Personal Recollections from 15 years of Monthly Meetings

[to appear in Essays in Honor of Herbert Simon,
eds Elisabeth Augier and James G. March, MIT Press, 2002]

Raul E. Valdes-Perez

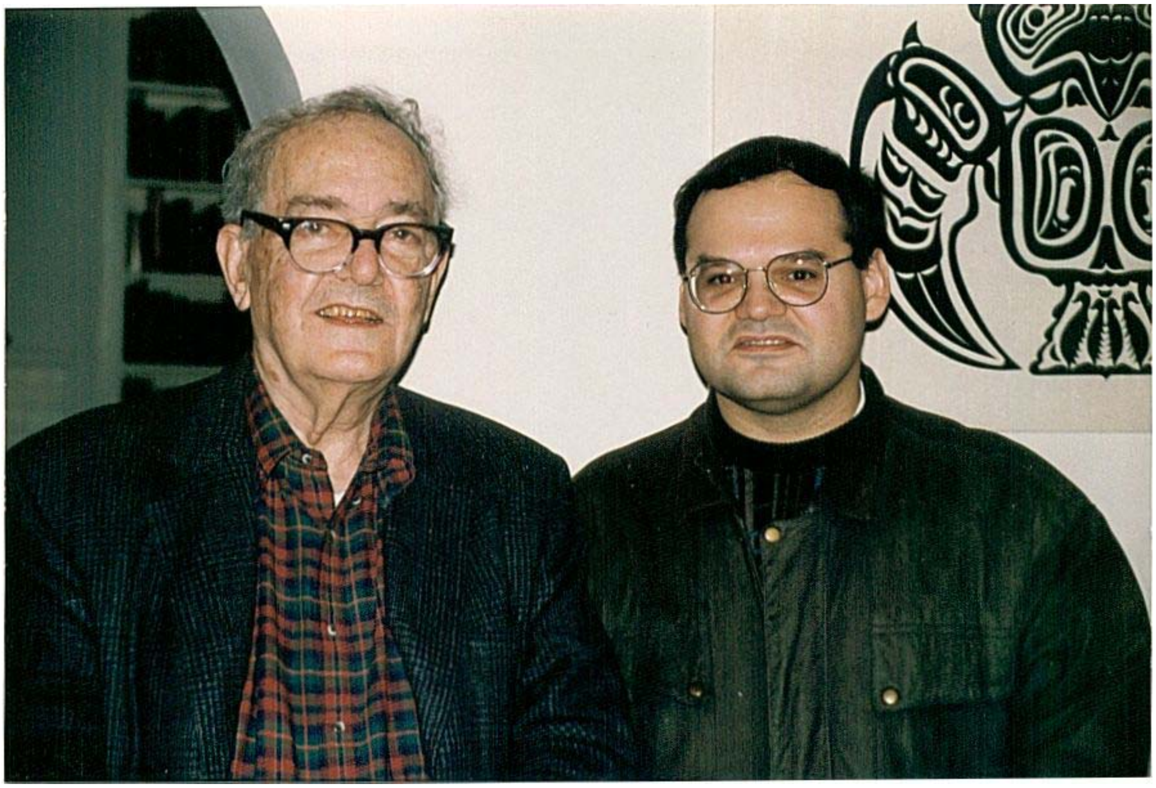
I had the personal fortune to know Professor Herbert Simon since1986, meeting with him monthly in my role as his computer science Ph.D. student, and since 1991 as a faculty member of the computer science department. After my graduation, I asked him whether he would accept continuing our meetings to discuss research and science in general, and he graciously agreed. We met around twice monthly for an hour in his Psychology department office over fifteen years. Hence, I have both good data and a good internal model of what Simon believed in science, although my recollections below should fairly be judged susceptible to my own biases and imperfect memory.

I will organize my personal observations around notes that I took during a lecture by Simon on the nature of research as part of Carnegie Mellon's computer science department Immigration Course in 1986, which introduced new doctoral students to faculty research projects as well as to broader issues, such as addressed by Simon's lectures. Simon gave this Immigration lecture nearly every year, and I attended many more of these over the following years. The 1986 version made the most impression on me, being a new student.

