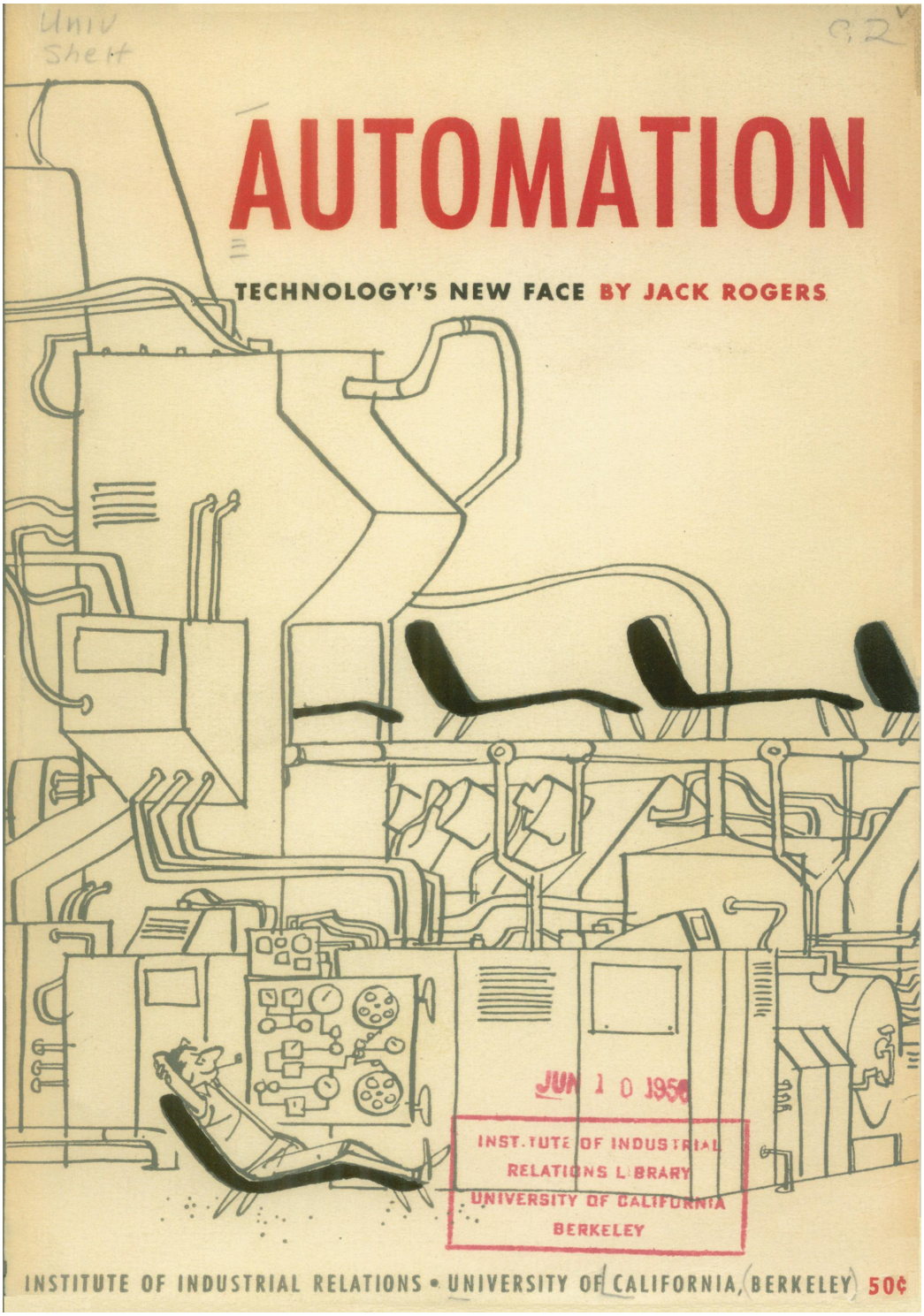
Herbert Simon and the Image of the Future

