2. Herbert Simon on mind as computer

Gerd Gigerenzer and Daniel G. Goldstein

> We know how people design strategies, and even how they learn language and make scientific discoveries. In all these cases, we have examples of computer programs that perform these tasks in humanoid ways.
> (Herbert A. Simon, 1992, p. 150)

In this chapter, we try to bring to life the voice of Herbert Simon. The raw material is a set of letters and emails in which Simon comments on a manuscript of ours, subsequently published as “Mind as Computer: Birth of a Metaphor” (Gigerenzer and Goldstein, 1996). In that paper, we drew attention to the social origin of Simon’s view of cognition as computation, and analyzed the initial resistance to and belated acceptance of that view by his peers. In short, we argued that the first computers were human systems and that machine computers were built in their image. Inspired by Adam Smith’s division of labor, the French engineer Gaspard de Prony built a human computer from a system of workers, which in turn inspired Charles Babbage to build a mechanical computer. Later, Simon reversed the analogy and modeled humans in the image of a computer program. We argued that psychologists’ view of the relation between intelligence and calculation shifted from identity (brilliant arithmetic as the mark of a genius) to opposition (calculation as mechanical and mindless) and back (cognition *is* computation).

Simon’s comments on our account allow a glimpse into his scientific persona – a Renaissance man with titanic ambitions who rebelled against disciplinary boundaries. Although he behaved kindly to his students, he was always ready to challenge others’ ideas and was no stranger to sharp disputes. Simon had the courage and self-confidence to preach the heresy of bounded rationality to economists, and the unorthodoxy to maintain that computer programs are psychological theories to psychologists. He provoked many by equating humans with machines, undercutting the mainstream faith that humans differ in intelligence, creativity and emotions. As his biographer Hunter Crowther-Heyck (2005, p. 74) noted, Simon already possessed enormous self-confidence when he, as a student in Rudolf Carnap’s course at the University of Chicago, wrote a letter to his famous professor in which he praised Carnap’s book *The Logical Syntax of Language* but also criticized him for overvaluing formal linguistic coherence above correspondence with facts. Fittingly, young Simon’s church youth group in Milwaukee proudly called themselves “The Heretics” (p. 27).

Throughout his life, Simon rarely let a comment on his work go unchallenged if he disagreed with it. So it happened with the subtitle of our paper, “Birth of a Metaphor”:

> I do not, by the way, regard it [mind as computer] as a metaphor. As you undoubtedly know from my papers, I regard the programs that simulate human thought as psychological theories. Computer programs are systems of differential equations, and have the same claim to be treated as theories as the systems of differential equations of physics and other sciences.[1]

For almost anyone but Simon that is a bold claim. Yet Simon had been making that claim since 1957, when he announced in an address to the Operations Research Society of America that "there are now in the world machines that think, that learn, and that create". In the same talk he predicted that "within ten years most theories in psychology will take the form of computer programs, or of qualitative statements about the characteristics of computer programs" (Simon and Newell, 1958a, pp. 7–8). His speech drew heavy criticism not only from psychologists, but from mathematicians as well. Richard Bellman, who introduced dynamic programming in 1953, rejected the 'mysticism' he believed Simon spread by using mental terms for machines. "I feel that to use terms like 'that think, that learn, and that create' without a careful statement as to what is meant by these terms in connection with machines is irresponsible" (Bellman, 1958, p. 449). To which Simon and Newell (1958b, p. 450) unflinchingly replied: "We used these terms without special explanations because we meant by them exactly what they mean when they denote behavior of humans."

FROM TOOLS TO THEORIES

Simon's friend and collaborator Allen Newell studied with the mathematician George Pólya, who worked on using heuristics to solve mathematical problems. Newell was more focused on formal languages and notation systems than Simon, while Simon was more interested in making sure that their models matched human behavior. Their claim that the same language can be used to talk about processes in computers and minds is known as the *physical symbol system hypothesis*: a physical symbol system, such as a digital computer, has the necessary and sufficient means for general intelligence (see the epigraph to this chapter). Both computers and minds are physical systems that process symbols, hence the concept of mind as computer.

The creation of the theory of *mind as computer* was the topic of our paper. Specifically, we argued that its discovery and its belated acceptance followed a process known as the *tools-to-theories heuristic* (Gigerenzer 1991). It comprises two steps:

1. *Discovery*. New scientific tools, once entrenched in a scientist's daily practice, suggest new theoretical metaphors and concepts.
2. *Acceptance*. Once proposed by an individual (or group), the new theoretical metaphors and concepts are more likely to be accepted by the scientific community if the members of the community are also users of the new tools.

Scientific discovery is often presented as a mystery, or narrated in the form of anecdotes: the story of Newton watching an apple fall in his mother's orchard while pondering gravitation, or those about Fechner, Kekulé, Poincaré and others that link discovery to the 'three b's': beds, bicycles and bathrooms. The tools-to-theories process of discovery, in contrast, describes a specific pattern in which a tool for data analysis turns into a theory. Consider the earlier (and related) proposal of the mind as intuitive statistician. Before inferential statistics – more precisely, a mixture of Fisher's and Neyman and Pearson's methods – were introduced to psychologists, theories that modeled cognitive processes as statistical inference were virtually absent. Only afterwards did researchers model cognitive processes – from stimulus discrimination to causal attribution – as intuitive statistical inference. Consistent with the second part of the tools-to-theories heuristic, these theories were generally accepted only after statistical inference became institutionalized in psychology around 1955 (Gigerenzer and Murray, 2015).