By no means do I imply from these notes that Simon, whom I could never bring myself to call "Herb" although most people did, believed that there is only one way of doing science. At least at the stage of his research career when I knew him, he was anything but dogmatic, since he always allowed for different research styles and methods. Consistent with this liberalness, a favorite criticism of his was the casual way in which using the English article "the" implied a unique property, as in "the key creative discovery". Simon liked to stress that scientific discoveries always involved different creative stages often carried out by different people: formulating the problem, finding the right approach, actually solving the problem, placing it within the context of existing knowledge, explicating the relationship to other fields, and so on.

1. A research problem is a question, not a topic.

Ask graduate students what they are working on, and most will reply with a topic like Artificial Intelligence, Human Memory, or Bayesian Networks. Probing further will get you sub-topics, and probing still deeper will extract tasks like a faster algorithm, a less memory-intensive one, data on human performance during recall, and the like. Instead, Simon emphasized formulating a question, partly I suppose because one can evaluate a question from many angles without actually doing any work on it. If the question is rejected, for example, by means of some of the tests in the next items below, then one saves much effort that is better expended elsewhere.



Another benefit of seeking a question is that one can, according to one's taste, emphasize finding a novel question. It is easier to make rapid progress when the question is new, because one doesn't have to improve on the many previous attempts to answer a question that is already well established. Of course, thinking about questions requires the student to have a good big picture of his or her discipline, in terms of what questions are of interest to the discipline or to science in general. This checklist provides some help in evaluating scientific questions.

One of my favorite examples of a once-novel question that was plainly available for anyone to formulate is: Why are the male testes always outside the body cavity in mammals? By finding an interesting, tractable question that nobody else has thought of, even modest progress will be a solid contribution.

In Simon's view, questions didn't need to be very specific. He was fond of examples like "What happens when you configure the world in this novel way and provide such-and-such stimulus?" which, according to him, fit the discovery of Ohm's law. One of Simon's earliest questions, from his own doctoral research into the processes of decision making, was: "How do playground administrators allocate their budgets?" The empirical observations that led him to his celebrated theory of bounded rationality began there, according to his autobiography.

2. **What would an answer to the question look like? How would you begin work on the problem?**

Simon's favorite example here was always a gravity shield. How to construct one is surely an interesting question, but if you cannot conceive of what form the shield will take, then you can scarcely figure out how to begin work on the problem.

From Simon's own research, another example is: Can one develop a cognitive model that would simulate, in humanly plausible and inspectable steps that are faithful to the historical facts, a celebrated discovery from the history of science? Answers to this question, as Simon and his students/collaborators Langley, Bradshaw, and Zytkow showed in their 1987 book Scientific Discovery, take the form of a computer program whose operation is consistent with knowledge in cognitive science, that explains the available historical circumstances about the discovery, and does not pre-suppose knowledge or data on the part of the discoverer that he or she did not possess.

3. **What test will determine whether the question was answered?**

I don't remember specific examples, but from other conversations with Simon, I believe that the many research articles with a title like "A Framework for …" suffer from this problem: what test would show whether you have the right framework? My search within the INSPEC literature database (computer science, physics, electrical engineering) for titles containing the word "framework" turned up 11,785 articles, so there is no shortage of them.

As an aside, Simon had a general suspicion of articles that begin "Towards …", presumably because he believed that authors should publish their conclusions when they arrive somewhere interesting, not when they've identified a direction and ventured a step or two. (INSPEC turns up 12,644 *Towards* articles.)

4. **Would anybody care whether I solved the problem? (e.g., would at least two people write you for reprints?)**

Working on problems that nobody cares about, no matter what creativity is required to solve them, is risky. Of course, the importance of the problem and its solution could emerge later, but probably most students do not have the well-formed scientific taste to pick the eventual winners. Simon doubtless intended this point and the others as heuristics that pointed the way to fruitful discovery, but which could be violated successfully.

I am reminded of another favorite theme of his: that the advice contained in popular proverbs often come in contradictory pairs, which shows their heuristic nature. For example, *Haste makes waste* and *Don't put off for tomorrow what you can do today*. Context is important.

5. **Does the present state of the art make the research feasible?**

I don't remember Simon supplying specific examples here, but I think he would agree that research in theoretical computer science that intended to prove that P equals NP (or its denial) was not very feasible, since there possibly was not enough theoretical knowledge available to make the problem soluble soon. An example closer to his own research would be attempts to show specific workings of the brain when the instruments to make the needed measurements have not been invented.

