Symbols and Search in Humans and Machines

Pat Langley
Institute for the Study of Learning and Expertise
Palo Alto, California 94306 USA
Center for Design Research, Stanford University
Stanford, California 94305 USA

– *A physical symbol system has the necessary and sufficient means for general intelligent action.*

– *A physical symbol system exercises its intelligence in problem solving by search – that is, by selectively generating and progressively modifying symbol structures until it produces a solution structure.*

Allen Newell and Herbert A. Simon (1976)

Abstract

This chapter examines two major contributions made by Herbert Simon, with his collaborator Allen Newell, to cognitive psychology and artificial intelligence. The first is the theory of physical symbol systems, which introduced the notion of symbol structures and proposed computational mechanisms that manipulate them. The second is the theory of heuristic search, which characterizes problem solving as guided exploration through a space of symbol structures that includes solutions. In each case, I review the structures and processes that the theory postulates, along with hypotheses about the intelligent behavior it supports. In addition, I consider limitations of the frameworks and extensions that address them, as well as alternative accounts of the target phenomena. These profound theories offered deep insights about the computational character of intelligence and the roles played by symbols and search in human and machine cognition.

Introduction

Herbert Simon had major impacts on multiple fields and no single chapter can hope to cover his influence on them all. He is best known as the discoverer and champion of *satisficing* in human decision making, for which he received the 1978 Nobel Prize in Economics. However, Simon's work on this topic is well documented elsewhere in this volume, so here I will focus on two other major contributions – *physical symbol systems* and *heuristic search*. He developed and elaborated these ideas jointly with Allen Newell, so I will often refer to 'Newell and Simon' in the pages that follow, since their names usually appeared in that order. Nevertheless, given his many collaborations with others in both areas, there is no question that Simon was an equal partner in these two lines of research. The ideas covered here are described most clearly in an article (Newell & Simon, 1976) linked to their joint receipt of the 1975 Turing Award, but this piece reviewed and generalized results developed in the 20 years that preceded its publication.

Before discussing the details of these contributions, I should provide some historical context. The team's work on symbols and search had major influences on two fields that, in the late 1950s and 1960s, were seen as intimately connected. Newell and Simon are considered co-founders of artificial intelligence because they developed the first list processing language – IPL-V (Newell & Shaw, 1957) – and because they designed and implemented the first AI system – the Logic Theory Machine (Newell, Shaw, & Simon, 1957), But

they also played critical roles in the cognitive revolution within psychology (Miller, 2003), especially in the use of digital computers to model and simulate human thinking. These were not separate efforts; the two innovators viewed their early AI systems as formal accounts of complex human cognition and many others in both fields adopted this perspective (Langley, 2012). For this reason, I will often talk about their contributions to these disciplines in the same breath to reflect the synergistic spirit of their work.

A few additional comments are in order. One is that Newell and Simon described their core ideas as *hypotheses*, but this undercuts their breadth and import. They are better seen as theoretical postulates about the computational nature of mind in both man and machine. Another is that the principles they proposed in the 1950s have become so central to cognitive psychology, AI, and computer science that few in these fields know their origins. Innumerable researchers have built on these ideas in the past 65 years without realizing it, so their inventors now receive much less credit than they deserve. At the same time, some researchers interpret Newell and Simon's claims more narrowly than justified and so insist that their own frameworks oppose them. This is due partly to the fact that Newell and Simon, as limited information processors themselves, could only explore a small part of the vast territory they opened up. Thus, the systems they implemented relied on assumptions that suggested their ideas were more limited than was actually the case. Finally, neither of their claims about the mind were entirely new, as they built on ideas introduced earlier in the century. Still, Newell and Simon expanded on these results in crucial ways and made genuine breakthroughs, especially in demonstrating the ability of digital computers to reproduce key facets of intelligence.

I have divided the chapter into two main sections linked to the two opening quotations, one focused on physical symbol systems and the other on heuristic search, which elaborates on the first. Each section reviews the theoretical postulates that Newell and Simon proposed in their 1976 article, separating them into assumptions about structures and ones about processes that operate on them. The chapter also examines their explicit claims – what they termed 'hypotheses' – about the types of behavior these elements enable and produce. After this, I consider alternatives to these theoretical ideas, although we will see that these are seldom as antithetical to their core assumptions as usually advertised. In addition, I discuss some limitations of the theories, but I also argue that these only indicate incompleteness and suggest ways that the research community might extend them. The chapter closes by summarizing the main points and reiterating the profound character of Newell and Simon's vision for symbols and search.

Physical Symbol Systems

Humans are distinctive in their ability to carry our extended intellectual activities. We can follow a story's narrative, engage in political debates, diagnose and repair broken devices, generate goal-oriented plans, derive theorem from assumptions, and design novel artifacts. Although executing such endeavors often requires concrete physical actions, the mental processing that supports them is typically quite abstract. Some animals show glimmers of such reasoning abilities, but *Homo sapiens*' excellence in this arena has led to our species' unique accomplishments. These abilities are the central concern of both AI and cognitive psychology, and early theories from Simon and Newell offer some of the most compelling accounts. As noted above, these date to their groundbreaking work on the Logic Theory Machine (Newell et al., 1957), often viewed as the first implemented AI system, but Newell and Simon (1976) described the framework most clearly. This paper introduced a theory of symbol processing that they intended to cover not only digital computers but also human cognition. In this section, I review their framework's assumptions and their claims about its capabilities, along with alternatives and limitations.

Symbol Structures and Symbol Processing

Newell and Simon's theory revolves around the notion of a *physical symbol system*. We can divide this framework into statements about the *structures* that arise in such a system and ones about the *processes* that operate over them. The first of these includes:

- *Symbols*, which are physical patterns that remain stable unless modified some activity. These can take different forms and manifest in different media. Classic examples are letters or digits written on paper, drawn in the sand, scrawled on a blackboard, or encoded in human or computer memory. Each symbol type is distinguishable from other types, although instances may have minor variations. Symbols are usually composed of simpler physical entities, but only their *pattern* is important.
- *Symbol structures*, which are organized sets of symbols that, taken together, form *expressions*. There is usually a fixed set of primitive symbols, but they can be assembled into an unlimited number of composite structures. Classic examples are words made from letters, integers from digits, sentences from words, equations from variables and mathematical operators, and diagrams from lines and labels. Expressions are also physical entities that can arise in different media while codifying the same basic content.
- *Designations*, which are symbol structures that link to – and *denote* – other structures that reside in memory stores or that exist in the external world. Classic examples of designating structures are words that stand for classes of objects or events, maps that represent spatial relations among entities, and scores that describe musical tunes or compositions. These give symbol systems the ability to *represent* other items, either within themselves or in the environment. An important special case is a structure that encodes a mental procedure, that is, a *stored program*.