Not all new theories are inspired by new tools, and tools-to-theories describes only one route travelled. In our paper, we argued that Simon and Newell had been guided by the tools-to-theories heuristic, and showed that psychologists accepted the hypothesis only after they had become familiar with personal computers for data processing. Simon was an actor in the discovery step, while his reluctant fellow psychologists were the actors in the acceptance step. Yet our analysis clashed with Simon's own account of discovery as a data-driven, not tool-driven process (Langley, Simon, Bradshaw, and Zytkow, 1987). His BACON program searches for invariants in data (such as constant differences or ratios) in order to rediscover laws such as Kepler's law or Ohm's law. How did Simon react to our defining his discovery as tool-driven?

> It is hard for one of the actors on the stage to comment on the drama, much less on the quality of the acting, but I can make a few remarks. You will not be surprised if I find the general thrust of your paper congenial, and say "right on!" The tools-to-theories heuristic is a valuable antidote to the ruling doctrine that theories, which descend from heaven (inspiration), are the beginning of science.[2]

Thus, Simon seemed to be content with our conjecture that mind as computer was inspired by a new tool rather than new data. But Simon would not be Simon if he had not immediately continued with a 'but': "Haven't you made the case a bit too one-sided, however?" After all, discovery has multiple causes, including the role of observations, which Langley et al. (1987) focused on in their book *Scientific Discovery*.

> Of course, the previous literature tends to find the first cause in theory […] Pointing to tools (as you do) and to data (as we do in SD) is an essential corrective. But let us not make the same mistake in reverse. Down with first causes![3]

Agreed. The question is not what came first. Rather, the question is how tools, data and theories are entangled in a process that more closely resembles a spiral than a straight line. And the dynamics of the spiral are the most interesting part. Consider again statistical tools. After these had inspired theories of mind, the kind of data obtained to test the theories changed. For instance, sensory discrimination, once understood in terms of absolute and differential thresholds, became modeled as statistical inference, known as *signal detection theory* (Tanner and Swets, 1955). This theory had exactly the same formalism as the tool Neyman–Pearson statistics. The new view that sensory discrimination is statistical inference introduced new cognitive concepts such as balancing false alarms with misses and, consequently, a new kind of data. Unlike in earlier studies on sensory discrimination, data about two kinds of errors were obtained, false alarms and misses, in order to test the predictions of the new theory that pictured the process as a cost-benefit analysis (Gigerenzer and Murray, 2015). In general, if a new tool inspires a new theory, it may require new kinds of data, which are then used to test the new theory. Discovery is not simply first, followed by justification, as has been taken for granted by both sides in the controversy over whether the context of discovery is actually relevant for science (Nickles, 1980). Popper famously argued that discovery is "irrelevant to the logical analysis of scientific knowledge" (1959, p. 31). We could not disagree more; discovery and justification are intertwined. If a theory is inspired by a new scientific tool, the tool has multiple functions: it provides the language for the theory, defines the kind of data obtained and serves to test the theory's predictions.

Simon's new information processing theory is another case in point. When the properties of computer programs became the new elements of the mind, computer simulation also became used to test the theory.

Discovery

Simon did not fall for the cult of statistical inference and *p*-values that was institutionalized in the mid-1950s and has survived to the present day (Gigerenzer, 2018). Statistical inference was neither his tool nor his theory of mind. Instead, Simon felt frustrated watching the mindless rituals in which his fellow psychologists routinely engaged: "The frustration lies not in the statistical issue itself, but in the stubbornness with which psychologists hold to a misapplication of statistical methodology that is periodically and consistently denounced by mathematical statisticians" (Simon, 1979, p. 261). For his research, the "familiar tests of statistical significance are inappropriate" (Simon, 1992, p. 159). As a consequence, his data-driven vision of scientific discovery was not modelled after statistical inference (Simon, 1973). Its logic was to take a data set and look for invariants or patterns, which recalls John Tukey's exploratory data analysis.

As mentioned, something other than patterns found in data led to the mind-as-computer hypothesis. In the summer of 1954, Simon learned to program IBM's first stored-program computer, the IBM 701 (Simon, 1991). Later in the summer, he began to think of humans and computers as systems that manipulate symbols. A tool, not data, inspired his hypothesis. Simon agreed, but with a caveat:

> Now back for a moment to the tools-theory-data theme. You are quite correct in saying that the computer was the crucial new thing that enabled the cognitive revolution. But you have to add to that the fact that only people who had already been wondering how to char-

> acterize humans as information processors could envisage using the computer in this way. Hence it is no accident that the first applications were made by people who were already deeply preoccupied with describing the information processing of decision makers – in the organization studies at Carnegie Tech, and in the big experimental study of an Air Force early warning station at Rand. Without that stimulus, I expect we would have been just as blind as the engineers and mathematicians. Even John McCarthy and Marvin Minsky, who certainly arrived independently at the idea of using computers to model symbolic processes, had no glimpse of heuristic search as the way to organize the programs – their model was formal logic and the lambda calculus of Church. Again, chicken and egg. The idea of thinking as information processing gave the idea of using the computer as tool, the tool made the idea tangible and implementable.[4]

Here, Simon emphasized that he and Newell thought about computer programs in a different way than McCarthy and Minsky, who initially tried to create rigorously logical programs. The idea of heuristic search rather than logical rules can be traced back on Newell's side to Pólya's heuristic methods and on Simon's side to his work on satisficing rules and bounded rationality (1955). For instance, their famous program for discovering proofs for theorems, the Logic Theorist (see below), proceeded by trial and error, guided by heuristics such as comparing the present state with the goal and selecting steps that reduce the difference – hence the term *heuristic problem solvers*. The essential point is that the very properties of the tool turned into those of the mind. In Simon's own words:

> Since computers are serial devices, requiring time to carry out their processes, it was natural to hypothesize the same seriality in human beings, and hence construct EPAM [Elementary Perceiver and Memorizer] in such a way that amount of learning would be roughly proportional to time. (Simon, 1979, p. 55)

Reluctant Acceptance

The acceptance step of the tools-to-theories heuristic predicts that the tool is unlikely to be accepted as a theory until the research community becomes familiar with the tool. Before the 1970s personal computer revolution, which enabled researchers to interact individually with a computer, the existing mainframe computers allowed them to interact solely via staff members. Moreover, those few familiar with mainframe computers were often frustrated. Consider Harvard's Center for Cognitive Studies, co-founded by George Miller, a proponent of the new information-processing psychology. The Center would appear a good candidate for the early adaption of Simon's new psychology. Miller et al.'s (1960) *Plans and the Structure of Behavior* was so near to Newell et al.'s (1958) ideas that Simon and Newell accused them of having stolen their ideas and getting them all wrong (the published book was eventually filled with citations of their work; Baars, 1986, p. 213). Yet when reading through the Center's annual reports of the 1960s, we found few papers or symposia dealing with computer simulation. The Center had a PDP-4C computer that served as a tool to run experiments, but as late as 1969, it conducted no cognitive simulations. One reason was frustration with the technology. From the average 83 hours the PDP was in use in an average week in 1965–1966, 56 hours were spent on debugging and maintenance. Getting a program to work could take months. Not surprisingly, a 1966 technical work was entitled *Programmanship, or How to Be One-Up on a Computer Without Actually Ripping Out Its Wires* (see Gigerenzer and Goldstein, 1996). Computers were a steady source of exasperation, delaying the move to the acceptance stage of the tools-to-theories heuristic. On this point, Simon agreed:

> I find your account of the reasons why the new tool was taken up so slowly to be convincing. The personal computers of the past decade or so, with their transportable disks, have made all the difference. … and the connectionists (whatever I think of their prospects) can thank the transportability of their programs for a great deal of their popularity.[5]

Previously, Simon himself had pointed to the link between psychologists' unfamiliarity with the tool and the hesitant acceptance of the theory: "Perhaps the most important factors that impeded the diffusion of the new ideas, however, were the unfamiliarity of psychologists with computers and the unavailability on most campuses of machines and associated software" (Simon, 1979, p. 365). Even with available funding for computer simulation of cognitive processes, Simon's students and colleagues at Carnegie-Mellon University showed little interest:

> Returning to CMU, the Psychology Department did not by any means receive the new IP religion with wholly open arms. Starting about 1961, Green, Newell and I received money from NIMH for research in computer simulation, and that project – which was for the whole cognitive part of the department – continued in one form or another for 20 years. Initially, we tried to persuade some of the younger people already in the department, or hired specifically for this purpose, to participate in the new activities on the grounds that IP psychology was the wave of the future; but most of them did not become believers, felt pushed, and stuck as close to classical Behaviorism as they dared.[6]

All that changed when personal computers began their invasion of the psychological laboratories in the late 1970s. For instance, Apple I and Apple II were first marketed in 1976 and 1977, respectively. Simon (1979) estimated that between about 1973 and 1979, the number of researchers at Carnegie-Mellon University in the new information processing psychology had "probably doubled or tripled" (p. 390). Once the tool was entrenched in psychologists' everyday practice, widespread acceptance of the vision of mind as computer followed. Since then, cognition is said to be computation. What else could it be?

SIMON'S FAITH IN CONTINUITY VERSUS NEWELL'S WORRIES

Information processing psychology introduced a theoretical language entirely foreign to psychology, borrowed from the programming world: "Explicit flow control, subroutines, recursion, iteration statements, local naming, production systems, interpreters, and so on. … We confess a strong premonition that the actual organization of human programs closely resembles the production system organization" (Newell and Simon, 1972, p. 803).

This new language went hand in hand with a change in methodology. The 1958 *Psychological Review* article by Newell et al. used standard terminology such as *experiment* and *subject*, but radically changed its meaning. 'Experiment' no longer meant a randomized controlled study but rather a computer simulation, and 'subject' no longer referred to a large number of undergraduates but instead either to an inanimate being, 'LT', as the Logic Theorist was known, or, in other research, to individual chess experts. The study of individual experts and the use of think-aloud-protocols was a break with the research practice of American psychology in the second half of the 20th century and more closely resembled that of German psychology in the first half of the same century. For instance, Wilhelm Wundt, considered to be the father of experimental psychology, typically served as the subject in his experiments (Danziger, 1990). In early German psychology, it was often the subject who published the article; in this very spirit, Newell and Simon named their subject the Logic Theorist, as a co-author of a paper submitted to the *Journal of Symbolic Logic*. Regrettably, the paper was rejected and the Logic Theorist never attempted to have anything published again.