Daniel Kelly, UC Berkeley



David McDonald was a native son of Pittsburgh and a steel man. Born in the city in 1902, McDonald was the President of the United Steelworkers, one of the largest and most powerful unions in Pennsylvania. He had been with the union through its first decades, as it flourished under the mostly sympathetic eyes of Roosevelt and Truman. The past few years had been hard ones, though, especially after an exhausting, months-long strike carried out in the teeth of opposition from President Eisenhower. In 1961, McDonald traveled to Washington to explain himself, his strike, and his growing fear for the future of labor.

His testimony before Congress was alarmist by design. At the heart of his presentation-- almost a jeremiad--was a new wave of self-regulating and computerized production processes known by the catch-all term "Automation."

To the modern ear, "Automation" is a commonplace word. In 1961, it was still new-fangled—suggestive of mechanical wonders, to be sure, but also vaguely threatening. It had been coined only nine years earlier, to designate the continuous-process and feedback-guided production systems which had become industry-standards after the war.[1] Yet by 1961 the word was everywhere. What historians call "The Automation Debate,"--a wide ranging discussion of postwar economic arrangements and their compatibility with technological change--was taken up by anyone with an interest in the direction of big-science capitalism, from mainstream and business presses to liberals and the New Left. It remained a lodestar of political discussion from roughly 1956, at the moment of the earliest mechanization strikes and congressional hearings, well through to the late 1960s, when Vietnam and economic crisis and all the other storms of the early 1970s inundated the political landscape, submerging this Fordist discourse beneath the waterline.[2]

Amongst the notables who took Automation to be a significant new historical force were management leaders like Peter Drucker and John Diebold, a host of liberal and leftist theorists, from Daniel Bell to Martin Luther King to the SDS, and even so idiosyncratic a thinker as Hannah Arendt, who framed the significance of her Magnus Opus, *The Human Condition*, around the dual shock of space travel and automated production, in her view the most historically significant events of the modern age. Professionally speaking, it was a flashpoint of concern for economists, labor leaders, and most of the American social science establishment. When, in 1962, the Department of Labor compiled a reference bibliography of the most important works on the subject of Automation, they discovered over 500 relevant titles since 1956 alone.[3]

McDonald's testimony was largely reflective of the rhetorical moves patterned into the Automation Debate. The American New Deal system was premised on full employment under the protection of large unions—what might machines do to this political settlement? McDonald warned the congressmen that Automation could, at least in theory, displace 90 percent of all Americans workers within a decade (the most likely source for this information was the "father of Cybernetics," the mathematical polymath

---

[1] John Diebold, *Automation: The Advent of the Automatic Factory* (New York: Van Nostrand, 1952).

[2] To learn more about the understudied history of the Automation Debate, please see the work of Howard Brick and David Steigerwald. I have also benefited from listening to the ideas of Angus Burgin.

[3] The Department of Labor, The Bureau of Labor Statistics. *Implications of Automation and Other Technological Developments: A Selected Annotated Bibliography, Bulletin #319*. (Washington DC, 1962) 1.

Norbert Wiener, who by then had turned against further scientific work on ethical and political grounds). At times McDonald almost seemed embarrassed by the fantastical possibilities he asked the congressmen to entertain—the technology to come, he worried, could be like something out of "Buck Rodgers," Science Fiction visions made real, at the expense of American Labor. Fantastical or no, the political, social, and cultural consequences of further technological change would be vast. Whether Man could accommodate to such plenty and plentiful leisure, or Congress assure their proper distribution amongst men, was up to the present generation. The coming society, McDonald analogized, would necessarily be a Roman slave society. The question was, Who would be the slave?—machines to men, or men to machines and their owners? However unlikely it may have seemed to Congress, his union had come to believe that automation was the major threat facing American labor and American democracy.[4]