Symbols and symbol structures are often interpreted more narrowly than they deserve. Their definitions are agnostic about the substrate for their physical patterns, so the same structures can occur in very different media, whether stone tablets, papyrus scrolls, concrete graffiti, vacuum tubes, silicon chips, or neuronal connections. Neither are they limited to 'symbolic' statements in logic or rules; as I argue later, multilayer neural networks, sparse encodings, and other 'distributed representations' also fall within the paradigm. In fact, because symbol structures are composed of simpler elements, they always have a distributed character.

A physical symbol system also incorporates *processes* that inspect, manipulate, and otherwise operate over its symbol structures. According to Newell and Simon's account, these include:

- *Interpretation*, which involves 'running' or 'executing' an expression that designates a multistep procedure. This may be a purely internal activity, such as performing mental arithmetic, or an external one that affects the environment, like following a cooking recipe. This builds directly on the notion of a stored program, mentioned earlier, that is central to computer science.
- *Creation and modification* of symbol structures, which is needed to carry out these stored programs. These processes may involve computing and storing intermediate results, as done when solving an algebraic expression, or altering existing structures, as done when replacing the current value for a variable with a new one. Because symbol structures are composed of individual symbols, the system can create new structures by combing them in novel ways.
- *Extended operation*, which means that the symbol system acts over time to produce an evolving collection of symbol structures. This can support activities like generating multi-step plans to achieve goals, understanding the narrative of a story, creating designs that satisfy specifications, and learning new concepts or skills. Structures that arise later can build on ones introduced earlier to organize content in a hierarchical manner, although other relations, such as analogy, are also possible.

Again, it is important not to interpret the idea of symbol processing too narrowly. Generating a sequence of logical expressions is a classic example of this activity, but simulating a set of differential equations to produce a continuous trajectory is equally valid. Creation and modification of lists, list structures, and strings count as symbol processing, but so does the manipulation of numeric variables and matrices. At a higher level, chaining and acquisition of relational rules are important instances, but so are spread of activation and parameter estimation in neural networks.

I should also note that Newell and Simon (1976) did not stop with a definition of symbol processing; they also made an audacious claim about its relation to intelligence. In particular, they stated (p. 116) that:

- *A physical symbol system has the necessary and sufficient means for general intelligent action.*

Here the words 'intelligent action' refer to the types of mental activities listed earlier that make human cognition so distinctive. The modifier 'necessary' means that any agent exhibiting such behavior must be an instance of a physical symbol system, whereas 'sufficient' means that no additional structures or processes are required. I will argue that the first half of this claim has been generally supported by the past 60 years of research in AI and cognitive psychology, as these fields have found no instances of intelligence that fall outside the definition in its broad sense. The second claim's status is more debatable, in that considerable elaborations of the basic theory seem necessary before we can provide a full account of human-like intelligence. Newell and Simon (1976) emphasized that their hypothesis was not a mathematical truth but rather subject to empirical test, as reflected in their article's main title, 'Computer Science as Empirical Inquiry'.

The collaborators also acknowledged that their insights about symbol systems built on decades of earlier results. These included significant advances in:

- *Formal logic*, which provided a formal basis for proof and deduction, including mechanisms for syntactic manipulation of complex symbol structures.
- *Turing machines and digital computers*, which introduced general methods for sequential processing of information and instantiated them in physical devices.
- *Stored programs* for digital computers, which let the latter encode and interpret procedures as data, an ability that became central not only to AI but to computer science in general.
- *List processing* for digital computers, which supported notions of designation, data types, and dynamic memory in programming languages like IPL-V (Newell & Shaw, 1957) and Lisp (McCarthy, 1960).

Newell and Simon argued that interpretation came jointly from the second and third items, whereas designation came from the final one. The theory of physical symbol systems, and its associated hypothesis, combined and generalized these intellectual threads into a single, unified framework. This breakthrough made it possible to create the first AI systems and, in parallel, supported the cognitive revolution in psychology.

Alternatives to Physical Symbol Systems

Now consider some apparent alternatives to the physical symbol system hypothesis. We should start with the classic adversary of cognitive theories, *behaviorism*, which dominated American psychology for four decades. As Miller (2003) notes, the cognitive revolution of the 1950s was largely a response to what many saw as limitations of this paradigm. Behaviorism postulated that all behavior, in both animals and humans, results from stimulus-response pairs that directly connect an organism's perceptions with its actions. Such associations are learned through experience in reaction to experienced reward, which can form chains of stimulus-response pairs to produce complex sequential behavior. The physical symbol system framework

allows such connections, but Newell and Simon believed that intelligent agents like humans also have internal mental structures, whereas mainstream behaviorist accounts explicitly forbid them. Later variations relaxed this restriction and included the possibility of internal stimulus-response bonds, making them similar to *production systems* (Newell, 1973; Neches, Langley, & Klahr, 1987), which are viewed as classic examples of symbol systems. A related movement, *situated cognition*, also posited that intelligent behavior relies heavily on interaction with the external environment. Advocates like Suchman (1987) cast this view as in opposition to Newell and Simon's ideas, but Vera and Simon (1993) argued that they are largely compatible. Indeed, symbol systems manifest not only in an agent's mind but in its environmental surroundings (Newell & Simon, 1972, pp. 800–803). Thus, behaviorism and related paradigms are not true competitors to physical symbol systems, even in their extreme forms, but instead are constrained instances of them.[1]

A second group of competitors acknowledges the existence of mental states but considers them to be points in a continuous space rather than as discrete symbolic structures. A prime example is *dynamic systems theory* (e.g., Spivey, 2007), which views cognitive processing as tracing a trajectory through this space over time, with each step determined by the current mental state and external stimuli. This framework often focuses on theories of situated and embodied cognition, but it acknowledges a core role for internal mental states, while also claiming that they are continuous rather than discrete in character. Dynamic systems accounts differ from other examples of physical symbol systems in their emphasis on continuous encodings, but they also hold much in common. Both characterize cognition as extended operation that produces a sequence of mental states. Moreover, the differential equations hypothesized by dynamic systems advocates as responsible for generating successor states are similar in spirit to the production rules adopted by many discrete theories of cognition. This mapping becomes more obvious when one realizes that equations are themselves parameterized symbol structures. The dynamic systems movement harks back to a much older movement, related to Gestalt psychology (Koffka, 1935; Kohler, 1940), that treats mental states as a *field* rather than as a point in continuous space. In this view, mental operations can sometimes lead to reorganization of the field, an idea to which I will return later.