Despite the fairly radical conceptual and methodological break with the psychology of the time, the 1958 *Psychological Review* paper stressed continuity, not change. That could be interpreted as a precautionary move at the time against rejection. But Simon remained adamant about continuity:

> The line we took in our 1958 Psych Review paper, stressing continuity rather than a break with the past, was not a ploy, but a genuine attempt to break the schools-of-psychology tradition based on a careful assessment of the work of Selz and Duncker, and of the solid base of experiments that had been laid down by the American behaviorists and functionalists (e.g., Woodworth). What people like Thomas Kuhn ignore when they write of scientific revolutions is that, while relativity and quantum mechanics may have represented a great break with the past, careful pains were taken to show how the classical theories provided excellent approximations as long as we didn't try to deal with the very fast (relativity) or the very small (quantum mechanics). There was no break, but an extension. So it is with IP [information processing] psychology.[7]

The reference to the German psychologists Otto Selz and Karl Duncker does indeed provide a sense of continuity. To study the nature of thought, Duncker asked subjects to 'think aloud' while trying to solve problems. These think-aloud protocols were the data from which he drew conclusions, such as that problem solving is a step-by-step process of successive reformulations of the problem until the final 'aha' moment arrives (Gigerenzer and Murray, 2015, ch. 5). Apart from this similarity, however, Simon's key message of thinking about mental processes in terms of computer processes was entirely unprecedented in both German and American psychology. Although one can debate what counts as a rupture with the past, Simon's mind-as-computer hypothesis is certainly a candidate. So why did Simon keep insisting on continuity? We think the answer can be found in his message of scientific progress in psychology. Throughout his career, Simon rejected the poor-mouthing of psychology as a discipline that reinvents the wheel, knows little today and is awaiting its own Newton or Einstein. He became quite impatient with the timidity of psychology in asserting its progress, knowledge and continuity:

> By any reasonable metric, we know more about the human mind and brain than geophysicists know about the plate tectonics that move the continents over the globe, far more than particle physicists know about elementary particles, or biologists about the processes that transform a fertilized egg into a complex multicellular organism. We discount our knowledge because some of it is so common-place, so familiar from our everyday acquaintance with ourselves and other people. (Simon, 1992, p. 1)

But here Simon and Newell disagreed. Where Simon saw continuity and progress, Newell saw a proliferation of phenomena that went in circles rather than spirals. In his paper "You can't play 20 questions with nature and win", Newell (1973) wrote that the one half of him believed all is well, but the other half was distressed and confused. In particular, the pessimistic Newell worried that explanations of phenomena proceed mostly by the construction of binary opposites, such as nature versus nurture, serial versus parallel processing, conscious versus unconscious and analog versus digital. With such general dichotomies, clarity is never achieved. Rather, "matters simply become muddier and muddier as we go down through time." (pp. 288–289). Simon was not pleased by these remarks, which led to the sole public disagreement between him and Newell in 35 years of collaboration. Responding to his friend, Simon asserted "that there *is* a way to win the that Twenty Questions game" (Simon, 1989, p. 20).

THE HUMAN ORIGIN OF COMPUTERS

From Sigmund Freud's principle of constancy of energy to Paul Samuelson's neoclassical economics, theories in the social sciences are sometimes modeled after the natural sciences, in particular physics. These origins are often presented as a qualification for being scientific. It is less often pointed out that the influence also works the other way round – another potential case of the lack of self-confidence in psychology and the social sciences in general about which Simon had complained. For instance, statistical mechanics was directly inspired by Adolphe Quetelet's (1969 [1842]) sociology (which he called *social physics*). Quetelet, who began as an astronomer, observed that the behavior of individuals (such as crime, marriage, suicide) was erratic and largely unpredictable, but viewed as a collective, their means and variances were stable and predictable. He used the normal distribution, which served astronomers as a tool for taming observational error, as a theory of society: the true position of a star (the mean) translated into *l'homme moyen*, the ideal average person, and observation errors into individual deviations from the ideal. Physicists Ludwig Bolzmann and James Clerk Maxwell pondered the erratic behavior of gas molecules and reasoned that these might behave as Quetelet's individuals did, predictable only when considered as a collective (Porter, 1986). By this route of discovery – from astronomers' model of observational errors to a theory of society to the behavior of gas molecules – the deterministic Newtonian world view was finally overthrown by a statistical view of nature (Gigerenzer et al., 1989).

The invention of the computer is another case where a social invention provided the inspiration for a physical one. During the French Revolution, engineer Gaspard de Prony organized the titanic project of calculating logarithmic and trigonometric tables with an unprecedented precision of up to 25 decimal places for the new decimal system (which was part of the revolution). Inspired by Adam Smith's division of labor, he divided the daring task into a hierarchy of subtasks. At the top were a few excellent mathematicians such as Adrien Legendre who devised the formulae, in the middle seven to eight persons trained in analysis and at the bottom 70 or 80 unskilled persons who performed millions of additions and subtractions. These *manufacturing methods*, as Prony called them, changed how people thought about calculation. In the Enlightenment, intelligence and even moral sentiment were viewed as combinatorial calculus, and great thinkers as proficient calculators (Daston, 1994). The story about Gauss's brilliance in arithmetic was one of the last of its kind. Prony's demonstration that calculation could be done by an assemblage of unskilled workers shifted calculation away from intelligence to that of mechanical work.

Charles Babbage (1994 [1812]) conceived of replacing these workers by machinery and built his first mechanical computers. In his chapter on the division of mental labor, Babbage explicitly referred to Prony's manufacturing methods as inspiration for a general science of machine intelligence (Schaffer, 1994). Thus, the machine computer was built in the image of the human computer.

When Newell and Simon tested their first computer program, the Logic Theorist, they, like Prony, used a human computer. The trial program consisted of human components, including Simon's wife, children and several graduate students. Each person became a subroutine of the program. The human origin of the machine computer was already known to Simon:[8]

> I liked your "computer as a factory" passage, and was glad to see Prony brought into the picture, and especially Adam Smith. I've always found that an excellent anecdote (I hope it is true) to show that the "hard" sciences really are founded on economics – it's great material for facing down the hubris of physicists.

