakingly constructed out of the mess of trace data, actually runs. The methodological issues in this step include the level of detail to be used in the comparison, and the choice of similarity metric. The reader is referred to Chapter 6 in HPS for a detailed exposition of this four-step strategy identification method; a modest replication is available in Ohlsson (1990a).

This four-step method was applied by the inventors themselves as well as by their collaborators, students and the widening circle of distant disciples who, like the current author, flocked to Carnegie-Mellon University for training in this exciting new way of studying thinking. Many of these distant disciples returned home to write doctoral dissertations that bewildered the professors at their home institutions (e.g., Ohlsson, 1980a). The resulting body of work demonstrated that it was indeed possible to use trace data to construct heuristic search explanations for how people behave in a wide range of problem solving tasks, including classical puzzle tasks like Missionaries and Cannibals[5] (Simon & Reed, 1976) and Tower of Hanoi[6] (Anzai & Simon, 1979).

In fact, the Newell-Simon paradigm addressed all the discontents of the era:

(a) Deep issues. The in-principle answer to how problem solving is possible is that people can make tentative decisions, anticipate and evaluate the outcomes of actions in the mind's eye, and change course in response to unfavorable evaluations. Problem solving is neither deductive nor random. It operates by bringing heuristic knowledge to bear to reduce, as far as possible, the uncertainty associated with each successive decision, and then proceeding tentatively. In an era when discourse about thinking was still focused on associations (Maltzman, 1955), deductive inferences (Henle, 1960, 1962), or the pathologies of thought (Rapaport, 1951), the heuristic search concept provided an entirely new way to think about thinking.

(b) Paradigmatic achievement. The bulk of the 920 pages of HPS consists of three applications of the authors' paradigm to cryptarithmetic, chess, and a logic-like symbol manipulation task. The three analyses were, and remain, the most careful analyses of problem solving ever carried out. In particular, the analysis of a single think-aloud protocol for the DONALD + GERALD = ROBERT cryptarithmetic puzzle in Chapter 6 of HPS is as thorough an explanation of a problem solving effort as one could possibly wish for. It

[5] Three missionaries and three cannibals want to cross a river. They have a boat that can carry no more than two persons across. If the cannibals outnumber the missionaries at any point in the crossing process, they will attack and eat the missionaries. By what sequence of crossings can all six cross in safety?

[6] There are three discs of different sizes on one of three pegs. Move all three to another peg, without moving more than one disc at a time, and without putting a larger disc on a smaller one.

still defines the upper limit of what is possible by way of explaining the twists and turns of the stream of thoughts behind a particular problem solving effort.

(c) Rigorous theorizing. The dictum that explanations are to be cast as running computer programs raised the formal rigor of theorizing about problem solving. Expressing a hypothesis in program code forces the theorist to be complete, explicit and precise. Running the program is an intersubjectively valid way of deriving the behavior it implies. The value of a hypothesis can be measured by whether its implementation in a simulation program improves the goodness of fit between the behavior of the simulation and the behavior of the simulated person (Gregg & Simon, 1967).

(d) Theory-driven methodology. HPS is often regarded as a theoretical work and there is indeed much theorizing therein, but most of the authors' innovations pertained to the collection and analysis of empirical data. The latter include procedures for collecting and analyzing verbal protocols, a formal notation for the specification of problem spaces, Problem Behavior Graphs for describing solution paths, a rule-based representation for heuristics, and several other tools and techniques. These were special-purpose tools for research on problem solving.

(e) Real-world relevance. Newell and Simon and other pioneers in cognitive science often dismissed, in no uncertain terms, the research strategy of starting with simple behaviors: "*Our theory posits internal mechanisms of great extent and complexity, and endeavors to make contact between them and the visible evidences [sic] of problem solving. That is all there is to it.*" (HPS, p. 10) From the beginning, their studies focused on complex tasks that people solve outside the psychologist's laboratory, including chess (Newell & Simon, 1972), thermodynamics (Bhaskar & Simon, 1977), and scientific discovery (Kulkarni & Simon, 1988; Simon, Langley, & Bradshaw, 1981).

These strengths notwithstanding, the influence of the Newell-Simon paradigm began to fade in the late 1980s and early 1990s. Although their re-enactment concept of explanation was widely adopted throughout the cognitive sciences, their unique methodological innovations were not, and the four-step strategy identification method is not used today to study problem solving. The next two sections diagnose this impasse and discuss how it might be resolved.

The Impasse

The Newell-Simon paradigm encountered two closely related problems, both unsolvable: how to aggregate trace data to reveal novel empirical regularities, and how to formulate a general, task-independent theory of problem solving. The attempt to address the latter issue by re-focusing on the cognitive architecture had the paradoxical effect of leading researchers away from the study of problem solving.

The Problems with Trace Data

Adoption of the Newell-Simon four-step strategy identification method by cognitive psychologists was limited and soon stopped. Today, researchers throughout the cognitive sciences frequently use verbal protocols as data. However, those protocols are typically subject to a code-and-count methodology that aims to generate dependent variables for use in traditional experimental group comparisons (Chi, 1997). The code-and-count and the Newell-Simon methods both begin with the problem solver's utterances, but code-and-count destroys the temporal sequence that Newell and Simon regarded as the main information contributed by a verbal protocol. To the best of my knowledge, nobody has carried out the four-step method recently, at least not in a published article.

One reason is that the Newell-Simon method yields explanations that are as task-specific as the behaviors they explain. After all, a problem space is a representation of a particular problem, and a strategy is a set of preferences and norms for how to search a given problem space. The explanation they provide is of the general form, *this person acted in such-and-such a way on problem X because he or she understood X in terms of this or that problem space, which was searched by strategy so-and-so, and it turns out that applying that strategy to that space generates precisely the sequence of steps that he or she was observed making*. The first encounter with such an explanation provides the thrill of seeing the mind at work up close. After repeated exposures, a nagging thought knocks on the door to consciousness: a strategy explains an *event*, namely, a particular problem solving effort. However, the purpose of collecting and analyzing scientific data is to identify *regularities*, empirical laws against which a general theory can be tested. Newell and Simon's four-step strategy identification method does not by itself reveal previously unsuspected regularities in problem solving.