***

Herbert Simon also lived in Pittsburgh, but he was no steel man. Instead, he was a professor in Carnegie Tech's School of Industrial Management, an experimental program which sought to weld the technical training of engineers to the traditional education of a business manager. A polymath and a social science iconoclast, Simon had begun his career as an evangelist for more formal, quantitatively sophisticated social sciences. He'd originally trained to be a political scientist, writing a dissertation and field-defining analysis of the structure of administrative organizations. His later work moved into economics, game theory, and the cognitive foundations of decision-making, especially the computational limits of making a correct choice, which Simon called "bounded rationality." By the time of David McDonald's Automation testimony, though, he'd become a computer scientist. He'd also become a prophet of a totally automated Mind—one of the first proponents, to be more specific, of a radical vision of Artificial Intelligence.

Because of this surprising mid-career turn, from student of administrative systems and critic of the economics profession, to developer of perhaps the most jarring automation of them all, the automation of thinking, it would have been reasonable to expect that when Herbert Simon sat down to write out his considered views on the topic of Automation, he would potentially agree with David McDonald about the scope of Automation's power and the need for greater public action.

As it happens, however, he did not. From his first writings on the subject in 1960, to the book he wrote on the effect of computers on the future of labor in 1965, Simon was an avowed skeptic of any great impact of computing on labor. Setting aside his belief that "in [his] time, computer will be able to do anything a man can do…read, think, learn, create," he still held that "mankind will not find the life of production and consumption in a more automated world greatly different from what it has been."[5]

In this paper I would like to trace how Simon came to hold these two very complicated, seemingly conflicting views—what he himself called his "technological

---

[4] House Committee on Education and Labor; Subcommittee on Unemployment and the Impact of Automation, Hearing on the "Impact of Automation on Employment." March 8th, 1961. Pg. 59.

[5] Herbert Simon, *The Shape of Automation for Men and Management* (New York: harper and Row, 1965) xiii.

radicalism" with his "economic conservatism." The tension between these views deserves attention, first, because Simon was recognized, as an economist and technologist, to be uniquely qualified to comment on the economic impact of technology, and his work as a prominent professor and consultant amplified the power of his voice.

Perhaps more importantly, however, is the way his own shifting views help illuminate the debate over technology itself, and the intellectual ambitions of Artificial Intelligence researchers within that matrix. Because of his mixed up and unusual career—marked as it was by continuously shifting, and I would argue, conflicting disciplinary loyalties—historians can use him to trace the professional pressures and contradictions, the intellectual flows and eddies, which carried along those who dove into the debate about technology. As an ambitious computer scientists, Simon displayed an almost Faustian belief in the computer and its power to revolutionize human life. As an economist and political scientist, he was loath to deny the predictive power of the social sciences, even if those sciences played down the impact of the machines he believed to be so revolutionary. When he tried to defend both disciplinary commitments, he produced a deeply revealing failure, like the debris of a smashed atom revealing its constituent parts.

If we want to see the debris, though, we must first put him in motion.

***

The way his most prominent student remembered it, Simon began the Spring 1956 semester by marching into the classroom and announcing "Over the Christmas holiday Al Newell and I invented a thinking machine!"[6] A computer that could think? The first digital computers had only come intro existence a few years before, mostly as tools for prosecution of the war, usually as slavish number crunchers, designed exclusively for the calculation of tedious equations. The idea of a "thinking machine" must have been quite a startling departure.