6. **If the field is crowded, what is my secret weapon that makes me think I can solve the problem where others have failed?**

I remember being surprised when hearing this, since it seemed to stress a competitive aspect to scientific research which, given my first-year student's idealism, seemed misplaced. However, failure and me-too success aren't much rewarded in science, so it's hard to dispute the heuristic wisdom of the advice. Simon pointed out then, and always, that "I'm smarter than other people" wasn't an acceptable answer, since it's rarely true. Much better answers could involve having access to unique apparatus that other people lack or, of course, possessing a new idea.

7. **Working on several tasks at once is good in order to avoid ruts.**

This was in reply to a question (mine) from the audience, about how to avoid getting stalled in one's research. I don't know whether Simon really intended this advice to apply to graduate students, who after their initial coursework are usually absorbed by one project, or to senior researchers who have the luxury of developing several simultaneous projects.

There is a relation between this advice and Simon's own research into the cognitive processes that underlie scientific discovery. As is widely known, Simon hypothesized that scientific discovery, as practiced by human beings, makes use of the same mental capacities that ordinarily problem-solving does, whether it involves doing crossword puzzles, solving exam problems, or inventing new cooking recipes. In fact, Simon argued that, if nothing else, assuming a same mental faculty was the right hypothesis for reasons of parsimony or Occam's Razor, i.e., of not multiplying entities beyond what are needed to explain the facts. In his view, there were no facts that proved that the mental discovery processes of great scientists were qualitatively different than the everyday mental processes of less celebrated people doing less exalted problem-solving tasks.

Like any theory, Simon's hypothesis that scientific discovery *was* problem-solving needed to be checked against the facts. In his view, the only significant hard fact about scientific discovery was that discoverers often reported that the insights that led to a major discovery happened suddenly while attending to something else, the so-called Eureka sensation or effect. So how could Simon's problem-solving hypothesis account for the Eureka effect?

Simon's explanation began by noting that discoveries came only after working on a problem for some time. His idea was that problem-solving (also known as heuristic search in problem spaces) leads the researcher in different directions, all of which are unsuccessful if the problem is yet to be solved. When the researcher puts the problem aside for a while and later starts, he begins in a new direction while leveraging all the clues and undertanding of the problem gained so far. Thus, the new direction of reasoning can lead quickly to a definitive insight and a Eureka sensation.

Simon would not have claimed that this explanation of the Eureka effect is "true". In fact, it is difficult to test. Rather, it simply fulfilled the need for a hypothesis, scientific discovery as problem-solving in this case, to admit a plausible explanation of whatever hard facts are available about the studied phenomenon, here, scientific discovery.

So, Simon's own research was consistent with this advice #7: work hard on a problem, but practice putting it aside as well.

8. **A Ph.D. thesis is a "progress report", that reports progress you have made on a problem.**

A consequence of this dictum is that research topics that are of an all-or-nothing character, i.e., the student either wins a grand prize or has nothing to show for it (except perhaps a consolation Master's degree) are very risky. I think that science students in longer-established fields understand this better than computer science or artificial intelligence graduate students, many of whom would set up very ambitious but practically insoluble research problems. Put another way by P.B. Medawar: science is the art of the soluble.

9. **Outstanding discoveries happen by luck, but only to prepared minds and with sufficient effort.**

This was not an original remark, but worth repeating to aspiring researchers. Among Simon's favorite examples was the discovery of Penicillin by Fleming, who famously noticed dead bacteria where living ones were to be expected. A less-prepared mind might have ignored the surprise and thrown out the Petri dish, but Fleming persevered, leading to his discovery of the anti-bacterial properties of penicillin.

This dictum hit home with great personal clarity in the context of a modest discovery I made some years later. I had been collaborating with a Bulgarian linguist Vladimir Pericliev, whom I had met at a Stanford symposium on scientific discovery that Simon and I had organized in 1995, on the problem of reconstructing, in computational terms, the logic underlying a discovery process that was popular in linguistics research of the 1950s and 60s (so-called componential analysis). In the componential analysis of the kinship terms used by a human language, the linguist would try to formulate concise descriptions of the kinship terms (e.g., English uncle, cousin, sister-in-law, etc.) that would demarcate each from every other term, using the attributes of sex, birth, marriage, and other relations between the kin and the speaker. This concern with conciseness, understandability, and demarcation led to the following discovery.