Applications of the four-step method brought home a second point that aggravated this problem: individual differences in problem solving strategies are ubiquitous and they are qualitative. Consider the reasoning problem in Figure 1. Some experimental subjects

---

A child is putting blocks of different colors on top of each other.

A black block is between a red and a green block.
A yellow block is further up than the red one.
A green block is bottommost but one.
A blue block is immediately below the yellow one.
A white block is further down than the black one.

Which block is immediately below the blue one?

Figure 1. An example of a verbally presented spatial reasoning problem to be solved without external memory aids. (Source: Ohlsson, 1990).

attack this problem by trying to visualize the ordering of all the blocks and reading off the answer from their mental model. Others try to use the given information piecemeal, to eliminate all alternative answers but one (Ohlsson, 1984c, 1990a). These are two qualitatively different ways to conceptualize the problem, leading to two different problem spaces.

This situation is entirely typical. In their book-length exposition of verbal protocols as a research methodology, Ericsson and Simon (1984) stated the problem clearly and bluntly: "*[At the level of detail provided by a protocol] we can expect to encounter significant interpersonal differences in processing. This makes it difficult to use a single computer model to predict or account for the detail of numbers of different protocols*" (p. 196). The empirical research available for review at that time strongly confirmed this observation. In task domain after task domain, Ericsson and Simon (1984) found that different individuals tend to take qualitatively different paths through the problem space (pp. 196–198).

The question then arises how to take the next step. How are multiple strategy analyses to be aggregated within and across subjects to reveal new regularities? Given a collection of solution paths, problem spaces, and strategies, what is the aggregation operation that extracts whatever is general across them? How are qualitatively different paths or strategies to be combined to form a description of problem solving that is more general than the individual problem solving performances themselves? The standard aggregation operation of experimental psychology—compute an arithmetic mean—is not applicable to these qualitative, complex, and symbolic constructions. How, then, is data aggregation to be accomplished? The four-step, bottom-up method laid out in HPS stops where an individual strategy has been identified and verified. No further step was prescribed. But without an aggregation step, the researcher arrives at the end of a long and arduous analysis without any other conclusion to report than that *this subject solved this problem in this way, and that subject solved the same problem in that way, and that other subject* . . . , and so on.

In the 1980s, a handful of researchers proposed that problem solving strategies might exhibit general properties that can be discovered by inspecting and comparing heuristics from different domains (Groner, Groner & Bischof, 1983). For example, Lenat (1983) made the interesting observation that heuristics exhibit a U-shaped relation between generality and usefulness: a very specific heuristic is useful because it provides detailed direction. For example, *to switch on a projector of brand X, push the red button located at the left front corner* tells the user exactly what to do. In contrast, a general heuristic is useful because it can be applied across a wide range of problems, even unfamiliar ones. *To turn on any electrical device, locate and press the power button* is an example. However, heuristics of intermediate generality do not seem to offer either the advantage of tight guidance or wide applicability; *to turn on any projector, press its power button* is a case in point. This regularity—I propose we call it Lenat's Rule—re-describes strategies at a higher level of

abstraction. A body of such regularities could inform and constrain a general theory of problem solving. But to the best of my knowledge, Lenat's Rule is one of a kind; no one has found a second empirical law of this sort. Either there are none, or researchers have not been looking.

In my first published research paper, I addressed the data aggregation problem by suggesting that a collection of strategies for a given problem type could be summarized in a construct I called a *strategy grammar* (Ohlsson, 1980b). The basic idea was that strategies in a domain share certain functions that are dictated by the nature of the task, but the functions might be implemented in different ways. A generative grammar therefore seemed a useful summarization tool: each component of the grammar could be broken down into its subcomponents by conjunctive replacement rules, and each subcomponent could be expanded in multiple ways according to disjunctive replacement rules. The set of strategies that could be derived from a grammar constituted the set of strategies that one would expect people to use on the relevant task, so the grammar could, in principle, predict the occurrence of behaviors that had not yet been observed. The idea was worked out for the type of verbally presented spatial reasoning problem exemplified by the problem in Figure 1. The construction of the strategy grammar is a data aggregation step; it proceeds bottom-up from a set of strategies generated through the four-step method to a more abstract, compressed description of the set of strategies.

However, even with the strategy grammar in front of me I could still not formulate a general conclusion: what does the strategy grammar *say* about problem solving? The features that are captured in the grammar—for example, the fact that all the strategies for this type of task have some criterion for when to consult the problem text—are more general than the specific heuristics, but they do not seem to be regularities in problem solving behavior. Conference audiences critiqued my technique, correctly I now think, for being "too syntactic" and I never did a second strategy grammar analysis.

The aggregation problem was not the only factor that limited the dissemination of Newell and Simon's four-step strategy identification method. Their four-step method is closer in spirit to archeology and natural history than to the experimental methodology of laboratory sciences like chemistry and physics. In natural history, you collect interesting specimens, dissect them carefully, and report what you find. This style of inquiry differs in several respects from the hypothesis-testing paradigm of experimental psychology: the researcher doesn't necessarily start with any hypothesis, although he or she might have a question in mind (*What kind of animal is this? What was this building used for? What is the central difficulty in this problem?*). There is typically no prediction and no manipulation of any independent variable; instead, there is the selection of an interesting specimen to dissect. The report of the dissection is long and full of factual details; it aims to be as complete as possible. If there is no initial hypothesis, then the conclusion cannot be about

falsification or verification; indeed, there is often no single-sentence conclusion. Rather, the thick description of the dissected specimen *is* the result of the study. Developing its implications, if any, for theoretical principles is not the sole responsibility of the researcher performing the dissection but a long-term responsibility shared by all researchers in the relevant field. The purpose of the report is to put all the facts about the dissected specimen on the table as raw material for theorizing. Although thick description is a perfectly respectable scientific activity, the fact that Newell and Simon's four-step method departed so radically from the standard methodology of experimental psychology affected the rate and extent of dissemination.