As it happened, the fact that Herbert Simon was working with computers *at all* must have come as a surprise, given where he began his academic career. He had originally come to Carnegie Tech from Chicago, where he had been something of a polymath, a star graduate of Charles Merriman's department of Political Science. At the time, Chicago was one of the leader's of the "Behavioral" revolution in political science. Simon's undergraduate and graduate work, which oftentimes transcended pure political science to include advanced quantitative training and logic, the philosophy of science, and economics, was guided from the beginning by the Behavioralist imperative to practice Political Science *as a science*. Traditional Political Science departments, in this view, were the provenance of dilettantes, dabbling widely in political philosophy, history, and uncritical descriptions of the law.[7] For Simon and his colleagues at Chicago, science meant neither mere description nor normative judgment. Although no single method emerged, Chicago Behavioralism tended to a focus on how organizational or macro-political phenomenon arose from the decisions of individual actors (usually defined by formally specified aims), and preferred to work with polling data, sampling, and other numerical and seemingly objective metrics of political reality. As a short hand, imagine

---

[6] The student was Edward Feigenbaum, who eventually became a computer scientist himself. The story is recounted in Herbert Simon, *Models of My Life* (New York: Basic Books, 1991) 206.

[7] Simon, *Models of My Life,* 55.

that the Behavioralists wished to make Political Science less like history, and more like economics.

In accord with this program, Simon's dissertation aimed to be a rigorous foundation for the study of bureaucracies, focusing not on their culture, habits, or history, but rather on how individuals within the bureaucracy were functionally slotted into the organization, came to inhabit their organizational roles, and acted as conduits of decisions--evaluating, choosing, and executing the plans which are the stuff of any administration. The book that resulted from this research, *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization* (1947), quickly achieved the status of a classic. Simon used his success to open the School of Industrial Management at Carnegie Tech, which he presided over essentially as director of research, parlaying the position into appointments with the Ford Foundation, the Rand Institute, a Nobel Prize, and much else that was influential in mid-Century social science.

His preference for a methodologically individualist analysis of collective organization put Simon in the company of rising powers, and the similarities to other works helped his colleagues recognize the value of his contribution. Indeed, at that time scholars from across the academy were becoming taken with Oskar Morgenstern and John Von Neumann's work on game theory, and there were a number of attempts to dissolve political organizations, voting mechanisms, bureaucracies, and much else to their individualist logics, whether in Economics, Public Choice, Social Choice, or Positive Political Theory. In all these, the Post-Keynesian emphasis on macro-units rather than micro-foundations was slowly being reversed.[8]

The similarities shouldn't be overemphasized, however. Simon travelled down a busy path, but from a different starting place, and towards a very different destination. For one thing, Simon's work was never motivated by political conservatism: he never questioned the legitimacy or efficacy of the bureaucracies he studied.[9] Unlike in Public Choice scholarship, for example, the reduction of bureaucracies to the individuals in them was not meant to disprove the workability or public-spiritedness of bureaucracy as such[10]—rather, the reduction was interesting to Simon *because the bureaucracies generally worked.* He was neither exposing nor debunking. Second, Simon's reductions did not reduce to the *homo economicus* of neo-classical theory. In contrast to the self-interested schemer who is economic man, Simon's individuals could adopt the bureaucracy's imperatives for their own reasoning ("organizational identification"), and they could act upon the commands of others without trying to gain from them

---

[8] For an account of the macro/micro question in economics, the social science hegemon par excellence, please see Daniel Rodgers, *The Age of Fracture* (Cambridge: Belknap Press, 2003), especially chapter 2.

[9] Simon's early political work was connected to Merriman's interest in the City Manager's Movement, a progressive campaign to rationalize the administration of city government. He later worked on projects at Berkeley intended to benefit New Deal programs, and even into late life he characterized his own politics as those of a "New Deal Liberal." His friendly interest in Socialism is also very much unlike the explicit anti-communism (and even anti-New Dealerism) one finds in Public Choice theory.

[10] For an account of the politics of Public Choice theorists, see Sonya Amadae's *Rationalizing Capitalist Democracy*. To summarize, the movement sought to undermine any notion of a "public" over and above individuals, denied the possibility of measuring their welfare or aggregating their democratic preferences (necessary for any sense of the public good or the public will), and was endlessly delighted by the supposed perversions of purpose which utility-maximizing individuals introduced into public administration.