In preparation for a trip to South Florida, I was casually browsing the web pages of the University of Miami and came across this statement: "The University of Miami is the youngest of 24 private research universities in the country that operate both law and medical schools." At once it occurred to me that, given a database of peers – such as universities and their attributes, a computer program could be written to generate such *niche* statements, which expressed concisely and in grammatical English how a single, chosen database entry was interestingly *unique*. This task turned out to be a new data-mining question to ask about a database, which seemingly has never been approached in any systematic way, despite the plethora of niche statements that are appear everywhere, at least in American culture, not excepting scientific culture. Time will tell how important this new data-mining question is, but there are promising applications of the idea in science, such as in genomic data analysis – in what interesting way is a specific gene unique?

Following Simon's dictum, this writer could have seen the University of Miami statement two years earlier and not given it a further thought.

10. To stimulate the imagination, arm yourself with knowledge from many fields, in order to approach the task from different angles.

Simon made a practice of having lunch at an open table with faculty members from other university departments. I recall him saying that a favorite conversational tactic was to ask "What's new in your field?" If his lunch-mate reported a new topic of interest in another field, in which Simon had expertise to bring, he could read the pertinent literature, formulate his own ideas, and submit them for publication, whether the field was something he had ever published in or not. In his last years, Simon began to take an interest in theoretical, qualitative questions in developmental biology and how to approach them from an evolutionary viewpoint that his previous work in other fields (e.g., economics) prepared him to answer. I did some programming for him in Mathematica on simulations of his ideas, but lack of time prevented my supporting him much in his efforts. Of course, Simon did not rely only on lunch-mates for new ideas.

Not everyone had the breadth of knowledge and confidence to do research in fields with any previous track record. The advice above refers to the other direction, in which knowledge of techniques from other fields could be newly imported into one's own discipline.

11. If he had his druthers, the department would accept no graduate students with undergraduate degrees in computer science.

Simon felt very strongly that computer science departments erred by accepting mostly students with little backgrounds in anything besides computer science. He did not change this opinion in later years, since we talked about it often. He believed that students with little outside knowledge to draw on would tend to follow the research steps already laid out by others. Also, interdisciplinary research, which Simon of course championed, would be stunted, since few such students would seek outward-looking opportunities. Needless to say, many other faculty disagreed with him. Today almost all computer science graduate students, at least at Carnegie Mellon, have almost entirely computer science backgrounds.

I end my remarks with several final anecdotes about conversations that greatly helped me understand how Simon thought about research.

Simon introduced the idea of *satisficing* in Artificial Intelligence: people solve everyday problems by spending some effort gathering information followed by making a decision that leads to a satisfactory outcome. In his autobiography <u>Models of My Life</u>, he tells about his field observation of how playground administrators solve the problem of deciding what to spend their budgets on. Contemporary thinking was that they found an optimal solution by considering their constraints and resources and optimizing their choices. As a student, I asked Simon whether one couldn't construe their decision-making processes in terms of optimizing, within which their time, probability of finding information, and other factors could be included in the optimization criteria. Simon replied that one could always conceive any problem as an optimization problem, but thinking in that way would lead the researcher to a very different path, such as estimating unknowable probabilities, quantifying the space of different purchasing decisions (which are boundless), and so on. The researcher who instead viewed the decision problem as one of *satisficing* would try to identify the heuristics and information-gathering steps used in practice, perhaps try to find better heuristics if his goal was to improve human decision making, and so on. Very different research would ensue. Simply put: conceiving of the problem in different ways led to different research, and Simon believed that the more fruitful research would be based on the *satisficing* interpretation. The question of whether decision makers *truly* optimized or not was problematic and not to pursued scientifically, since what test would answer it (see #2 above)?

As a student, I read many of Simon's early papers, chosen somewhat at random, simply because I was intrigued by the man and his work. In the early 50's he wrote an article that used the formalism of differential equations to model aspects of social interaction. Upon finishing the article, it struck me as excessively conjectural, hard to test, and generally uncharacteristic of the Simon that I knew. At my next meeting I asked why he had written such an article. He replied that his goal had always been to introduce more rigor into the social sciences, and in the early 50's differential equations were available to him, but not computers. When he came into contact with computers, Simon quickly recognized them as the formal tool he was seeking, since computers were a highly flexible experimental instrument that he could use to model intelligence or other social science phenomena.

I wrote only a few minor, short papers with Simon, since my own research on scientific discovery did not attempt to model human processes in any way, but to improve on them using computers, which was of interest to Simon but was not something he did himself.

But once when we were discussing how to conclude an editors' introduction for a special issue of a journal, Simon asked: "What would Cicero write?" I never did figure out precisely what Professor Simon meant by that, since he just wrote the conclusion himself, and quite well, as always.