Practical considerations also hindered the widespread and continued use of the four-step method. Those of us who tried the method soon discovered that this is a labor-intensive enterprise. A 20-minute think-aloud protocol can take ten times as long or more to transcribe and segment into individual utterances. The identification of the solution path and the invention and verification of the relevant strategy is likely to take longer. A single problem solving behavior—a single data point—can thus require 10–30 hours of analytical work. The labor intensive nature of the method is an obstacle for young researchers setting out to meet the ever-escalating demands on scholarly productivity. For example, the analysis in Chapter 6 of HPS of a single 20-minute think-aloud protocol for the DONALD + GERALD = ROBERT cryptarithmetic puzzle runs to 95 pages; VanLehn (1991) spent 47 pages reporting the analysis of a 90-minute protocol; and Ohlsson (1990a) required 46 pages to report the analysis of a single 4-minute protocol. In general, the number of pages required to describe even a modestly complex problem solving behavior is likely to shock any space-conscious journal editor into rejection mode. Responsible Ph.D. advisors steered their students away from this enterprise, and inter-generational transmission of the four-step method was interrupted.

In retrospect, the main point of carrying out the four-step strategy identification process was to prove that it could be done; that is, to show that the higher-order cognitive process of solving an unfamiliar problem could indeed be analyzed into a sequence of information processing steps that could be re-enacted by a computer program. This was a marvelous achievement at the time, even though subsequent progress in the cognitive sciences makes it seem mundane today. As each successful simulation of problem solving made that point one more time, it became less and less clear what was gained by making it yet again. It is highly implausible, Lenat's Rule notwithstanding, that we will ever discover any general properties of problem solving strategies.[7] Strategies are as infinitely variable as the tasks to which they apply, and for any one task they vary from person to person

[7] We can reconceptualize Lenat's Rule as an observation about the world rather than about human psychology: We live in a universe such that the rule holds. If we lived in a different universe, heuristics of intermediate generality might be the most useful, and Lenat's Rule would not be true.

as well as over time. Fitting simulation models to the step-by-step moves of individual experimental subjects is a case of overfitting theory to local and largely meaningless variations in data.

The Problems with General Mechanisms

The goal of research on problem solving is a general theory, that is, a set of principles that capture the essential properties of the cognitive processes by which human beings solve unfamiliar problems and explain any empirical regularities in their behavior while doing so. The principles should give researchers a deeper understanding of how people are able to solve problems, why some problems are more difficult than others, why some individuals solve them more efficiently than others, why people at the same performance level solve them in qualitatively different ways, how problem solving ability can be trained, and so on. In Newell and Simon's work, the principle of heuristic search was central. Almost everything they wrote about problem solving concerned, directly or indirectly, the question of how cognitive agents bring knowledge to bear to constrain and guide search. But this is a single, abstract principle. Does it capture everything there is to say about problem solving at a general level, or can the theory be fleshed out with auxiliary principles? What might those auxiliary principles be like? What *kind* of principles should they be?

Newell and Simon's first attempt at a general theory was to posit that there exists an *Ur*-strategy, a search mechanism that can be applied to any task whatsoever. The search strategies discovered in empirical analyses are task-specific instances of this general mechanism. This concept of generality guided the design of their second problem solving program, which was consequently called the General Problem Solver (GPS; Ernst & Newell, 1969). The specific strategy implemented in this system was called *means-ends analysis.* It works by (a) computing an important difference between the current problem state and the goal state; (b) looking up that difference in an difference-operator table which specifies which operator (action) is useful for reducing which type of difference; (c) choosing among the retrieved operators; (d) applying the chosen operator; and (e) iterating until all differences between the current state and the goal have been eliminated. GPS was successful in multiple task domains, and when UNESCO organized the first international conference ever on information processing in Paris in 1959, GPS was the closest thing to an operational artificial intelligence that the world had yet seen (Newell, Shaw & Simon, 1960).

From the point of view of psychology, this approach to generality failed. In many task domains, people do not engage in means-ends analysis but use forward search, hill climbing, reasoning by analogy, or some other type of strategy. For example, Greeno (1974) found that GPS did not provide a good explanation for what he called the "Hobbits and Orcs Problem" (more commonly known as the Missionaries and Cannibals Problem;

see footnote 5). He found that people construct the solution to this problem in terms of short sequences of forward-looking, correct moves, not in terms of a stack of subgoals that successively reduce the differences between the current problem state and the top goal.

In addition, the GPS mechanism was not in fact task-independent: the difference-operator table was a crucial but task-specific component that had to be re-created anew for each type of task. It was not clear how the difference-operator table should be interpreted in a psychological context. (To claim that it was learned in prior experience was not an option for a theory of how people solve problems for which they have no prior experience.) The ambition to explain problem solving in complex real world domains also highlighted the importance of task specific knowledge. Attempts to program computers to re-enact the performance of experts showed that one can do so without much attention to general principles of problem solving, if one can identify the specific knowledge the expert brings to bear (Buchanan & Feigenbaum, 1978; Feigenbaum, 1989; Shortliffe, Axline, Buchanan, Merigan, & Cohen, 1973). Means-ends analysis, in the precise form in which it was implemented in the GPS program (Ernst & Newell, 1969), faded from view, and a related but simpler concept of backward chaining took its place as one strategy among others in a repertoire of general strategies (later known as *weak methods*). Attempts were made to re-capture generality by specifying a universal weak method from which all the weak methods can be derived (Laird & Newell, 1993), but this concept has so far had little impact on psychology.

If there is no general, task-independent problem solving mechanism, then what type of general principles about problem solving are researchers supposed to look for? Newell and Simon's second approach was to attribute generality to the machinery that executes the various strategies. If a strategy is seen as a piece of software, it becomes natural to view the brain as the hardware; the mind then becomes the operating system. Although originally called "the information processor" in HPS, this entity came to be known under the catchier title *the cognitive architecture* (Anderson, 1983). The architecture concept slices cognition into the general, constant, and presumably innate basic processes, on the one hand, and the acquired and infinitely variable task strategies on the other. A general theory is limited to the former; there is no reason to expect generality in the latter. The reader can see this intellectual move emerge in the theory chapter in HPS (pp. 788–808), and it was fully explicit in two papers by Allen Newell that introduced the first implemented production systems architecture (Newell, 1972, 1973).[8]

The move was successful. There now exist several serious theories of the human cognitive architecture (Anderson, 2007; Langley, Choi, & Rogers, 2009; Newell, 1990; Sun, 2007).

[8] One can argue that Miller, Galanter, and Pribram (1960) was the first attempt to describe the cognitive architecture, but they did not explicitly discuss the architecture concept itself, and their description did not result in an implemented model.

