The Subtlety of Sameness

A Theory and Computer Model of Analogy-Making

Robert M. French

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

Foreword

If somebody asked me to design a book that would introduce the most important ideas in artificial intelligence (AI) to a wider audience, I would try to work to the following principles:

1. Go for details. Instead of presenting yet another impressionistic overview of the field, concentrate on the details of a particular AI model, so that the readers can see for themselves just how and why it works, seeing its weaknesses and boundaries as well as its showcase triumphs.
2. Model something we all know intimately. Choose a psychological phenomenon that is familiar to everyone—and intrinsically interesting. Not everybody plays chess or solves route-minimization problems, and although we almost all can see, unless we are vision scientists, we have scant direct familiarity with the details of how our visual processes work.
3. Explain exactly how the particular model supports or refutes, supplements or clarifies the other research on the same phenomenon, including work by people in other disciplines.
4. Give concrete illustrations of the important ideas at work. A single well-developed example of a concept applied is often better than ten pages of definition.

This book by Bob French all about his Tabletop model fills the bill perfectly, so when I read an early draft of it (I was a member of his Ph.D. dissertation committee), I encouraged him to publish it and offered to write a foreword. From its easily read pages, you will come to know the model inside out, not only seeing *that* it comes up with recognizably human performance but seeing—really seeing—*how* it comes up with its results. And what does it do? It does something we all do every day: it appreciates analogies. It creates them and perceives them, in a manner of speaking. The simple setting of the task is inspired: a game of "Do this!" whose point you will get in an instant, but whose richer possibilities are not only surprising, but quite inexhaustible.

You get to tackle all the problems yourself and think about them "from the first-person point of view." Something goes on in you when you do these problems. What on earth is it? It seems at first the farthest thing from mechanizable—"intuitive," quirky, fluid, aesthetic, quintessentially human—just the sort of phenomenon that the skeptics would be tempted to brandish, saying, "You'll never get a computer to do *this*!" Or, more cautiously, "You'll never get a computer to do this the way *we* do it!" If you are such a skeptic, you are in for a surprise.

Most AI programs model phenomena that are either highly intellectualized thinking exercises in the first place—like playing chess or constructing proofs—or else low-level processes that are quite beneath our ken—like extracting three-dimensional information about the visible world from binocular overlap, texture gradients, and shading. French's program, in contrast, models a phenomenon that is neither difficult nor utterly invisible but rather *just* out of reach to the introspecting reader. We can *almost* analyze our own direct experiences into the steps that French's model exhibits. AI workers love acronyms, and I hereby introduce the term AIGLES—almost-introspectible, grain-level events—as the general term for the sort of high-level psychological phenomenon French has modeled. If there *were* a Cartesian Theater in our brains, across whose stage the parade of consciousness marched, his would be a model of something that happens immediately backstage. (Those who join me in renouncing the all-too-popular image of the Cartesian Theater have a nontrivial task of explaining why and how French's model can avoid falling into that forlorn trap, but this is not the time or place for me to discharge that burden. It is left as an exercise for the reader.)

From the particulars, we can appreciate the general. French introduces, exemplifies, and explains some of the most important and ill-understood ideas in current AI. For instance, almost everybody these days speaks dismissively of the bad old days in AI and talks instead about "emergence," while waving hands about self-organizing systems that settle into coherent structures and so forth. (I myself have spoken of multiple drafts in competition, out of which transient winners emerge, a tantalizingly metaphorical description of the processes I claim are involved in human consciousness.) French provides a no-nonsense model of just such a system. When posed a problem, the answer it arrives at is "the emergent result of the interaction of many parallel unconscious processes" (p. 20). So here is a fine place to see what all the hoopla is about. You get to see how the currently popular metaphors—a batch of cooling molecules or a system of interconnected resonators coming to vibrate in unison, for instance—apply to an actual nonmetaphorical reality, a system engaged in doing some undeniably mental work.

His model also illustrates a version of "dynamic" memory structures, which deform to fit the current context, and it achieves its results by exploiting a "parallel-terraced scan." It accomplishes "implicit pruning," which must somehow be what we manage to do when we ignore the irrelevancies that always surround us. It does this by building (and rebuilding and rebuilding) the relevant structures on the fly, thereby avoiding at least some of the "combinatorial explosions" that threaten all AI models that have to ignore most of what they *could* attend to without catastrophically ignoring the important points. The central theme of the book is that the processes of producing mental representations and manipulating them are inextricably intertwined. As French puts it, "You *must* take the representation problem into consideration *while* you are doing processing." When you understand this paragraph in detail (and you will when you have read the book), you will have a good grip on some of the central ideas in recent AI.

These ideas are not just French's of course. His work grows out of the family of projects undertaken in recent years by Douglas Hofstadter's group at Indiana University and, as such, provides a fine demonstration of the powers of that school of thought in AI. Many skeptics and critics of AI from other disciplines have surmised there was something profoundly wrong about the hard-edged, inert (but manipulable) symbols of the "physical-symbol systems" of traditional AI, and hence they have been intrigued by Hofstadter's radical alternative: "active symbols." Active symbols sound great, but what are they, and how on earth could they work? This book takes us a few sure steps toward the answer. French provides a judicious comparison of his own work—which has plenty of its own originality—to that of others who have worked on analogy, in Hofstadter's group, in AI more generally, and in psychology.