("authority"). And when these individuals reasoned about options, it was not in search of a utility-maximizing best-case scenario, as only a greedy genius might find, but among relatively clear alternatives designed to satisfy felt needs ("bounded rationality" intended to "satisfice.")

This last point was decisive. Having reduced a social whole, an administrative bureaucracy, down to the people and procedures that made it go, Simon discovered—or probably invented—a conception Man as a computationally imperfect decider. Simon's man chose amongst limited numbers of options that were good enough, rather than effortlessly discovered the best of all possible options. He was, in short, a man, and not a God. And when Simon considered how this limited creature could reason so well as he did, he decided this reasoning process was also worthy of further decomposition and reduction to smaller, more limited parts. And to do this, he decided, his man would be a computer.

***

The thinking machine Simon introduced to his class in January 1956 was not actually an up-and-running electronic computer, but rather was a "paper-machine," a program of computer pseudo-code detailed enough (but also general enough) to simulate how a computer would run a program. That is, it was a program for a man simulating a machine simulating a man. The qualification hardly dampened Simon's enthusiasm. At the heart of his enthusiasm were two redefinitions, of man and computer, which would be the foundation of his work for the rest of his career. First, the computer. Early computers had been numerical devices, essentially calculators, and they took their name from the female workers who'd previously performed the same function before she'd been automated from the office. For these reasons, Simon was surprised to discover that the computer he was introduced to at a summer program at Rand, the JOHNNIAC, could in principle manipulate any symbol. By understanding the machine as a physically-instantiated symbol manipulator, especially one limited in computational speed and memory, Simon, along with many others in the mid-1950s, believed they had a machine whose power and limits mimicked man's own.[11]

The man who was to be mimicked was more familiar. He was the individual of Simon's own work.

Indeed, Simon's work on thinking machines, quickly recast as "Artificial Intelligence" by the 1956 summer conference at which he displayed his work, was based on a collapsing definition of man as decision-maker, thinker, and problem solver, words he used interchangeably through his research and work. Two of his earliest programs, the Logic Theorist (LT) and the ambitious General Problem Solver (GPS) give a clear vision of how this Artificial Intelligence was intended to function.

In Simon's programs, intelligent decision meant taking a system of symbols, on one end, and an ideal symbolic goal, on the other, and searching through the various permissible lines by which the starting symbols could be transformed until they satisfied the goal. His program therefore had to have the capacity to represent systems of symbols, to compare provisional systems to the goal state, and to properly transform a system into

[11]Indeed, this belief was basic to the so-called "cognitive" revolution in psychology, linguistics, and much else.

a new system for further exploration, to "explore a search space" of possible lines of transformation. In practice, this could mean arranging axioms into a logical proof, as he did when he trained the Logic Theorist to mechanically solve the problems in Chapter Two of Russell and Whitehead's *Principia Mathematica*.

Because the lines of symbolic transformation would multiple exponentially, like a tree trunk endlessly branching as it grew, Simon's programs needed to evaluate the likelihood of any branch's success, prior to exploration. The key evaluating feature of the Logic Theorist, and of the later GPS, was its "heuristics," the equations by which it estimated, roughly, the "distance" or "difference" between a state and the goal state, characterized the nature of the difference, and if promising, selected a transformation strategy deemed appropriate to a difference of that kind.

The General Problem Solver was Simon's most ambitious work, a program which he sometimes called an intelligence, and at other times a Model of an intelligence, a simulation of an intelligence, or an embodied theory of the mind (that is, the program itself *was* the theory and its test all in one). It was general to the extent that it followed the general pattern of problem solving outlined above, carrying it from task to task. The only addition required for tackling a new kind of task, Simon hoped, was a heuristic unique to that task domain, whether it be playing chess or solving geometric puzzles. Indeed, Simon became almost notorious for his optimism, repeatedly predicting that a computer would become a champion chess player, an original mathematician, a competent translator, perhaps even a great scientist, all within a decade or less.[12] All that stood in the way was a sufficiently robust stable of heuristics.