These proposals are computer systems that can be used as high-level, special-purpose programming languages for modeling human cognition. In some cases, researchers have been able to predict quantitative performance measures with high accuracy using this type of simulation tools (Anderson & Lebiere, 1998). Lately, architecture theorists have begun to ground their proposals in neuroscience, mapping components of their architectures onto particular brain centers (Anderson, 2007; Just & Varma, 2007).

From the point of view of problem solving research, the architecture approach to generality comes with a price: the general principles do not address problem solving per se. The cognitive architecture is the machinery that underlies all of cognition, including attention, language, learning, memory, perception, and so on. The architectural processes are basic in the sense that they constitute the computational substratum in which all the higher cognitive functions are implemented. For example, a key principle of several architectural proposals is that working memory has limited capacity. Another is that the probability of retrieving an item from long-term memory is a function, in part, of past attempts to retrieve that item. These properties of the architecture influence problem solving (so we believe), but they affect any other type of cognitive processing as well, and hence are not principles of problem solving per se. Precisely because they are equally true of problem solving, discourse comprehension, decision making, mental arithmetic, and any other type of thinking, they do not say anything specific about how people go about solving an unfamiliar problem. In the architectural framework, problem solving has no independent description; heuristic search can only be understood as the composite result of thousands of applications of the basic processes. If there are as yet undiscovered principles of problem solving, they cannot be stated in the conceptual vocabulary of the cognitive architecture.

The components of the architecture are also basic in the sense of being of short duration. Estimates range between 50 milliseconds and a few tenths of seconds (Newell, 1990). The focus on the 50 ms time scale affects the style of empirical research. Competitive testing of different architectural proposals requires precise quantitative measures. To make such tests rigorous, researchers have to focus on simple tasks in which behavior only lasts for fractions of a second or at most a few seconds. The Stroop task, working memory span tasks, and the anti-saccade task exemplify a large class of tasks that intuitively appear to exercise the basic processes of the cognitive architecture. These tasks are not problems in the sense in which that term is used in problem solving research. There is no goal that the subject does not know how to reach; on the contrary, experimental subjects are instructed and even trained in what to do. The ambition to ground the cognitive architecture in quantitative measures pulled researchers' attention away from the analysis of qualitative, information-rich traces of complex problem solving behaviors, the very enterprise that Newell and Simon pioneered.

In summary, Newell and Simon's first concept of generality, codified in the General Problem Solver, failed as a psychological theory because it is not true: there is no single problem solving mechanism, no universal strategy that people apply across all domains and of which every task-specific strategy is a specific instance. Their second concept of generality initiated research on the cognitive architecture. The latter is a successful scientific concern with many accomplishments and a bright future. But it buys generality by focusing on a time band at which problem solving becomes invisible, like an elephant viewed from one inch away.

Newell and Simon's many contributions continue to influence the cognitive sciences as well as economics and the philosophy of science. However, the influence of their most unique methodological innovation, the four-step strategy identification method, waned as the problem of data aggregation remained unsolved and the problem of generality was reformulated in terms of the cognitive architecture. Researchers, including Newell and Simon themselves, began posing new questions that required different empirical pursuits.

Impasse Resolution

A natural response to the decline of problem solving research is to reach for a new theory that can inspire a new research agenda. But research did not decline because the principle of heuristic search was falsified. There are no data that disprove the claim that people engage in heuristic search, and much data that support it. Indeed, heuristic search is unlikely to ever be falsified in the Popperian manner (Ohlsson, 2011): when a problem is unfamiliar, decisions are necessarily tentative, so wrong choices are unavoidable. When an outcome is unfavorable, a cognitive agent—animal, human, or machine—has only two options: try something else or give up. Any agent that successfully solves an unfamiliar problem must be capable of varying its behavior in the face of negative outcomes. But tentative actions, evaluation of outcomes, and variability of action are the key components of heuristic search. Hence, heuristic search is not an empirical hypothesis but a necessary component of any system that can solve unfamiliar problems. If so, heuristic search will be a central component of any future theory of problem solving. Newell and Simon's enduring contribution is to have articulated this necessary truth and given it a precise, formal expression.

If we retain the theory of heuristic search but reject strategy identification as the proper pursuit for empirical studies, what are researchers supposed to do? Which alternative pursuits would deepen our understanding of how people solve unfamiliar problems beyond the state of the theory at the end of the 1970s? One answer is suggested by the observation that the theory of heuristic search is seriously incomplete. To search, one has

to construct a search space and assemble a search strategy. Where do problem spaces and search strategies come from, keeping in mind that "prior experience" is not an admissible answer in the case of an unfamiliar problem? Simon and Newell (1971) stated the problem concisely: *"The initial question we asked in our research was: 'What processes do people use to solve problems?' The answer we have proposed is: 'They carry out selective search in a problem space that incorporates some of the structural information of the task environment.' Our answer now leads to the new question: 'How do people generate a problem space when confronted with a new task?' Thus, our research, like all scientific efforts, has answered some questions at the cost of generating some new ones."* (Simon & Newell, 1971, p. 154)

Although they posed the problem of where problem spaces and strategies come from clearly enough, their discussions of it were either short (Simon, 1978, pp. 284–286; Simon & Newell, 1971, pp. 155–156) or unconvincing (Newell & Simon, 1972, pp. 847–867). They never formulated a research agenda for this problem of comparable originality, scope, and power to their agenda for problem solving proper. They said as much: "*This part of [our] theory is both tentative and incomplete*" (Newell & Simon, 1972, p. 847). After 40 additional years of research, it might be fruitful to once again try to formulate such an agenda. In the following, I break down the overarching question of where problem spaces and search strategies come from into five specific research questions.

Where Do Problem Spaces Come From?

A problem space is defined by mental representations of the initial problem situation, a set of relevant actions, and a goal. The question of origin arises with respect to each component.

Problem perception

The perception of a problem presumably engages the same perceptual apparatus as all other forms of perception. We do not have one mental process for recognizing a pair of pliers as a pair of pliers in an unproblematic situation, and a second process for recognizing the pliers as such in a problem situation. Perhaps problem perception is nothing but ordinary perception, applied to problem materials. If so, the question of how people perceive problems might be answered by a general theory of perception, and there is nothing to say about problem perception per se.