A third alternative instead postulates that cognitive structures are *iconic* in character, involving a form of mental images. For example, Shepard and Metzler (1971) hypothesized mental rotation of such internal representations to explain results on comparison of objects in different orientations, while Kosslyn (1980) proposed a matrix of grid cells that supported various processes related to mental imagery. Similar data structures are commonly used to encode cognitive maps in robotics research (e.g., Yamauchi, Schultz, & Adams, 1998). Although Newell and Simon did not discuss iconic theories and other papers (e.g., Simon, 1978) suggest that they did not intend them as examples of their framework, we can nevertheless view them as a specialized form of symbol structure that stores content and changes over time. Neither is the framework antithetical to classic symbolic paradigms, as shown by an extension to the Soar architecture (Laird, 2012) that includes linked memories for list structures and for spatial maps. These different representations, and their associated processes, each have strengths that offset the other's weaknesses.

A final theory, currently very much in vogue, claims that long-term structures are encoded in a *neural network* consisting of a set of nodes and weighted links that connect them (Hinton & Anderson, 1981). Processing involves the spread of numeric activations from input nodes to outputs, passing through intermediate levels along the way. Such 'connectionist' accounts emphasize distributed representations and parallel computation, which are often contrasted with 'symbolic' approaches that rely on logic-like encodings and

[1] Research in the AI paradigm of reinforcement learning (Sutton & Barto, 1998) is a computational descendent of behaviorism with links to production systems that allows for mental states but that also favors close ties to the environment.

sequential processing. Connectionist models are described as 'subsymbolic' because their components do not map onto familiar words or concepts, but, taken together, their nodes, links, and weights certainly satisfy the definition of symbol structures. Newell and Simon (1976) did not mention such systems, and Richman and Simon (1989) treat them as competitors, but this does not invalidate the argument to include them. The distinction is blurred further by recurrent neural networks (e.g., Elman, 1990), which operate in discrete cycles, with outputs from one round becoming inputs for the next, so that node activations serve as a dynamic short-term memory akin to that in production systems. Recent work has even introduced mechanisms that focus attention on some elements over others (e.g., Seo et al., 2017). There remain important differences between the paradigms, but both count as instances of symbol systems.

In summary, the theory of physical symbol systems – as usually interpreted – has alternatives, but closer inspection suggests that these frameworks are not actually incompatible. Despite common claims to the contrary, they are not true adversaries to Newell and Simon's theory. A more accurate statement would be that each one is a special case of physical symbol systems that imposes constraints which are open to debate. This suggests that Newell and Simon's key contribution was moving beyond such specialized accounts to offer a broader view that includes logic, rules, grammars, and other symbol structures not supported by the narrower theories. They claimed that digital computers – and the human mind – are not mere stimulus-response machines, number crunchers, or image manipulators. Instead, they are *general symbol processors*, which is what underlies their distinctive ability to exhibit flexible, intelligent behavior.

Elaborations on Physical Symbol Systems

We should also examine some apparent limits of the symbol systems paradigm. The history of science includes many cases in which an initial theory was extended and elaborated over time in response to new phenomena and new insights. For instance, Dalton's atomic theory focused originally on the number of atoms for each element in given types of molecules, but others revised his specific models (e.g., for water) as new data became available. Decades later, the community augmented the framework to incorporate molecular structure when confronted with results about organic chemicals like benzene. Similarly, Pasteur's germ theory of disease dealt initially with associations between microorganisms and pathologies, but it was afterward elaborated to include immune responses and their acquisition. In the same way, it is natural to consider how one might extend the theory of physical symbol systems as the basis for intelligent behavior.

Further analysis of human cognition suggests additional constraints that we can place on the structures and processes that underlie intelligence without reducing generality. These stronger theoretical assumptions about the infrastructure of the mind are often collectively referred to as the *cognitive architecture* (Newell, 1990; Anderson, 2007). Research on this topic typically adopts five postulates:

- *Memories are collections of distinct, modular elements that are encoded as discrete symbol structures.*
- *Long-term memories store stable elements that change slowly through learning, whereas working memories contain dynamic elements that change rapidly during performance.*
- *Elements in long-term memories are accessed by matching their structures against the contents of elements contained in working memories.*
- *Cognitive processing occurs in cycles that match long-term structures, select a subset to apply, and execute them to update working memories or the environment.*
- *Cognition dynamically composes mental structures: performance processes produce new short-term elements and learning creates or alters long-term elements.*

This paradigm has led to many candidate architectures that share these high-level assumptions but differ in their details (Langley, Laird, & Rogers, 2009; Langley, 2017a). The most common subclass, also proposed first by Newell (1966, 1973), is known as *production systems*. These encode long-term knowledge as a set of condition-action rules that match against and alter working-memory elements upon application. Simon used production systems in later modeling efforts and he supported the cognitive architecture movement, but he never focused on them to the same extent as Newell.

On another front, we have seen that a number of alternative frameworks emphasize the notion that human cognition is *embodied*, in that internal representations are grounded in perception and action. However, it is more accurate to treat this as a *complementary* theory, as Newell and Simon's account was agnostic on the matter. Their treatment certainly did not exclude the idea that cognition often involved directly interaction with the world, as documented by their work on puzzles and chess (Newell & Simon, 1972, Simon, 1975). This suggests another elaboration of the theory, which I will call the *symbolic physical system* hypothesis:

- *The elements in a physical symbol system denote structures or processes that occur in the external world and interpreting them involves simulating them mentally or executing them physically.*

This statement imposes stronger constraints than the original theory, as it only allows mental structures that map onto external configurations or activities. These need not be concrete or fully instantiated; grounding does not rule out storage and manipulation of abstractions or generalizations. The extended theory forbids mental structures and processes that lack any environment connections, but it does *not* claim that all cognition involves direct external activity, so it is not limiting in the same sense as behaviorism. Intelligence still relies heavily on internal content, but this is always linked to perceived, inferred, or imagined entities and activities in the world. This assumption channels the framework to explain abstract activities like logic and mathematics by connecting them to manipulation of external symbols or to mental models of them.

An important corollary addresses the content of environmental descriptions. These include the size, shape, and organization of physical objects, as well as the spatial regions in which they occur. They also describe dynamic processes and actions that transform these situations over time, along with the events that characterize these changes. Again, these can range from generic statements for abstract categories and procedures to detailed specifications for concrete entities and activities. In other words, every cognitive structure and process maps onto situations or transformations that could occur in physical environments, which in turn involve distinct objects with spatial relations that change over time. Research on qualitative physics (Faltings & Struss, 1992) has produced formalisms for specifying such mental models and interpreters for simulating them. This paradigm has focused on the qualitative behavior of dynamical systems, but there has also been work on qualitative representation and reasoning about space (Cohn & Hazarika, 2001). These elaborations of the symbol system framework clarify further how it can support embodied cognition.[2]

Another area in which Newell and Simon's vision for symbol systems arguably falls short is its emphasis on isolated agents. Embodiment addresses connections between internal cognition and the environment, but it still ignores the social character of much human experience. Their theory allows for symbol systems that interact with others, as demonstrated by many artifacts implemented within the framework, but it does not call out this important capacity. Langley (2017b) proposes another claim that remedies this omission:

- *Intelligence depends on the ability to represent models of other agents' mental states, generate and reason over such models, and use them for informed interaction.*

[2] Of course, they do not specify detailed mechanisms for perception and action that let cognition interface with the environment, which remain major scientific challenges, but they place some constraints on these processes.