***

In a paper written for the School of Industrial Administration, the first time he tackled the Automation issue directly, Simon considered the question, "would the corporation be managed by machines?"[13] The title alone indicated the radical forms of automation Simon had already begun considering, well outside the familiar categories that structured the Automation Debate. In most instances, the Pollyannas of automation had presumed that machines would replace only the dreariest drudgery, like assembly line work and the rote calculation of figures. By contrast, the most likely move for critics to make, whether scholars or labor leaders, was to recognize that automation seemed to have no immediate "skill bias"—that it might automate men out of good jobs, too. Simon's vision went well beyond them—not just to a few "skill" positions, but through the entirety of management. The corporation of the future would be "a highly-automated man-machine system" by 1985, by which time society would have "the technical capability of substituting machines for any and all human functions."[14]

In instance after instance, Simon emphasized the most rather than the least radical vision of future production. Mechanics?—"The genuinely automated factory—the workerless factory that can produce output and also…maintain and repair itself—will be

---

[12] The ambition of Simon's predictions were so infamous (and their failures so conspicuous) that he was used as a ridiculous foil by the philosopher and A.I. skeptic Hubert Dreyfus, in his *What Computers Can't Do*.

[13] Simon, *The Shape of Automation*, 26.

[14] Simon, *The Shape of Automation*, 27, 30.

technically feasible."[15] Psychiatry?—He imagined how one psychiatrist could do the work of ten.[16] Management?—with fewer employees, management too would become redundant.[17] "A physician, a corporate vice-president…a college teacher"—all could in principle be automated, whatever their status or desirability.[18] In a few cases, he inflated his estimates, the better to illustrate the power of his economic principles. But on the whole, the collected effect of the performance was to emphasize the almost limitless potential to duplicate man-the-worker, "a pair of eyes and ears, a brain, a pair of hands, a pair of legs, and some muscles for applying force."[19]

"Duplicating the problem-solving and information-handling capabilities of the brain is not far off; it would be surprising if it were not possible within the next decade."[20] In the light of such predictions, what hope was there for the flesh and blood man? In his two most widely published accounts of the problem, Simon said there was reason for hope, even for optimism.

In "Will the Corporation be Managed by a Machine," and his later entry, "The Long Range Economic Effects of Automation," Simon constructed a highly numerate foundation for correct prognostication and forecasting, one which was rhetorically structured as a master discourse over the effects of his own startling imagery. To predict long-run equilibrium, Simon explained, one had to distinguish between the variables—here his own machines—and the invariables—the economic situation that would presumably hold into the future.[21] In his first, less impressive run, Simon simply presumed long-run full employment, projecting to a future in which whatever displacement automation may yet come is already a thing of the past—an essay written in the future perfect.[22] The best justifications he offers for this view rest in the application of Ricardo's Law of Comparative Advantage. Machines, Simon reasons, may soon be superior to man in all respects, but they will be *comparatively* superior in only some domains. If a computer is a thousand times the calculator man is, but only one hundred times the stenographer, expect to see more human stenographers in the future.[23]

"Long Range Effects," the second and more gaming exploration, begins with a disquieting thought: The horse, long the finest employee in Western Europe, disappeared without a trace as a result of the tractor. Could the same happen to human workers? In a long and somewhat tedious construction of a hypothetical single-technology economy, that of the "beanbrick," Simon models the effects of rapid increases in the effectiveness of beanbrick production, the availability of beanbricks as capital, and the impact upon labor income as a share of national wealth. His takeaway, that assuming no excessive rents in land, labor must necessarily take the majority of any increased technological productivity, at least managed to address the issues in the Automation debate without

---

[15] Simon, *The Shape of Automation*, 34.
[16] Simon, *The Shape of Automation*, 36.
[17] Simon, *The Shape of Automation*, 43.
[18] Simon, *The Shape of Automation*, 42.
[19] Simon, *The Shape of Automation*, 37.
[20] Simon, *The Shape of Automation*, 39.
[21] Simon, *The Shape of Automation*, 29.
[22] This was exactly the sort of "long run" which is the butt of Keynes's famous joke.
[23] Simon, *The Shape of Automation*, 33.

hand waving them away, even if it was remarkably ahistorical about the kind of technology at stake, no doubt by design.