But at least one observation we made was new to him:[9] "I have never thought about the parallel between our human simulation of LT and Prony's bureaux de calculs, and was delighted by the thought."

When Babbage attempted to replace the human computer with a machine, he reinforced Prony's divorce between intelligence and calculation. This divorce subsequently had its effect on psychological theory. For instance, calculation was absent both when Adolphe Binet and Théodore Simon devised the first test of intelligence in 1905 and in Louis Terman's Stanford-Binet test of 1916. Similarly, in the first half of the 20th century, psychologists understood thought as the association of ideas or successive restructuring of a task into a solution, not as statistical or algebraic calculation. Calculation had shifted from the company of 'great men' to a group denied prestige under sexism: women. The Manhattan Project at Los Alamos, where the atomic bomb was built in the 1940s, relied on human computers, mostly low-paid women, who repetitively performed the same calculations, such as cubing a number and passing it on to another person (Gleick, 1992). In the early 1980s, almost 40 percent of US graduates in computer science were women. Only when personal computers became available during that decade, marketed to men who enjoyed playing games, and computer science evolved into a highly paid profession, did men push their way into the profession. By 2020, the presence of women in US computer science departments had declined to 18 percent.[10]

Despite insisting on continuity with psychological theory, Simon turned the wheel back to the Enlightenment view of the mind, once again equating computation with human intelligence. Yet what computers in fact did was no longer mere calculation, but information processing of all kinds of symbols.

Is the Computer Like the Brain or the Mind?

Simon's mind-as-computer hypothesis was not the first analogy drawn between humans and the modern digital computer. In his 1958 Silliman lectures, which illness prevented him from finishing and delivering, the Hungarian mathematician John von Neumann pointed to similarities between the nervous system and the computer, as well as between the neuron and the vacuum tube, adding a word of caution about their differences. Turing (1950), in contrast, found similarities in hardware superficial. The fact that the modern computer and the nervous system are both electrical did not impress him, and he pointed to Babbage's purely mechanical computer. Whereas von Neumann thought about 'computer as brain', Turing thought about 'computer as mind'. Here is Simon's view:

> I could never understand how von Neumann landed on the wrong side of this issue. After all, he started out life in logic and with a good understanding of Goedel and Turing. The confusion has something to do – I haven't quite figured out what – with the idea that has been rampant in mathematics for 90-odd years that the whole purpose of logic (i.e., "mind") is to provide firm foundations for mathematics. Hence the logicians seized on Turing's characterization of Turing machines as generators of partial recursive functions, and forgot, or never knew, that any dynamic system – arithmetic or not – is computing a partial recursive function, but that it is probably not profitable to characterize it that way. Turing had two faces, and the formalists only saw one of them.[11]

Simon, like Turing, looked for similarities in function, not in hardware. But Turing looked in the opposite direction from Simon. He contemplated to what degree computers are like minds, whether computers can imitate humans and whether one could teach machines with the same techniques used to teach children. That is, Turing asked whether computers have features of the human mind, but not whether the mind is like a computer. For him, it was obvious that a human cannot imitate a computer;

a question such as “what is the square root of 365,364,363?” would quickly detect any human presenting as a computer. Simon, in contrast, tried to understand the human mind in the image of computer programs. Turing’s analogy was ‘computer as mind’, while Simon’s was ‘mind as computer’.

CHESS AND THE LIMITS OF SIMON’S ARTIFICIAL INTELLIGENCE

As mentioned above, in 1957 Simon made the bold prediction that within ten years most theories in psychology would take the form of computer programs. In the same talk, he also forecast that “within ten years a digital computer will be the world’s chess champion, unless the rules bar it from competition” (Simon and Newell, 1958a, p. 7). Neither prophecy was fulfilled. However, the second forecast did come about exactly three decades later than predicted, in 1997, when IBM’s computer program, Deep Blue, beat chess world champion, Gary Kasparov.

The victory did not happen in the way Simon expected. His vision of artificial intelligence (AI) was to extract the strategies (heuristics) expert chess players use and teach the computer how human experts solve the problem. The computer was the student and the human, the teacher. The computer could then execute the human strategies faster and with fewer errors. Let us call Simon’s vision *psychological AI*, where the ‘I’ refers to human intelligence. Deep Blue, however, was strikingly different. When Joe Hoane, one of Deep Blue’s programmers, was asked how much of his work was devoted specifically to artificial intelligence in emulating human thought, he responded: “It is not an artificial intelligence project in any way […] we play chess through sheer speed of calculation and we just shift through the possibilities and we just pick one line” (Krauthammer, 1997).

Unlike Simon, IBM did not try to build a computer in the image of the human mind. Nor did Google when it built AlphaZero in 2017. AlphaZero only needs to know the rules of the game and uses brute computing power to play millions of games against itself and learn how to win at chess, Go and shogi (Japanese chess) by reinforcement and by trial and error. These machine learning systems are also often called AI, but here the ‘I’ no longer refers to human intelligence. Simon had predicted the win, not the route towards it. Advances in hardware engineering rather than psychological AI won out in chess.