This conjecture builds directly on symbol systems, as it refers to that framework's structures and processes, but it goes further to suggest that full intelligence *requires* the ability to designate and interpret models of others' mental states. Like the symbolic physical system hypothesis, this conjecture introduces no new types of symbol structures, but it takes a strong position about the content they encode. This *social cognition* hypothesis has been central to AI research on dialogue (e.g., Allen et al., 1995; Mcshane & Nirenburg, 2012) and multi-agent coordination (Huhns & Singh, 1998), but it has not received the same broad support as Newell and Simon's original claim about symbol systems.

Problem Solving as Heuristic Search

Humans have another, more specific, distinctive ability: they can solve nontrivial problems that they have never encountered. We can decipher and unravel challenging puzzles, generate multistep plans to achieve our goals, design new artifacts that serve a desired function, play games against unfamiliar opponents, and find ways to make broken devices work again. This ability is impressive not only in that it can address quite complex tasks, but also in its breadth and generality. A few animals show rudimentary capacities to solve novel problems, but they never approach the richness or flexibility seen in our species. For this reason, problem solving was a major topic during the AI and cognitive psychology revolutions of the 1950s, and the Newell-Simon collaboration made key breakthroughs in this area. Their paper on the Logic Theory Machine (Newell et al., 1957) reported the first computational artifact that exhibited multistep problem solving. This system focused on proving theorems in the propositional calculus, but the approach had much broader potential, as shown in their General Problem Solver (Newell, Shaw, & Simon, 1960). Both efforts revolved around what their creators called *heuristic search*. In this section, I review discussion of this framework in their 1976 article, along with variations on the idea and limitations that require further attention.

Problem Spaces and Heuristic Search

The central concept in Newell and Simon's second theory is the *problem space*, which specifies a set of possible situations in which solutions to a problem may be found. In some settings, like chess, this space can be incredibly large, which means that one cannot store its elements in advance or simply enumerate them exhaustively. Thus, typically a problem space is not itself a symbol structure and the only practical approach is to define it in terms of more basic elements. This leads naturally to postulates about the structures that underlie problem solving, which include:

- *Situations*, which are structures that may be problem solutions or that fall along paths to solutions. Examples include locations in a travel plan, assignments to cells in a Soduko puzzle, and layouts for a circuit design. Newell and Simon (1972, p. 810) referred to them as *states*, presumably because they sometimes map onto physical situations (e.g., places along a route); others call them *nodes* in the problem space.
- *Tests*, which are symbol structures that let one determine whether a situation is acceptable as a solution. Examples include goal descriptions for planning problems, specifications for design tasks, and thresholds for numeric evaluation criteria. These are often stated in the same or similar formalism as situations, which simplifies their comparison, but they may instead be opaque processes.
- *Generators*, which are structures that indicate how to produce new situations from existing ones. Examples here include the conditional effects of actions for use in planning, components available for design tasks, and even sets of letters for crossword puzzles. A more common term is *operators* (Newell & Simon, 1972, p. 810). Like tests, generators are often described in a notation similar to that for situations.

- *Heuristics*, which are symbol structures that characterize the desirability or quality of situations or the generator instances that produce them. These may take the form of rules or constraints that specify whether an option is allowable, rejectable, or preferable. Alternatively, they may be stated as mathematical functions that return numeric scores reflecting a candidate situation's quality.

One can represent each of these constituents in different ways, but they are invariably symbol structures, making the heuristic search theory an elaboration of the symbol system hypothesis. The idea of a problem space has figured prominently in studies of human problem solving and in multiple subfields of AI, where it has led to both conceptual advances and practical applications.

Newell and Simon's theory of problem solving also postulates a set of interacting processes. These include three mechanisms that inspect and manipulate the structures just discussed:

- *Generating new situations* from existing ones, which explores regions of the problem space in search of solutions. This process applies one of the generators to one or more known situations to produce a new candidate. Examples include adding an action to a plan, filling a cell in a Sudoku puzzle, and modifying an algebraic expression. This activity gradually makes explicit a portion of the implicit problem space.
- *Testing situations* to determine whether they constitute acceptable solutions. This mechanism applies the test criteria to structures produced by generation to tell the problem solver whether it can terminate with success. For instance, in planning one checks whether a situation satisfies the target goals, in Sudoku whether all cells are filled and all constraints met, and in algebra whether the variable is isolated.
- *Using heuristics* to guide generation and selection of situations to focus search. This may involve applying discrete rules or constraints to reject undesirable structures or invoking numeric functions to evaluate alternatives. The problem solver may use heuristics to evaluate either candidate solutions (e.g., situations in planning) or generator instances (e.g., actions that produce them).

Together, these three mechanisms work together to produce *heuristic search*. Although not guaranteed, this often leads to focused exploration of the problem space, which can be very large, while generating and testing only a small fraction of situations implicit in its specification. This ability to guide search in promising directions often makes it far more effective at finding solutions than nonheuristic, 'blind' methods. Search can occur without heuristics, but relying on them can make it much more tractable and, in some cases, they focus attention so effectively that behavior mimics that of a specialized algorithm.

As before, Newell and Simon (1976) did more than simply define the structures and processes they postulated were involved in this important facet of intelligence. They also made an explicit claim (p. 120) about its abilities that they called the *heuristic search* hypothesis:

- *A physical symbol system exercises its intelligence in problem solving by search – that is, by selectively generating and progressively modifying symbol structures until it produces a solution structure*.[3]

Again, they intended this claim to hold for problem solving in both humans and machines, although they did not state this as overtly as with their previous one. In the 65 years since its introduction, the theory of heuristic search has been used to explain human cognition in theorem proving, puzzle solving, plan generation, game playing, conceptual design, and many other settings. The framework has also been so widely adopted that few researchers recall its origins, but they have explored elaborations and variations of the core idea, to which I will turn shortly. There is no question that the theory has been immensely successful and that there is overwhelming evidence for its key role in human and machine intelligence.

[3]Their statement also included a sentence about solutions being symbol structures, which I have omitted here because it is definitional. For some reason, they did not mention heuristics except in the hypothesis' name.