***

We needn't belabor the extent to which Simon's explicit arguments did, and did not, attend to the complaints made by organized labor and its representatives. Suffice to say that in reducing the power of industrial and intellectual automation to previous technical change or to technological change as such—that is, in construing "technical change" as an old, abstract, and continuous process, rather than one which was historically specific and economically episodic, he tended to perform economic argumentation in exactly the style that was of least help to workers on the ground, in specific industries and a specific moment, many of whom needed the rhetoric of change and discontinuity to have their complaints heard.

Perhaps of more interest, though, is the implicit style of Simon's argumentative moves.

Not unlike David McDonald, Simon understood the rhetorical value of fear, shock, the construction of historical epochs, and the encounter with technological estranging. In his own public pronouncements, even during his discussion of the Automation question, he took a puckish relish in disclaiming the uniqueness of man, a very personal kind of iconoclasm.

Nevertheless, the force of Simon's writerly performance came from the contrast he created between the image of the machine-as-maker, and the power of economics-as-predictor. He constructed, as any writer might (but as the father of a logistic conception of intelligence must), a hierarchy of explanation, for which the power of imagery was merely emotive. Having conjured startling images of robot workers, equal to man in every way, he explained (but would not describe) that all would be well, his equations and his numeric models proving it to be so. Throughout, the supremacy of numerical and formal modeling over the hazy day-terrors of the image was assured, even if temporarily suspended for effect.

It is no great surprise that Simon would hold to such a hierarchy. Most scientists would, and given his Behavioralist training, the explication of the economic system must necessarily proceed from history and image to model and form. Even after he had left political science and economics behind as something of a renegade and internal critic, the idea of science as fundamentally anti-historical and descriptive remained.

Perhaps more unique to Simon, though, was a tension, however submerged, in terms of the kind of revolution, the kind of history, we could expect Artificial Intelligence to make. For Simon, McDonald and Labor had misunderstood the meaning of automation, and especially of Artificial Intelligence. If it was to be a revolution, it would be a revolution in our capacity to *understand*. Artificial Intelligence was to give us a new grasp on the nature of man. To vaguely traffic in the idea of a new man and a new historical epoch, as McDonald did, was to suggest a social revolution distinctly at odds and in tension with the rapid expansion of scientific understanding and control. Artificial Intelligence was supposed to explain man and manage his world—to posit that on the eve of this *conceptual* revolution, man or his world were to be qualitatively transformed—that was a revolution too far.

What, in the end, should we make of Herbert Simon's contribution to the Automation Debate? I think it is useful to consider his essays a specific sort of genre, not mere policy papers, but not quite science fiction, either. Simon's essays, his startling promise of progress, and in this case, his sober recalibrations, all these were opportunities to draw before the mind's eye something quite spectacular—an image of the world to come. And in constructing such a fantasy, it was oftentimes satisfying to encounter specific resistances, the better to give their ideas shape, and to avoid the nagging sense that all this dreaming is idle work or Utopian fiction. In fact, like all historically influential utopias, Simon's ideas needed to specifically engage with a literary representative of reality, in this case signaled by the concern with timeless economic law, the better to produce a sense of their victory and plausibility.

Artificial Intelligence labs and the literary works they produced were a space for such dreaming. Their images of a future world were presented as anticipations, guides, and models. In retrospect they may have been consolation for a future which, even as it was announced, was slipping away. Regardless, the historicizing of this space of dreaming is one goal of my work.