Known as one of the fathers of AI, Simon did not himself coin the term. In the *Sciences of the Artificial* he states that he prefers the terms *complex information processing* and *simulation of cognitive processes*, while adding a qualification:

> But then we run into new terminological difficulties, for the dictionary also says that “to simulate” means “to assume or have the mere appearance or form of, without the reality; imitate; counterfeit; pretend.” At any rate, “artificial intelligence” seems to be here to stay, and it may prove easier to cleanse the phrase than to dispense with it. (Simon, 1969, p. 7)

Yet cleansing is not what happened. Rather, the phrase *AI* became used for projects unrelated to human intelligence, such as deep neural networks, random forests and other machine learning techniques that work without direct inspiration by theories of the mind. These projects are fueled by Moore's law, according to which computing power doubles every two years or so. That development raises the question, does Simon's psychological AI have a future?

We will never know how Simon would have responded had he lived to witness the success of deep artificial neural networks in the early 21st century. In the next and final part, we nevertheless take the liberty of suggesting an answer that rescues but also limits his program of psychological AI.

Well-Defined Versus Ill-Defined Problems

We will distinguish three possible answers to the question of the future of Simon's AI. The first would be to persevere with Simon's program of analyzing experts' heuristics in order to make computers smart, no matter what the domain. Yet such a view is no longer tenable now that machine learning has outwitted this program in several areas beyond chess. The second response would be that psychological AI is inferior to machine learning in all domains and generally inept at designing smart algorithms; its role should therefore be restricted to the description of psychological processes. However, we do not see any evidence for such a general claim. Instead, we have evidence of machine learning systems performing excellently in some domains but less so in others. This brings us to a third possible answer, that there is a particular domain in which psychological AI can match or outperform machine learning algorithms while having additional advantages such as transparency, but that it is not a general problem solver in all domains.

What then is its domain? In our view, the strength of psychological AI lies in ill-structured problems, and less so in well-structured problems (Gigerenzer, 2022). Chess is a well-structured problem, where all possible actors (the pieces) are known, their movements are governed by a set of unambiguous rules that are stable in the sense that they hold both today and tomorrow, and the two players have a single goal, to checkmate the other's king. In other words, a well-structured problem is a stable world where nothing unexpected can ever happen. A game of roulette also represents a stable world, but a probabilistic rather than a deterministic one, where all possible future events (the individual pockets where the ball can land), their consequences (the pay-offs) and the probabilities are all known. Frank Knight, professor at the Chicago Department of Economics when Simon was a student, used the term *risk* for such situations, to be distinguished from *uncertainty*, where complete knowledge is not attainable. Under Knightian uncertainty, rules are ambiguous, partially known and can be negotiated and violated.

Simon's early work on decision making in administrations dealt with ill-structured problems. For instance, while working on budget decision problems in his native Milwaukee's recreation department, he concluded that the economic framework of utility maximization "was hopeless" (Simon, 1988, p. 286). Instead, he learned that managers relied on heuristics such as adding incremental changes to last year's budget. That led to his work on satisficing and bounded rationality.

Note that we did not have the opportunity to discuss this issue with him, and he might have disagreed. At least in earlier writings, he tried to play down the difference between well-defined and ill-defined problems:

> I will try to show that there is no real boundary between WSPs [well-structured problems] and ISPs [ill structured problems], and no reason to think that new and hitherto unknown types of problem solving processes are needed to enable artificial intelligence systems to solve problems that are ill structured. (Simon, 1973b, p. 182)

In this article, Simon went as far as to argue that chess was not a well-defined game because it is intractable. Yet a game can be proven intractable only when it is well-defined in the first place, such as in the case of Tetris or backgammon.

The Future of Simon's Psychological AI

Here, Simon appears to have overlooked a fundamental issue. Human intelligence evolved to solve ill-structured problems, not well-structured ones. These are the proper domain of Simon's psychological AI, a domain in which it can compete with machine learning (Gigerenzer, 2022).

Our research group has demonstrated that in ill-structured situations, simple heuristics of the type Simon envisioned can perform at the level of state-of-the-art machine learning systems, or of complex statistical algorithms in general. Here are some illustrations. Consider first a class of heuristics called *fast-and-frugal trees*, which model how human experts make decisions under uncertainty in a sequential way. These simple decision trees can match random forests in predictive accuracy and have the advantage that they are transparent, easy to understand and learn, and can be quickly executed in emergency situations (Katsikopoulos et al., 2020). Next, consider the highly publicized Google Flu Trends, an algorithm designed to forecast the spread of flu-related doctor visits, which is based on the analysis of 50 million search terms and 450 million algorithms. In our research we found that a simple rule that uses only a single data point, the *recency heuristic*, forecast flu-related doctor visits substantially more accurately than Google's big data algorithm (Katsikopoulos et al., 2022). Finally, consider the US presidential elections of 2016, where big data analytics and election markets had predicted Hillary Clinton's victory by a large margin. In contrast, a simple *tallying heuristic* that only counts reasons for and against the contestants correctly predicted Trump's win and had previously predicted all but one presidential election correctly since 1984 (Lichtman, 2016).

The important point is that in each of these cases, the starting point is an analysis of human heuristics and content-related knowledge, which are then programmed into a computer to make decisions or predictions. Big data and machine learning, in contrast, pay little attention to the evolved heuristics that humans use and often ignore the content of the problem. Note that simple heuristics work in situations of uncertainty, where unexpected changes occur and people's behavior is guided by unpredictable factors. In well-structured problems such as chess, however, simple heuristics cannot easily compete with machine learning methods (Gigerenzer, Hertwig, and Pachur, 2011).