Of course, Newell and Simon did not pull their theory of heuristic search from thin air; they precipitated ideas that had been in the intellectual atmosphere for some time, even if they were not fully conscious of them. I have noted its strong connection to symbol systems, so it shares the precursors of that framework, but it also mirrors insights from earlier studies of problem solving. In particular, Selz (1927) presented one of the first accounts of this ability in humans, stating clearly that it involves a stepwise process of moving from an initial situation to a goal state. As Simon (1981) has noted, Selz also outlined key aspects of means-ends analysis, although he appears to have downplayed the role of search, which he denigrated as trial and error. However, the importance of guided search is clear in later analyses of chess playing by De Groot (1965), one of Selz's students. Duncker (1945) also incorporated his predecessor's ideas about the stepwise, directed character of problem solving and championed the use of heuristic methods to guide search.[4] Of course, these early researchers did not formalize the accounts of human cognition on digital computers, which did not yet exist, but their core ideas prefigured those in Newell and Simon's theory of heuristic search.

As Langley (2017c) has noted, the literature has used *heuristic* in three distinct ways. One sense refers to a strategy for organization of decision making or problem solving. Examples of such methods include the *take-the-best* heuristic for choice tasks (Gigerenzer & Goldstein, 1996) and techniques like hill climbing and means-ends analysis for problem solving. The latter are usually domain independent and quite general; Newell (1969) has called them *weak methods* to contrast them with more powerful approaches with narrower application like linear programming. This is the sense of 'heuristic' that Polya (1945) used is his analysis of mathematical thinking. Another meaning refers to a domain-specific rule or constraint for recognizing an option that one should select, reject, or prefer over alternatives. A final sense, related to the second, is a numeric function, typically specific to a domain, used to evaluate situations or actions. The last two have become dominant in the AI literature, almost to the exclusion of the first meaning, although Newell and Simon appeared to include them all. I should also note that heuristics, as originally presented, offered no guarantees for finding optimal solutions or, indeed, any solution at all, but they often made difficult tasks tractable. Unfortunately, many AI researchers now focus on so-called 'admissable heuristics' that, under certain conditions, produce an optimal solution, which is a subversion of the original concept.

Variations on Heuristic Search

As with physical symbol systems, the heuristic search theory is incomplete, but researchers have developed many extensions and variations. One of the earliest elaborations, *means-ends analysis*, came from the same team (Newell et al., 1960). This posits that problem solving involves selecting some operator O to reduce differences between a current state S and a goal description G. This in turn generates two subproblems, one to transform S into another state that satisfies the conditions of O and another to transform the state produced by applying O into one that satisfies G. Means-ends search has been implicated repeatedly in studies of human problem solving, but it does not occur in all situations and the original method has difficulty with some planning tasks, which led some AI researchers to reject it. However, there exist more flexible variants (e.g., Langley, Barley, & Meadows, 2018) that use different criteria for selecting operators to overcome this drawback and that can produce not only backward chaining from goals but also forward chaining from states. Means-ends analysis offers an account of novice problem solving, but a related approach, *problem-reduction search*, offers a way to incorporate domain knowledge to further guide the exploration process.

[4] Selz and De Groot were in the 'structuralist' school, whereas Duncker was a Gestaltist, but they agreed on many core points.

A more distinctive alternative, *analogical problem solving*, uses a different way to draw on knowledge to generate candidate solutions (e.g., Veloso & Carbonell, 1993). Classic work on heuristic search starts from an 'empty' structure and adds elements, such as steps in a plan, one at a time. In contrast, analogical approaches retrieve from long-term memory a complete solution and adapt its elements as needed to the new task. The adaptation process may require removing components that are not relevant, adding new ones not present in the retrieved solution, or both. This still involves a form of search, but the starting point is quite different and processing has a different character than creation of a solution from scratch. In particular, it relies on heuristics for retrieving one or more candidate solutions from memory and on ones for deciding which elements to remove and which to add. In summary, analogical problem solving arguably falls within the heuristic search framework, but Newell and Simon did not emphasize it when they formulated their theory. An intermediate approach, *analogical search control*, comes closer to their vision in that it uses retrieved decisions to guide problem solving from scratch (e.g., Jones & Langley, 2005).

Another distinction concerns the class of tasks addressed by the problem-solving process. Newell and Simon's studies focused on *achievement* tasks in which it can be difficult to find any solution that meets the specified goals. Examples of this category include solving classic puzzles and proving mathematical theorems. However, there exists another broad class of tasks in which there are many solutions, some of which are more desirable than others, with examples like job scheduling and route planning. These are often referred to as *optimization* problems and they can be solved used very different methods. One common technique is to search a space in which each candidate is a complete solution (e.g., a schedule with all jobs assigned) and use operators that modify this structure in some way. As in analogy, problem solving starts a fully specified solution, although this might be generated randomly rather than retrieved from memory. Also, rather than adding or removing steps, search typically involves replacing one choice with another and uses numeric functions as heuristics that guide search toward nodes with higher scores. Again, solving optimization tasks fall within the framework of heuristic search, but Newell and Simon did not emphasize them in their theory. Finally, note that optimization focuses on the quality of solutions, but it does *not* require that problem solving find an optimal one, which is an essential distinction.

Despite their differences, the frameworks above all search through a space of symbol structures and thus build on the physical symbol system hypothesis. A more distinctive alternative replaces these discrete structures with points in a continuous vector space, in which each dimension is some parameter of interest. Problem solving in this paradigm still involves an iterative process of generating new situations that are evaluated for desirability, but the details are quite different. As in some approaches to discrete optimization, search begins from a fully specified situation that is often produced randomly. Rather than generate discrete alternatives and selecting among them, a *gradient descent* mechanism uses a numeric function to compute a change to the current situation to find a new situation that scores better on some criterion. This stepwise process continues until no further improvement occurs, so it is naturally viewed as a form of optimization. The trajectory may end in a local optimum, so it is common to run gradient descent repeatedly from different randomly selected starting points, although even this does not guarantee global optimality.

As with symbol systems, each of these search methods are special cases of Newell and Simon's more general characterization. Their formulation also covered older techniques, like depth-first search, which explores an arbitrary problem space exhaustively, and linear programming, which finds optimal solutions to highly constrained problems very efficiently. They were also aware of minimax evaluations based on lookahead search in game playing. The key innovation was not search itself but the inclusion of heuristic methods that, although offering no guarantees, often work well in practice and that match behavior observed

in human problem solving. Their new theory explained how people, despite their limited resources, can solve challenging tasks and how to devise AI systems with the same ability. Heuristic search lets the problem solver, whether human or machine, navigate between the Scylla of exhaustive search techniques, which are general but intractable, and the Charybdis of specialized methods, which are efficient but narrow.

Limitations of Heuristic Search

The theory of heuristic search had major, lasting impacts on both AI and cognitive psychology, providing insights not only into areas like planning and game playing, but even into creative behavior like design (Simon, 1969) and scientific discovery (Langley, Simon, Bradshaw, & Żytkow, 1987). However, these successes do not imply that it had no limitations, many of which Ohlsson (2012) has discussed in an excellent review of research on problem solving. As before, such drawbacks do not indicate that the framework is incorrect as much as they show incompleteness, in which case it is appropriate to elaborate it in ways that address them. Ohlsson's analysis focused on the origin and transformation of problem representations and search strategies. I will recount his main ideas but organize their presentation somewhat differently.