On the other hand, it is possible that there are regularities in problem perception that are not salient in the perception of other types of situations and hence might not be addressed in research on perception per se. Consider the *insight sequence*: when faced with a problem that requires a creative response, a person sometimes experiences successive alterations of mode and tempo of their thinking: there is steady progress in exploring the initial options; there is an impasse and possibly cessation of problem solving activity; a

new option comes to mind; and search resumes. From the point of view of heuristic search, such alterations are mysterious. Why would a search process run into an impasse? How could such an impasse be resolved? The explanation for the insight sequence that I and others have proposed is that the initial perception of the problem did not lead to a problem space in which the problem can be solved. By re-perceiving the problem, the initial state and hence the problem space is revised, possibly bringing previously unheeded options to mind (Ohlsson, 1984b, 1990b, 1992, 2011).

The question is how re-perception is accomplished. What cognitive processes and mechanisms underlie this phenomenon, and which general principles do they instantiate? Perception researchers might not give a high priority on this problem, because they are busy explaining how anything can be perceived at all. Also, everyday perception is characterized by the striking stability of our percepts under ever-changing conditions of distance, lighting conditions, viewing angle, and so on. Hence, perception researchers might see explaining stability as a more central problem than explaining the variability that is required for insight. I have proposed a theory of re-perception elsewhere based on a redistribution principle; the reader is referred to Ohlsson (2011, Chap. 4) for details and to Ohlsson (2008) for a model of one component of re-perception. The point for present purposes is that understanding re-perception is more important for understanding problem solving than for understanding perception per se, so a comprehensive theory of problem solving needs to include auxiliary principles to explain regularities that pertain to this process.

Action retrieval

To solve a problem is to do something about it. In many cases, thinking of the right action is the key. The competence of an average adult encompasses hundreds, perhaps thousands of actions (grasp, hit, roll, throw, etc.), so the set of actions considered during heuristic search will necessarily be a small subset. How do people know which subset to activate in the context of an unfamiliar problem? It is possible that action retrieval is nothing but a specific case of memory retrieval. If so, the answer to the question will be forthcoming from memory research, and there is nothing specific to say about action retrieval in the context of problem solving.

Although there are thousands of experimental studies of memory, most of them concern the retrieval of declarative, episodic, or autobiographical content. Very few have been concerned with the representation and retrieval of actions. It is possible that there are phenomena that pertain to action retrieval per se. For example, consider *functional fixedness*, the tendency to retrieve only the most frequent and familiar actions that can be performed on an object (Adamson, 1952; Birch & Rabinowitz, 1951; Duncker, 1935; German & Barrett, 2005). This phenomenon is useful for understanding behavior vis-à-vis

problems in which an object is to be used in an unusual or novel way, such as the pliers in the Two-String Problem.[9] Functional fixedness was discovered in the course of research on problem solving, not memory, but the general problem of why we do not always retrieve the action or actions that would be useful to consider is shared with research on mindlessness (Langer, 1989) and mental ruts (Smith, 1995).

Gibson (1977) invented the useful term *affordances* to refer to the set of actions that can be performed on an object or situation. This concept has turned out to be useful in robotics (e.g., Sun, Moore, Bobick, & Rehg, 2010). A synthesis of ecological perception, robotics and problem solving is likely to reveal other, as yet undiscovered regularities with respect to action retrieval. The principles we invent to explain them might constitute yet another component of a comprehensive theory of problem solving.

Goal setting/Problem finding

It is common to describe complex behavior as goal-driven and hierarchically organized. The ability to solve unfamiliar problems poses the question of how people unpack a goal into its subgoals *for the very first time*. The standard answer has its roots in the General Problem Solver and it is implemented in cognitive architectures such as Icarus (Langley, Choi & Rogers, 2009): retrieve the actions that have the current goal among their consequences; identify the preconditions of those actions; choose one action and pose its preconditions as conjunctive subgoals; iterate until the current subgoal can be accomplished with a single primitive action; and, if a precondition cannot be accomplished, select another action.

The question that remains unanswered is where the top goal comes from. In a laboratory setting, the experimenter informs the subject about the goal (e.g., *move all four discs to Peg C*). But in everyday life, people have to conceptualize the problems they face on their own. Why does one person tackle a task that others reject (e.g., *let's build a flying machine*)? Why do people set themselves different goals in response to one and the same problem situation (*the job is to reconcile astronomical observations with the circular orbits of the heavenly bodies* versus *the job is to figure out the true shape of those orbits*)?

Viewed in this way, goal setting is closely related to problem finding and problem recognition (Runco, 1994). The latter is a neglected research topic, perhaps because it takes the researcher out of the laboratory, which makes studies more complicated and resource demanding. The main empirical regularity with respect to problem finding that has been documented so far is that people differ in their ability and disposition to recognize a fruitful problem. It is likely that there are other, as yet undiscovered regularities in problem finding behavior and that novel theoretical principles are needed to explain them.

[9] Two strings hang from a ceiling, too far apart for a person to reach one while holding the other. The goal is to tie the strings together. The room is bare, except for a chair, a pair of pliers, a newspaper, and an umbrella.

Where Does the Initial Strategy Come From?

Action selection

Heuristic search is not random trial and error. We know this because actions are not evenly distributed over the possible options even in unfamiliar problem spaces. For example, the first time a person sees the Nine Dot Problem[10], he or she draws almost all lines inside the square formed by the nine dots, and among those lines there are many more vertical and horizontal than diagonal lines (Kershaw & Ohlsson, 2004). The question is why a person walks into a psychologist's laboratory with such preferences in place. If he or she has no prior experience of the problem, how come the person has any preferences at all?

The standard answer in the cognitive psychology literature is that prior experience of past situations, or types of situations, inserts itself into the current situation through a process called *transfer*, defined as the application of knowledge acquired in one situation to another, qualitatively different situation. There are two traditional explanations for transfer. The principle of abstraction has been with us since Plato and Aristotle debated the nature of knowledge on the streets of Athens, and it says that knowledge learned in a past situation applies to a current situation because what was learned is abstract, i.e., ignores some features of the situation while retaining others. In contrast, Thorndike's identical elements principle claims that what is learned in one situation applies to a future situation to the extent that the two situations are identical (Thorndike & Woodworth, 1901). Other theories of transfer have been proposed more recently (Nokes, 2009). If transfer during problem solving happens in the same way as in other types of situations, then the explanation for initial action selection biases will be forthcoming from transfer research and there is nothing specific to say about action selection in unfamiliar situations.