In *Simple Heuristics That Make Us Smart* (Gigerenzer, Todd, and the ABC Research Group, 1999), we proposed the study of heuristics people use to deal with situations of uncertainty. Simon read the proofs and seemed quite pleased about our extension of his work, writing in a blurb that the book "offers a fascinating introduction to this revolution in cognitive science, striking a great blow for sanity in the approach to human rationality" (Simon, 1999). At the same time, he remarked:

> I was glad to see the page devoted to emotions, social norms, and imitation. My own approach to introducing emotion into the picture is to introduce focus of attention as a major component of bounded rationality. I'll put in the mail for you a piece I did a few years ago. As to social norms, I think one of the deficiencies of current theory is the failure to introduce identification (a.k.a. "loyalty") to groups as a major element of human motivation, and a major explanatory mechanism for the understanding of organizations (to say nothing of ethnic conflicts!).[12]

His praise concerning emotions, social norms and imitation was, typically, double-edged. He was glad to see the page – the only one on the topic in the entire book. But he also pinpointed an important unresolved problem: how to model emotions as heuristic search or stopping procedures. We had made some rather concrete suggestions, without following directly up on these. Imitation, at least, is comparatively easy to model in the context of simple heuristics (e.g., Garcia-Retamero, Takezawa, and Gigerenzer, 2009).

Simon observed the rising power of programs that take computational efficiency instead of psychological reality as their criterion. In the wake of this, he emphasized that Newell and he "were interested in simulating human problem solving, and not simply demonstrating how computers could solve hard problems" (Simon, 1991, p. 209). At that time, he had to acknowledge that machine learning and automatic programming could design faster computer proof procedures than the Logic Theorist, who was denigrated as primitive by his critics (Simon, 1991, p. 209). Yet he responded that these critics simply misunderstood that his objective was to study the mind, not to construct a program that can solve a problem. In our opinion, the distinction between ill-structured and well-structured problems does not restrict Simon's AI program to psychological models but makes them a strong competitor in engineering solutions when predicting uncertain events.

CONCLUSION

AI has come to mean many things, including big data analytics and machine learning. For Simon, it meant something different and quite specific: computer programs that perform tasks the way humans do (hence the 'I' for human intelligence). He believed that the nature of the intelligence of humans and computers was essentially identical. Thus, computers could be made smart by human heuristics, and the mind could be understood in terms of what computers do. In this chapter, we ask where Simon's vision of the mind as computer originated. Unlike in Simon's own work on discovery, the BACON programs, which look for patterns in data, his theory of mind as computer was inspired by new tools rather than by new data. We have used his comments on our 1996 paper to bring his voice to life, a voice that believes in progress in psychology, expresses impatience with fellow psychologists who are too timid to assert their knowledge and sees continuity where others see breaks. His voice is also that of a Renaissance scholar, equipped with enough knowledge to cross the borders of disciplines while understanding what they have in common. In Simon's travels through economics and artificial intelligence, the common denominator is psychology.

NOTES

1. H. A. Simon, letter to Daniel Goldstein, Max Planck Institute for Psychological Research, Munich, dated 14 March 1996.
2. H. A. Simon, email to Gerd Gigerenzer, University of Chicago, 30 September 1994.
3. H. A. Simon, email to Gerd Gigerenzer, University of Chicago, 30 September 1994.
4. H. A. Simon, email to Gerd Gigerenzer, University of Chicago, 30 September 1994.
5. H. A. Simon, email to Gerd Gigerenzer, University of Chicago, 30 September 1994.
6. H. A. Simon, email to Daniel Goldstein, University of Chicago, 22 July 1994.
7. H. A. Simon, email to Gerd Gigerenzer, University of Chicago, 30 September 1994.
8. H. A. Simon, email to Gerd Gigerenzer, University of Chicago, 30 September 1994.
9. H. A. Simon, email to Gerd Gigerenzer, University of Chicago, 30 September 1994.
10. https://www.computerscience.org/resources/women-in-computer-science/
11. H. A. Simon, email to Gerd Gigerenzer, University of Chicago, 30 September 1994.
12. H. A. Simon, email to Gerd Gigerenzer, Max Planck Institute for Human Development, 4 May 1999.

REFERENCES

Baars, B. J. (1986). *The cognitive revolution in psychology*. Guildford.

Babbage, C. (1994). *Passages from the life of a philosopher* (M. Campbell-Kelley, Ed.). IEEE Press. (Original work written ca. 1812).

Bellman, R. (1958). On "heuristic problem solving," by Simon and Newell. *Operations Research*, *6*, 448–449.

Crowther-Heyck, H. (2005). *Herbert A. Simon: The bounds of reason in modern America.* Johns Hopkins University Press.

Danziger, K. (1990). *Constructing the subject: Historical origins of psychological research.* Cambridge University Press.

Daston, L. (1994). Enlightenment calculations. *Critical Inquiry*, *21*, 182–202.

Garcia-Retamero, R., Takezawa, M., and Gigerenzer, G. (2009). Does imitation benefit cue order learning? *Experimental Psychology*, *56*, 307–320.

Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, *98*, 254–267.
Gigerenzer, G. (2018). Statistical rituals: The replication delusion and how we got there. *Advances in Methods and Practices in Psychological Science*, *1*, 198–218.
Gigerenzer, G. (2022). *How to stay smart in a smart world.* MIT Press.
Gigerenzer, G., and Goldstein, D. G. (1996). Mind as computer: Birth of a metaphor. *Creativity Research Journal*, *9*, 131–144.
Gigerenzer, G., and Murray, D. J. (2015). *Cognition as intuitive statistics*. Psychology Revivals Series. Psychology Press.
Gigerenzer, G., Hertwig, R., and Pachur, T. (Eds.) (2011). *Heuristics: The foundations of adaptive behavior.* Oxford University Press.
Gigerenzer, G., Swijtink, Z., Porter, T, Daston, L., Beatty, J., and Krüger, L. (1989). *The empire of chance: How probability changed science and everyday life.* Cambridge University Press.
Gigerenzer, G., Todd, P. M., and the ABC Research Group. (1999). *Simple heuristics that make us smart*. Oxford University Press.
Gleick, J. (1992). *Genius. The life and science of Richard Feynman*. Pantheos.
Katsikopoulos, K., Şimşek, Ö., Buckmann, M., and Gigerenzer, G. (2020). *Classification in the wild.* MIT Press.
Katsikopoulos, K., Şimşek, Ö., Buckmann, M., and Gigerenzer, G. (2022). Transparent modeling of influenza incidence: Big data or a single data point from psychological theory? *International Journal of Forecasting*, *38*, 613–619.
Krauthammer, C. (1997, May 26). Be afraid. *Washington Examiner.* Retrieved from https://www.washingtonexaminer.com/weekly-standard/be-afraid-9802
Langley, P., Simon, H. A., Bradshaw, G. L., and Zytkow, J. M. (1987). *Scientific discovery: Computational explorations of the creative processes.* MIT Press.
Lichtman, A. J. (2016). *Predicting the next president: The keys to the White House*. Rowman and Littlefield.
Miller, G. A., Galanter, E., and Pribram, K. H. (1960). *Plans and the structure of behavior.* Holt, Rinehart & Winston.
Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing: Proceedings of the Eighth Annual Carnegie Symposium on Cognition* (pp. 283–308). Academic Press.
Newell, A., Shaw, J. C., and Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review*, *65*, 151–166.
Newell, A., and Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.
Nickles, T. (1980). Introductory essay: Scientific discovery and the future of philosophy of science. In T. Nickles (Ed.), *Scientific discovery, logic, and rationality* (pp. 1–59). Reidel.
Popper, K. (1959). *The logic of scientific discovery*. Basic Books.
Porter, T. (1986). *The rise of statistical thinking, 1820–1900*. Princeton University Press.
Quetelet, L. A. J. (1969). *A treatise on man and the development of his faculties* (R. Knox, Trans.). Scholars' Facsimiles and Reprints. (Original work published 1842)
Schaffer, S. (1994). Babbage's intelligence: Calculating engines and the factory system. *Critical Inquiry*, *21*, 203–227.
Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*, 99–118.
Simon, H. A. (1969). *The sciences of the artificial*. MIT Press.
Simon, H. A. (1973a). Does scientific discovery have a logic? *Philosophy of Science*, *40*, 471–480.
Simon, H. A. (1973b). The structure of ill-structured problems. *Artificial Intelligence*, *4*, 181–201.

Simon, H. A. (1979). *Models of thought* (Vol. I). Yale University Press.
Simon, H. A. (1988). Nobel laureate Simon "looks back": A low-frequency mode. *Public Administration Quarterly*, *12*, 275–300.
Simon, H. A. (1989). How to win at twenty questions with nature. In H. A. Simon, *Models of thought* (Vol. II, pp. 20–29). Yale University Press.
Simon, H. A. (1991). *Models of my life*. New York: Basic.
Simon, H. A. (1992). What is an "explanation" of behavior? *Psychological Science*, *3*, 150–161.
Simon, H. A. (1999). Blurb for Gigerenzer, G., Todd, P. M., and the ABC Research Group. In *Simple heuristics that make us smart*. Oxford University Press.
Simon, H. A., and Newell, A. (1958a). Heuristic problem solving: The next advance in operations research. *Operations Research*, *6*, 1–10.
Simon, H. A., and Newell, A. (1958b). Reply: Heuristic problem solving. *Operations Research*, *6*, 449–450.
Tanner W. P. Jr., and Swets, J. A. (1955). A decision-making theory of visual detection. *Psychological Review*, *61*, 401–409.
Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*, 433–460.
von Neumann, J. (1958). *The computer and the brain*. Yale University Press.

PERSONAL NOTE

I was fortunate enough to interact with Herbert Simon regularly between 1975 and 1984, first as a graduate student in Carnegie Mellon University's psychology department and then as a collaborator on joint projects. My presence at CMU was not an accident; I had come specifically to work with him on a particular topic – scientific discovery.

My undergraduate interests were wide ranging, but it seemed that in every area I found intriguing – human cognition, artificial intelligence, philosophy of science, econometrics – he had done seminal work, so I naturally concluded that I needed to study with this daunting polymath. My thinking was especially influenced by a 1966 chapter, "Scientific Discovery and the Psychology of Problem Solving", in which Simon dared to suggest that we might explain creativity in science as a form of problem solving, that is, as heuristic search through a space of symbol structures.

My dissertation research took this idea as its starting point, and the result was a set of AI systems that rediscovered laws from the history of physics. These defined new terms as algebraic combinations of existing ones and carried out heuristic search for terms with constant values. Symbols and search were central to this research program.

Along the way, I became convinced that the two metaphors were central to understanding both human and machine cognition. I went on to look at their dual role in explaining other phenomena, such as concept formation, syntax acquisition and even learning to improve problem solving itself.

As my career continued, these initial interests expanded to include other themes – such as memory, reasoning and expertise – that appeared necessary for a full account of cognition, but symbols and search have remained the core of my scientific world view and they continue to serve me well.

Pat Langley