The first issue deals with generating an initial problem space for a given task. According to Newell and Simon's account, this means selecting a way to represent the initial and other situations, encode the generators, and specify the tests or termination criteria. A person can draw on different sources of information to formulate a problem space. In some situations, he reads or hears instructions about a task, using language-understanding abilities to construct a formulation. Hayes and Simon (1974) reported an initial computational model of this process, but its abilities were limited and we need more work in this area. In other cases, a person is shown a diagram or physical display, which he must translate into an internal encoding of situations and operators. Technologies for image processing and sketch understanding are relevant here, but there have been few links to problem-space construction. Recent work on 'interactive task learning' has combined these two sources of content and produced promising results on heuristic search in game playing (e.g., Hinrichs & Forbus, 2014; Kirk & Laird, 2014). A third option involves manipulating physical objects to acquire abstract operators from experience. Research on learning action models (e.g., Wang, 1995) is relevant here, but it has seldom been applied to constructing problem spaces. In summary, there are promising avenues for extending the heuristic search theory in this direction, but it merits more attention than given to date.

A second drawback concerns when to create and eliminate problems. People identify and pursue new top-level tasks; they also decide when problems have been solved to their satisfaction and when to abandon them. Classic work on problem solving offers ways to generate subproblems (e.g., Newell et al., 1960), but introducing top-level tasks is another matter. Research on 'goal reasoning' (Aha, Cox, & Muñoz-Avila, 2013) has addressed this issue, often by positing generic rules or 'motives' that produce concrete goals under specified conditions. These shift the focus of an agent's attention to new problems that drive cognitive and physical behavior in different directions. The second topic, when to declare success, is a distinct issue that relates to Simon's (1956) theory of *satisficing*, which states that people halt when they encounter an option that is 'good enough'. Classic work on problem solving assumes that goal descriptions are matched in an all-or-none manner, but many settings diverge from this assumption. Using numeric evaluation functions to guide search lends itself to other termination schemes, such as halting when the value meets some threshold, which maps onto Simon's notion of an *aspiration level* in satisficing. The concept has typically been applied to simple choice tasks, in which the decision maker selects one candidate from a set, but it is equally relevant to problem solving, where it can serve as a criterion for success. The literature includes only a few studies of

this topic (e.g., Gajderowicz et al., 2018), so this is another area that needs more research.[5] Simple accounts treat aspiration level as constant, but it can clearly vary over time, as when repeated failures lead a person to revise expectations downward or successes upward. Repeated failures can also cause the problem solver to abandon a task when he decides that expending more effort is unlikely to produce results.

A third limitation involves the inability of heuristic search, at least on its own, to explain *insight*, in which someone repeatedly fails to solve a problem until they suddenly and unexpectedly 'see' the solution that evaded them before. Laboratory examples include the nine-dot puzzle and matchstick problems, which people find difficult to solve without hints, but insight has also been linked repeatedly to creativity in science and design. This phenomenon was a central concern of Gestalt theorists (Koffka, 1935; Kohler, 1940; Wertheimer, 1959), who posited that it involves *restructuring* the problem so that, perceived in a new way, the answer is obvious. Ohlsson (1984a) provides an excellent reconstruction of Gestalt theory and how it differs from Newell and Simon's framework. They share the assumption that unsolved problems have 'gaps' that must be filled, but they propose distinct mechanisms. In Gestalt restructuring, a problem structure is subject to 'forces' that, when unbalanced, draw it inevitably toward a new structure that is in a dynamic equilibrium, much like a physical system. In contrast, Newell and Simon allowed for multiple paths in a problem space that heuristic search must explore. A few efforts have extended the latter framework to explain insight effects. Simon (1966) suggested that extensive search followed by selective forgetting could reveal a previously obscured solution, whereas Jones and Langley (2005) reported an account that combined means-ends analysis with spreading activation retrieval. Ohlsson (1984b) proposed search through an augmented problem space that included operators for altering representations. In the AI literature, Amarel (1968) noted the influence of representations on problem solving and, more recently, Riddle, Barley, and Franco (2013) have analyzed meta-level operators that revise them. However, work on this topic has been limited and it deserves much more attention from the research community.

Symbols, Search, and Discovery

As we have seen, the symbols and search theories are very general, and Herbert Simon applied them to many research problems, but one of them – *scientific discovery* – merits special attention. There is little question that science is one of the pinnacles of human achievement and thus worthy of substantial attention and study. Yet philosophers of science, and many others, had long attached a mystical aura to the act of discovery, claiming that it requires a 'vital spark' not subject to rational analysis. However, Simon (1966) felt otherwise and claimed that discovery is a variety of problem solving and thus relies on heuristic search through a space of symbol structures.

Moreover, to make this claim operational, he adopted the same strategy that had proved so effective in other, more mundane, areas of cognition. That is, Simon and his collaborators developed computational models of scientific discovery processes and demonstrated their behavior on relevant tasks. To this end, they found ways to encode candidate laws and models as symbol structures and they devised heuristic methods to guide search through a space of such hypotheses. The resulting systems served two complementary functions: they offered plausible accounts of human scientists' activities and they revealed how one could automate scientific discovery on digital computers.

[5] Work on *partial satisfaction planning* (Benton, Do, & Kambhampati, 2009), which finds solutions that achieve a subset of target goals, is related but typically insists on optimality, which runs directly counter to Simon's definition.

Early efforts along these lines focused on the discovery of numeric laws from quantitative data. The best-known examples involved the Bacon family of systems (e.g., Simon, Langley, & Bradshaw, 1981; Bradshaw, Langley, & Simon, 1983), which found relations like Kepler's third law of planetary motion and Black's law of specific heat. These operated in a data-driven manner to induce descriptive summaries for their observations, introducing new theoretical terms in search of ones with constant values. However, the researchers also developed other systems that did more than summarize data; they searched a space of models that could explain phenomena in terms of unobserved structures and processes (e.g., Żytkow & Simon, 1986; Kulkarni & Simon, 1988).

Initial demonstrations took examples from the history of science, but later results showed that heuristic search through a space of laws and models could produce novel discoveries deemed worthy of publication in the scientific literature (e.g., Valdés-Pérez, 1994; Langley, 2000). Other researchers outside Simon's immediate circle built on these efforts to create more sophisticated systems and generate additional results. This line of research has been so successful that few people today would argue with the claim – once audacious – that discovery is problem solving writ large. Thus, it counts as another success story for symbols and search, this one especially impressive because so many had once claimed it was impossible.