However, on closer inspection, transfer principles throw limited light on action preferences during problem solving. What would the identical elements or the abstractions be that supposedly make people prefer to draw lines within the square and prevent them from drawing diagonal lines when solving the Nine Dot Problem? Either principle seems an implausible explanation for these peculiar biases. Other transfer theories fare no better. For example, what analogy could be operating in this case? Subjects' initial choices in this and many other problem spaces appear to be shaped by unconscious dispositions that have little rational basis. We understand next to nothing about the origin and operation of such dispositions. It is plausible that there are as yet undiscovered regularities with respect to initial action selection and that new theoretical principles are needed to explain them.

[10] Faced with a 3 by 3 arrangement of nine dots, draw four straight lines that pass through all nine dots without lifting the pen and without back tracking.

Outcome evaluation

Wrong moves are unavoidable when solving an unfamiliar problem, by definition of "unfamiliar", so search is necessarily guided, in part, by the evaluation of action outcomes. The question arises how the problem solver decides whether his or her last step brought him or her closer to or further from the current goal. An *evaluation function* is a mapping from problem states to some metric of promise such that the problem solver can decide whether an action was a step in the right direction. The prototype for such a function is the method that chess players use to assess the strength of a board position. Chess players apply this method, or some mental version thereof, when deciding which move to make (Holding, 1985, Chap. 8).

Where does such an evaluation function come from? The answer is unlikely to be forthcoming from research on judgment per se. Because evaluation functions are task specific, one plausible hypothesis is that people derive such functions from goals. For example, in Missionaries and Cannibals the goal is to ferry people from one bank of a river to the other. It seems to follow that problem states with more people on the far side of the river are closer to the goal than those with fewer people on the far shore, so the relative value of two problem states can presumably be determined by counting the

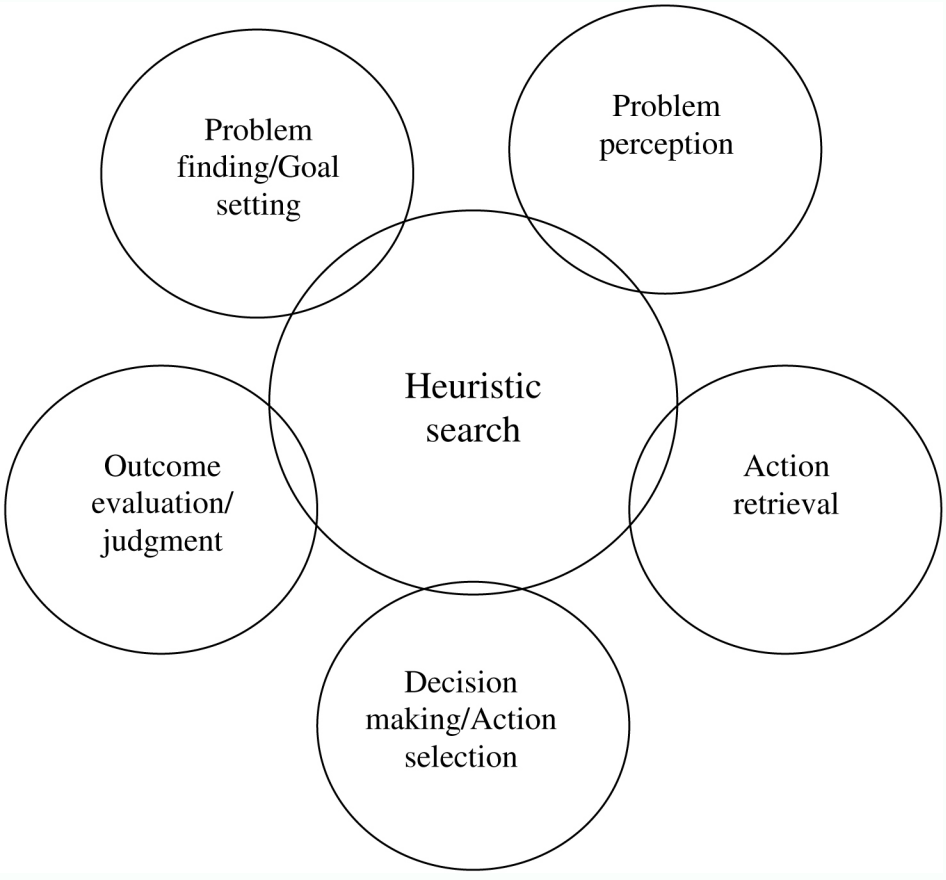

Figure 2. The structure of a hypothetical future theory of problem solving.

number of people on the far shore. This is a reasonable initial approach, but it turns out to be incorrect in some problem states, which explains why even very smart people can remain stuck on this little puzzle for minutes on end (Simon & Reed, 1976). The general question is by which cognitive processes a person assembles an appropriate evaluation function for a given problem for the first time. It would be surprising if there were no empirical regularities to be discovered with respect to such judgments. The theoretical principles needed to explain those regularities would be yet another component of a future theory of problem solving.

Summary

To engage in heuristic search, people have to perceive the problem situation, retrieve relevant actions, conceptualize the top goal, activate and apply action selection preferences, and assemble a way to evaluate problem states. Theories of these five cognitive functions might be forthcoming from psychological research on perception, memory, intentionality, decision making, and judgment, in which case there is nothing specific to say about them in the context of problem solving. If so, problem solving is not a natural kind in the study of cognition, and there might not be a theory of problem solving to be discovered. The reason to believe otherwise is that each of the five functions exhibits phenomena that are more salient or important in problem solving than in other contexts, or even unique to problem solving. Explanations for these phenomena are unlikely to be forthcoming from other areas of research, so they will have to be explained by problem solving researchers. The structure of a future theory of problem solving might therefore map onto the diagram in Figure 2. At the center is the core principle of heuristic search. Grouped around it are the five sets of yet-to-be-found principles that explain how people get to search. I conjecture that Newel and Simon would have agreed that such an expansion of their theory would significantly advance our understanding of how people solve unfamiliar problems.

Acknowledgments

The preparation of this paper was supported, in part, by award # N00014-1-09-1025 from the Office of Naval Research (ONR). No endorsement should be inferred.