Concluding Remarks

In this chapter, I examined two major contributions made by Herbert Simon and his collaborator, Allen Newell. They arrived at these insights initially in the 1950s, when developing the first AI system, which they also presented as a computational account of human problem solving. Along with others, they elaborated on these ideas in the decades that followed, creating additional systems that reproduced many distinctive features of human cognition and that offered compelling explanations for many observed phenomena. The early results by Newell and Simon, which figured prominently in both the founding of artificial intelligence and the cognitive revolution in psychology, have come to serve as reliable foundations for computational understanding of the mind. This makes them eminently worthy of retrospective analysis.

As we have seen, the first contribution was the theory of physical symbol systems. This postulated the existence of symbols, which are stable physical patterns, symbol structures, which are organized sets of symbols, and designations, which are symbol structures that denote other structures. The framework also assumed processes for interpretation of symbol structures, creation and modification of such structures, and extended operation over time. Taken together, these elements make up a physical symbol system, which Newell and Simon hypothesized provide the necessary and sufficient conditions for general intelligent action. The theory drew on earlier ideas about formal logic, Turing machines and digital computers, stored programs, and list processing, but it brought them together in a unified theory with broad implications. I also examined classic alternatives to symbol system accounts of intelligence, including stimulus-response pairs in behaviorism, differential equations in dynamic systems theory, and iconic processing of mental images. Each is actually a special case of the symbol systems account, which moves beyond them to claim that humans and digital computers are general symbol manipulators. Finally, I argued that apparent limitations of the framework are rather signs of needed elaborations, which research on cognitive architectures, embodied intelligence, and social cognition have – to a certain degree – addressed.

The second contribution that I reviewed was the heuristic search theory of problem solving. This relied on the notion of a problem space that is specified implicitly in terms of situations, which are possible problem solutions, test criteria, which indicate whether a candidate solves the problem, generators, which

describe how to produce new situations from existing ones, and heuristics, which characterize the desirability of situations or generators. The framework also included mechanisms for exploring the problem space by creating new situations, testing them to determine if they are solutions, and using heuristics to guide the search process. Moreover, each of these elements takes the form of symbol structures. They work together to support heuristic search, which enables problem solving by selectively generating and modifying situations until it finds a solution. Their theory incorporated ideas from early psychologists but combined them with insights about symbol systems, which led to embedding them in running AI systems. I also considered variations on heuristic search, including means-ends analysis and analogy, as well as discrete and continuous optimization. These count as special cases of heuristic search, which offers a more general theory that supports problem solving over the discrete structures implicated in planning, game playing, and other settings. Finally, I discussed limitations of the heuristic search paradigm, borrowing from Ohlsson (2012), such as the inability to create an initial problem space, to generate and eliminate top-level goals, and to explain insight problem solving, although researchers have made some progress on these fronts.

Given the overwhelming adoption of these two theories, and the decades that have passed since their introduction, it is natural than many have forgotten their origin and taken them for granted. The limited character of human cognition makes such dismissal understandable, but that makes it no less an error. The computational frameworks put forward by Herbert Simon and Allen Newell in the 1950s – physical symbol systems and heuristic search – were profound contributions that have enabled both sophisticated AI applications and detailed models of human thinking. They can be viewed as groundbreaking as Newton's gravitational theory and Dalton's atomic theory were in their times, and they laid the foundation for equally productive scientific disciplines. These complementary theories have given us deep insights about the computational nature of minds and the central roles of symbols and search in humans and machines.

Acknowledgements

This chapter was supported by Grant N00014-20-1-2643 from the Office of Naval Research and by Grant FA9550-20-1-0130 from the Air Force Office of Scientific Research, which are not responsible for its contents. I thank Jan Auernhammer, John Laird, Stellan Ohlsson, and Paul Rosenbloom for insights about historical influences on Simon's thought and for constructive comments on an earlier draft.

References

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York: Oxford University Press.

Aha, D. W., Cox, M. T., & Muñoz-Avila, H. (2013). (Eds.). *Goal reasoning: Papers from the ACS workshop*. Baltimore, MD.

Allen, J. F., Schubert, L. K., Ferguson, G., Heeman, P., et al. (1995). The TRAINS project: A case study in building a conversational planning agent. *Journal of Experimental and Theoretical Artificial Intelligence*, *7*, 7–48.

Amarel, S. (1968). On representations of problems of reasoning about actions. In D. Michie (Ed.), *Machine intelligence 3*. New York: American Elsevier Publisher.

Benton, J., Do, M., & Kambhampati, S. (2009). Anytime heuristic search for partial satisfaction planning. *Artificial Intelligence*, *173*, 562–592.

Bradshaw, G. L., Langley, P., & Simon, H. A. (1983). Studying scientific discovery by computer simulation. *Science*, *222*, 971–975.

Cohn, A. G., & Hazarika, S. M. (2001). Qualitative spatial representation and reasoning: An overview. *Fundamenta Informaticae*, *46*, 1–29.

Duncker, K. (1945). On problem solving. *Psychological Monographs*, *58*, 1–113. Whole No. 270.

De Groot, A. (1965). *Thought and choice in chess*. The Hague, Mouton. Translated from *Het denken van den schaker*. Amsterdam, Noord-Hollandsche Uitgevers Maatschappij, 1946.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Faltings, B., & Struss, P. (Eds.) (1992). *Recent advances in qualitative physics*. Cambridge, MA: MIT Press.

Gajderowicz, B., Fox, M. S., & Grüninger, M. (2018). The role of goal ranking and mood-based utility in dynamic replanning strategies. *Advances in Cognitive Systems*, *9*, 211–230.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669.

Hayes, J. R., & Simon, H. A. (1974). Understanding written problem instructions. In L. W. Gregg (Ed.), *Knowledge and cognition*. Lawrence Erlbaum.

Hinrichs, T. R., & Forbus, K. D. (2014). X goes first: Teaching simple games through multimodal interaction. *Advances in Cognitive Systems*, *3*, 31–46.

Hinton, G. E., & Anderson, J. A. (Eds.). (1981). *Parallel models of associative memory*. Hillsdale, NJ: Lawrence Erlbaum

Huhns, M. N., & Singh, M. P. (Eds.) (1998). *Readings in agents*. San Francisco: Morgan Kaufmann.

Jones, R. M., & Langley, P. (2005). A constrained architecture for learning and problem solving. *Computational Intelligence*, *21*, 480–502.

Kirk, J. R., & Laird, J. E. (2014). Interactive task learning for simple games. *Advances in Cognitive Systems*, *3*, 13–30.

Koffka, K. (1935). *Principles of Gestalt psychology*. London: Routledge & Kegan Paul.

Kohler, W. (1940). *Dynamics in psychology*. New York: Liveright.

Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.

Kulkarni, D., & Simon, H. A. (1988). The processes of scientific discovery: The strategy of experimentation. *Cognitive Science*, *12*, 139–175.