References

Adamson, R. E. (1952). Functional fixedness as related to problem solving: A repetition of three experiments. *Journal of Experimental Psychology*, *44*(4), 288–291. http://dx.doi.org/10.1037/h0062487

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195324259.001.0001

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

Anzai, Y., & Simon, H. A. (1979). The theory of learning by doing. *Psychological Review*, *86*(2), 124–140. http://dx.doi.org/10.1037/0033-295X.86.2.124

Bhaskar, R., & Simon, H. A. (1977). Problem solving in semantically rich domains: An example from engineering thermodynamics. *Cognitive Science*, *1*(2), 193–215. http://dx.doi.org/10.1207/s15516709cog0102_3

Birch, H. G., & Rabinowitz, H. S. (1951). The negative effect of previous experience on productive thinking. *Journal of Experimental Psychology*, *41*(2), 121–125. http://dx.doi.org/10.1037/h0062635

Buchanan. B. G., & Feigenbaum, E. A. (1978). Dendral and meta-dendral: Their applications dimension. *Artificial Intelligence*, *11*(1–2), 5–24. http://dx.doi.org/10.1016/0004-3702(78)90010-3

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences*, vol 6, pp. 271–315. http://dx.doi.org/10.1207/s15327809jls0603_1

Craik, K. J. W. (1943/1967). *The nature of explanation* (Rev. ed.). Cambridge, UK: Cambridge University Press.

Conway, F., & Siegelman, J. (2006). *Dark hero of the information age: In search of Norbert Wiener, the father of cybernetics*. New York: Basic Books.

Duncker, K. (1935). *Zur Psychologie des produktiven Denkens [Psychology of Productive Thinking]* (Dritter Neudruck). Berlin, Germany: Springer Verlag. [English version: Duncker, K. (1945). On problem-solving. *Psychological Monographs*, vol. 58, Whole No. 270.]

Dyson, G. (2012). *Turing's cathedral: The origins of the digital universe*. New York: Pantheon.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Ernst, G. W., & Newell, A. (1969). *GPS: A case study in generality and problem solving*. New York: Academic Press.

Feigenbaum, E. A. (1989). What hath Simon wrought? In D. Klahr and K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 165–182). Hillsdale, NJ: Erlbaum.

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution*. New York: Basic Books.

German, T. P., & Barrett, H. C. (2005). Functional fixedness in a technologically sparse culture. *Psychological Science*, *16*(1), 1–5. http://dx.doi.org/10.1111/j.0956-7976.2005.00771.x

Ghiselin, B., (Ed.), (1952). *The creative process: A symposium*. New York: New American Library.

Gibson, J. J. (1977). The theory of affordances. In R. E. Shaw and J. Bransford (Ed.), *Perceiving, acting and knowing: Toward an ecological psychology* (pp. 67–82). Hillsdale, NJ: Erlbaum.

Gregg, L. W., & Simon, H. A. (1967). Process models and stochastic theories of simple concept formation. *Journal of Mathematical Psychology*, *4*(2), 246–276. http://dx.doi.org/10.1016/0022-2496(67)90052-1

Greeno, J. G. (1974). Hobbits and orcs: Acquisition of a sequential concept. *Cognitive Psychology*, *6*(2), 270–292. http://dx.doi.org/10.1016/0010-0285(74)90014-0

Groner, R., Groner, M., & Bischof, W. F., (Eds.), (1983). *Methods of heuristics*. Hillsdale, NJ: Erlbaum.

Hadamard, J. (1949/1954). *The psychology of invention in the mathematical field* (enlarged ed.). New York: Dover.

Henle, M. (1960). On errors in deductive reasoning. *Psychological Reports*, *7*, 80.

Henle, M. (1962). On the relation between logic and thinking. *Psychological Review*, *69*(4), 366–378. http://dx.doi.org/10.1037/h0042043

Holding, D. H. (1985). *The psychology of chess skill*. Hillsdale, NJ: Erlbaum.

Just, M. A., & Varma, S. (2007). The organization of thinking: What functional brain imaging reveals about the neuroarchitecture of complex cognition. *Cognitive, Affective & Behavioral Neuroscience*, *7*(3), 153–191. http://dx.doi.org/10.3758/CABN.7.3.153

Kershaw, T., & Ohlsson, S. (2004). Multiple causes of difficulty in insight: The case of the nine-dot problem. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *30*(1), pp. 3–13. http://dx.doi.org/10.1037/0278-7393.30.1.3

Kulkarni, D., & Simon, H. A. (1988). The processes of scientific discovery: The strategy of experimentation. *Cognitive Science*, *12*(2), 139–175. http://dx.doi.org/10.1207/s15516709cog1202_1

Köhler, W. (1924). *Die physischen Gestalten in Ruhe and im stationären Zustand: Eine naturphilosophische Untersuchung*. Erlangen, Germany: Verlag der Philosophischen Akademie.

Köhler, W. (1927). *The mentality of apes* (2nd ed.). New York: Harourt, Brace & Company.

Laird, J. E., & Newell, A. (1993). A universal weak method. In P. S. Rosenbloom, J. E. Laird, and A. Newell (Eds.), T*he Soar papers: Research on integrated intelligence* (Vol. 1, pp. 245–292). Cambridge, MA: MIT Press.

Langer, E. J. (1989). *Mindfulness*. New York: Addison-Wesley.

Langley, P., Choi, D., & Rogers, S. (2009). Acquisition of hierarchical reactive skills in a unified cognitive architecture. *Cognitive Systems Research*, *10*(4), 316–332. http://dx.doi.org/10.1016/j.cogsys.2008.07.003

Lenat, D. B. (1983). Toward a theory of heuristics. In R. Groner, M. Groner and W. F. Bischof (Eds.), *Methods of heuristics* (pp. 351–404). Hillsdale, NJ: Erlbaum.

Maltzman, I. (1955). Thinking: From a behavioristic point of view. *Psychological Review*, *62*(4), 275–286. http://dx.doi.org/10.1037/h0041818

Mandler, G. (2002). Origins of the cognitive (r)evolution. *Journal of the History of the Behavioral Sciences*, *38*(4), 339–353. http://dx.doi.org/10.1002/jhbs.10066

Mandler, G. (2007). *A history of modern experimental psychology: From James and Wundt to cognitive science*. Cambridge, MA: MIT Press.

Miller, G. A. (2003). The cognitive revolution: a historical perspective. *TRENDS in the Cognitive Sciences*, *7*(3), 141–144. http://dx.doi.org/10.1016/S1364-6613(03)00029-9

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. New York: Holt, Rinehart and Winston. http://dx.doi.org/10.1037/10039-000

Newell, A. (1972). A theoretical exploration of mechanisms for coding the stimulus. In A. W. Melton and E. Martin (Eds.), *Coding processes in human memory* (pp. 373–434). New York: Wiley.

Newell, A. (1973). Production systems: Models of control structures. In W. G. Chase (Ed.), *Visual information processing* (pp. 463–526). New York: Academic Press.

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

Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review*, *65*(3), 151–166. http://dx.doi.org/10.1037/h0048495

Newell, A., Shaw, J. C., & Simon, H. A. (1960). Report on a general problem-solving program. *Information Processing: Proceedings of the International Conference on Information Processing, UNESCO, Paris 16–20 June 1959* (pp. 256–264). Munich, Germany: UNESCO.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.

Nokes, T. J. (2009). Mechanisms of knowledge transfer. *Thinking & Reasoning*, *15*(1), 1–36. http://dx.doi.org/10.1080/13546780802490186

Ohlsson, S. (1980a). *Competence and strategy in reasoning with common spatial concepts* (Working Papers from the Cognitive Seminar No. 6). Stockholm, Sweden: Department of Psychology, University of Stockholm.

Ohlsson, S. (1980b). Strategy grammars: An approach to generality in computer simulations of human reasoning. *Proceedings of the AISB-80 Conference on Artificial Intelligence, Amsterdam, The Netherlands, July 1–4*.

Ohlsson, S. (1984a). Restructuring revisited I. Summary and critique of the Gestalt theory of problem solving. *Scandinavian Journal of Psychology*, *25*(1), 65–78. http://dx.doi.org/10.1111/j.1467-9450.1984.tb01001.x

Ohlsson, S. (1984b) Restructuring revisited II. An information processing theory of restructuring and insight. *Scandinavian Journal of Psychology, 25*(2), 117–129. http://dx.doi.org/10.1111/j.1467-9450.1984.tb01005.x

Ohlsson, S. (1984c). Induced strategy shifts in spatial reasoning. *Acta Psychologica*, *57*(1), 46–67. http://dx.doi.org/10.1016/0001-6918(84)90053-2

Ohlsson, S. (1990a). Trace analysis and spatial reasoning: An example of intensive cognitive diagnosis and its implications for testing. In N. Frederiksen, R. Glaser, A. Lesgold and M. G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 251–296). Hillsdale, NJ: Erlbaum.

Ohlsson, S. (1990b). The mechanism of restructuring in geometry. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society, Cambridge, Massachusetts, July 25–28* (pp. 237–244). Hillsdale, NJ: Erlbaum.

Ohlsson, S. (1992). Information-processing models of insight and related phenomena. In M. T. Keane and K. J. Gilhooly (Eds.), *Advances in the psychology of thinking* (vol. 1, pp. 1–44). New York: Harvester/Wheatsheaf.

Ohlsson, S. (2008). How is it possible to have a new idea? In D. Ventura, M. L. Maher and S. Colton (Eds.), *Creative intelligent systems: Papers from the AAAI Spring Symposium* (Technical Report SS-08-03, pp. 61–66). Menlo Park, CA: AAAI Press.

Ohlsson, S. (2011). *Deep learning: How the mind overrides experience*. New York: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511780295

Poincaré, H. (1908/1952). *Science and method* (F. Maitland, Trans.). New York: Dover.

Rapaport, D., (Ed.). (1951). *Organization and pathology of thought*. New York: Columbia University Press.

Runco, M. A., (Ed.). (1994). *Problem finding, problem solving, and creativity*. Norwood, NJ: Ablex.

Shortliffe, E. H., Axline, S. G., Buchanan, B. G., Merigan, T. C., & Cohen, S. N. (1973). An artificial intelligence program to advise physicians regarding antimicrobial therapy. *Computers in Biomedical Research*, 6, 544–560. http://dx.doi.org/10.1016/0010-4809(73)90029-3

Simon, H. A. (1978). Information-processing theory of human problem solving. In W, K. Estes (Ed.), *Handbook of learning and cognitive processes: Vol. 5. Human information processing* (pp. 271–295). Oxford, UK: Erlbaum.

Simon, H. A., Langley, P. W., & Bradshaw, G. L. (1981). Scientific discovery as problem solving. *Synthese, 47*(1), 1–27. http://dx.doi.org/10.1007/BF01064262

Simon, H. A., & Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist*, *26*(2), 145–159. http://dx.doi.org/10.1037/h0030806

Simon, H. A, (1992). What is an "explanation" of behavior? *Psychological Science*, *3*(3), 150–161. http://dx.doi.org/10.1111/j.1467-9280.1992.tb00017.x

Simon, H. A., & Reed, S. K. (1976). Modeling strategy shifts in a problem-solving task. *Cognitive Psychology*, *8*(1), 8–97. http://dx.doi.org/10.1016/0010-0285(76)90005-0

Smith, S. M. (1995). Getting into and out of mental ruts: A theory of fixation, incubation, and insight. In R. J. Sternberg and J. E. Davidson (Eds.), *The nature of insight* (pp. 229–251). Cambridge, MA: MIT Books.

Sun, J., Moore, J. L., Bobick, A., & Rehg, J. M. (2010). Learning visual object categories for robot affordance prediction. The International Journal of Robotics Research, *29*(2–3), 174–197. http://dx.doi.org/10.1177/0278364909356602

Sun, R. (2007). The importance of cognitive architectures: An analysis based on CLARION. *Journal of Experimental and Theoretical Artificial Intelligence*, *19*(2), 159–193. http://dx.doi.org/10.1080/09528130701191560

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. (I). *Psychological Review*, *8*(3), 247–261. http://dx.doi.org/10.1037/h0074898

VanLehn, K. (1991). Rule acquisition events in the discovery of problem-solving strategies. *Cognitive Science*, *15*(1), 1–47. http://dx.doi.org/10.1207/s15516709cog1501_1

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, *20*(2), 158–177. http://dx.doi.org/10.1037/h0074428