Langley, P. (2000). The computational support of scientific discovery. *International Journal of Human-Computer Studies*, *53*, 393–410.

Langley, P. (2012). Intelligent behavior in humans and machines. *Advances in Cognitive Systems*, *2*, 3–12.

Langley, P. (2017a). Progress and challenges in research on cognitive architectures. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence* (pp. 4870–4876). San Francisco: AAAI Press.

Langley, P. (2017b). Interactive cognitive systems and social intelligence. *IEEE Expert*, *32*, 22–30.

Langley, P. (2017c). Heuristics and cognitive systems. *Advances in Cognitive Systems*, *5*, 3–12.

Langley, P., Barley, M., & Meadows, B. (2018). Adaptive search in a hierarchical problem-solving architecture. *Advances in Cognitive Systems*, *6*, 251–270.

Langley, P., Laird, J. E., & Rogers, S. (2009). Cognitive architectures: Research issues and challenges. *Cognitive Systems Research*, *10*, 141–160.

Langley, P., Simon, H. A., Bradshaw, G. L., & Żytkow, J. M. (1987). *Scientific discovery: Computational explorations of the creative processes*. Cambridge, MA: MIT Press.

McCarthy, J. (1960). Recursive functions of symbolic expressions and their computation by machine, Part I. *Communications of the ACM*, *3*, 184–195.

Mcshane, M., & Nirenburg, S. (2012). A knowledge representation language for natural language processing, simulation and reasoning. *International Journal of Semantic Computing*, *6*, 3–23.

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

Neches, R., Langley, P., & Klahr, D. (1987). Learning, development, and production systems. In D. Klahr, P. Langley, & R. Neches (Eds.), *Production system models of learning and development*. Cambridge, MA: MIT Press.

Newell, A. (1966). *On the analysis of human problem solving protocols*. Technical Report, Department of Computer Science, Carnegie Institute of Technology, Pittsburgh, PA. Reprinted in J. C. Gardin & B. Jaulin (1968), *Calcul et formalisation dans les sciences de l'homme*, 146–185. Paris.

Newell, A. (1969). Heuristic programming: Ill structured problems. In J. Aronofsky (Ed.), *Progress in operations research III*. New York: John Wiley.

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

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

Newell, A. (1985). Duncker on thinking: An inquiry into progress in cognition. In S. Koch & D. E. Leary (Eds.), *A century of psychology as science* (pp. 392–419). Washington, DC: American Psychological Association.

Newell, A., & Shaw, J. C. (1957). Programming the Logic Theory Machine. *Proceedings of the Western Joint Computer Conference* (pp. 230–240). Los Angeles: IRE.

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

Newell, A., Shaw, J. C., & Simon, H. A. (1957). Empirical explorations of the Logic Theory Machine. A case study in heuristic. *Proceedings of the Western Joint Computer Conference* (pp. 218–230) New York: Institute of Radio Engineers.

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

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, *19*, 113–126.

Ohlsson, S. (1984a). Restructuring revisited: I. Summary and critique of the Gestalt theory of problem solving. *Scandinavian Journal of Psychology*, *25*, 65–78.

Ohlsson, S. (1984b). Restructuring revisited: II. An information processing theory of restructuring and insight. *Scandinavian Journal of Psychology*, *25*, 117–129.

Ohlsson, S. (2012). The problems with problem solving: Reflections on the rise, current status, and possible future of a cognitive research paradigm. *The Journal of Problem Solving*, *5*, 101–128.

Polya, G. (1945). *How to solve it*. Princeton, NJ: Princeton University Press.

Richman, H. B., & Simon, H. A. (1989). Context effects in letter perception: Comparison of two theories. *Psychological Review*, *96*, 417–432.

Riddle, P. J., Barley, M. W., & Franco, S. M. (2013). Problem reformulation as meta-level search. In *Poster Collection of the Second Annual Conference on Advances in Cognitive Systems* (pp. 199–216). Baltimore, MD.

Selz, O. 1964 [1927]). The revision of the fundamental conceptions of the intellectual processes. In M. Mandler & G. Mandler (Eds.), *Thinking: From association to Gestalt*. New York: John Wiley.

Seo, M., Kembhavi, A., Farhadi, A., & Hajishirzi, H. (2017). Bi-directional attention flow for machine comprehension. *Proceedings of the Fifth International Conference on Learning Representations*. Toulon, France.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*, 701–703.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*, 129–138.

Simon, H. A. (1966). Scientific discovery and the psychology of problem solving. In R. G. Colodny (Ed.), *Mind and cosmos*. Pittsburgh, PA: University of Pittsburgh Press.

Simon, H. A. (1969). *The sciences of the artificial*. Cambridge, MA: MIT Press.

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology*, *7*, 268–288.

Simon, H. A. (1978). On the forms of mental representation. In W. Savage (Ed.), *Perception and cognition*. Minneapolis, MN: University of Minnesota Press.

Simon, H. A. (1981). Otto Selz and information-processing psychology. In N. H. Frijda & A. D. De Groot (Eds.), *Otto Selz: His contribution to psychology*. The Hague: Mouton Publishers.

Simon, H. A. (1981). In N. H. Frijda & A. deGroot (Eds.). *Otto Selz: His contribution to psychology*.

Simon, H. A., Langley, P., & Bradshaw, G. L. (1981). Scientific discovery as problem solving. *Synthese*, *47*, 1–27.

Spivey, M. (2007). *The continuity of mind*. New York: Oxford University Press.

Suchman, L. A. (1987). *Plans and situated action: The problem of human-machine communication*. New York: Cambridge University Press.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Valdés-Pérez, R. E. (1994). Human/computer interactive elucidation of reaction mechanisms: application to catalyzed hydrogenolysis of ethane. *Catalysis Letters*, *28*, 79–87.

Veloso, M. M., & Carbonell, J. G. (1993). Derivational analogy in PRODIGY: Automating case acquisition, storage, and utilization. *Machine Learning*, *10*, 249–278.

Vera, A., & Simon, H. A. (1993). Situated action: A symbolic interpretation. *Cognitive Science*, *17*, 7–48.

Wang, X. (1995). Learning by observation and practice: An incremental approach for planning operator acquisition. *Proceedings of the Twelfth International Conference on Machine Learning* (pp. 549–557). Tahoe City, CA: Morgan Kaufmann.

Wertheimer, M. (1959). *Productive thinking*. Tavistock Publications.

Yamauchi, B., Schultz, A., & Adams, W. (1998). Mobile robot exploration and map building with continuous localization *Proceedings of the 1998 IEEE International Conference on Robotics and Automation* (pp. 3715–3720). Leuven, Belgium: IEEE.

Żytkow, J. M., & Simon, H. A. (1986). A theory of historical discovery: The construction of componential models. *Machine Learning*, *1*, 107–136.