If you don't already appreciate it, you will come to appreciate the curious combination of ambition and modesty that marks most work in AI and the work of Hofstadter and his colleagues in particular. On the one hand, the models are tremendously abstract, not tied at all to brain architecture or to the known details of such processes as "early vision." All the important questions in these research domains are simply sidestepped. That's modesty. On the other hand, the models purport to be getting at something truly fundamental in the underlying structure and rationale of the actual processes that must go on in the brain. That's ambition. Like more traditional AI programs, they often achieve their triumphs by heroic simplification: helping themselves to ruthless—even comical—truncations of the phenomena (more modesty), in order, it is claimed, to provide a feasible working model of the essential underlying process (more ambition). The reader is left, quite properly, with an unanswered question about just which helpings of simplification might be poisoned. Are any of these bold decisions fatal oversimplifications that could not possibly be removed without undoing the ambitious claims? There is something that is both right and deep about this model, I am sure, but saying just what it is and how it will map onto lower-level models of brain function is still beyond me and everybody else at this time, a tantalizing patch of fog.

French's program doesn't learn at all—except what it could be said to learn in the course of tackling a single problem. It has no long-term memory of its activities, and it never gets any better. This might seem to be a horrible shortcoming, but it has an unusual bonus: his program never gets bored! You can give it the same problem over and over and over, and it never rebels but always takes it in a fresh spirit. This is excellent for "rewinding the tape"—looking, counterfactually, at what *else* a system might do if put in the same situation again. Heraclitus said that you can never step into the same river twice, and this is particularly true of human beings, thanks to our memories. Aside from a few famous amnesiacs, we normal human beings are never remotely in the same state twice, and this seriously impedes scientific research on human cognitive mechanisms. Is investigating a system with total amnesia, like French's, a worthy substitute for non-doable human experiments, or does the absence of memory and learning vitiate his model? French shows that AI fell into a trap when it opportunistically separated the building of representations from their processing; will some meta-French soon come along to show that he fell into just as bad a trap by setting long-term learning aside for the time being? A good question—which is to say that no one should think that the pessimistic answer is obvious. In the end, at some level, no doubt just about everything is inextricably intertwined with everything else, but if we are to understand the main features of this tangled bank, we must force some temporary separation on them.

A standard conflict in AI is between the hard edges and the fuzzies, a conflict fought on many battlefields, and some of the niftiest features of French's model demonstrate what happens when you slide back and forth between hard edges and fuzzy edges. There are knobs, in effect, that you can turn, thereby setting parameters on the model to give you nice sharp edges or terribly fuzzy edges or something in between. Probability plays a deep role in French's model, which makes it imperative for him to test his model in action many, many times and gather statistics on its performance—something that would not be at all motivated in most traditional AI. But if you set the model so that some of the probabilities are very close to 1 or 0, you can turn it into what amounts to a deterministic, hard-edged model. Or you can explore the trade-off between depth-first search and breadth-first search, by adjusting the "rho" factor, or you can create what French calls a semi-stack, another fuzzy version of a hard-edged idea. This is a particularly attractive set of features, for one of the things we know about ourselves is that the *appearance* of determinism and indeterminism in our mental life is highly variable.

AI is a seductive field. Even a book as scrupulously written as French's may mislead you into ignoring deep problems or deficiencies in the model, or—a very common foible—it may encourage you to overestimate the actual fidelity or power of the model. Here, for the benefit of neophytes, are a few of the tough questions you should keep asking yourself as you read. (You will get a better sense of the genuine strengths of French's model by making sure you know just what its weaknesses are.)

French claims his domain, the (apparently) concrete world of Tabletop, is rich enough to "ground" the symbols of his model in a way that the symbols of most AI programs are not grounded. Is this really so? We·viewers of Tabletop *see* the knives, forks, spoons, cups, bowls, and so forth, vividly laid out in space, but what does the model really understand about the shape of these objects? Anything? Does Tabletop know that a spoon, with its concavity, is more like a bowl than a knife is? *We* can see that a spoon is a sort of bowl on a stick, but that is utterly unknown to Tabletop. What other sorts of obvious facts about tableware are left out of Tabletop's semantics, and how could they be added? Perhaps it is here that we see most clearly what the model leaves out when it leaves out learning—in the real world of concrete experience. But what difference, if any, does this make to the groundedness of Tabletop's symbols? Is there some other sense in which Tabletop is clearly superior in "groundedness" to other programs? (I think the answer is yes. Can you see how?)

Tabletop gets its basic perceptual accomplishments for free. It cannot mistake a knife for a fork out of the corner of its eye or fail to see the second spoon *as* a spoon (if it ever directs its attention to it). Everything placed on the table is, as it were, legibly and correctly labeled according to its type. So what? (Might some of the combinatorial explosions so deftly avoided by Tabletop come back to haunt it if this gift were revoked? Again, so what?)

What would it take to add learning to Tabletop? What would it take to expand the domain to other topics? What would happen if you tried to add episodic memory? Could you readily embed Tabletop in a larger system that could face the decision of whether or not to play the Tabletop game, or play it in good faith? (A human player could get fed up and start giving deliberately "bad" answers to try to drive "Henry" into one amusing state of frustration or another. Is this a feature whose absence from Tabletop could be readily repaired, or would a model builder have to start over from scratch to include it?)

Finally, the granddaddy of all challenging questions for any AI program: Since this model purports to be right about something, how could we tell if it was wrong?

What sort of discovery, in particular, would refute it? The boring way of responding to this question is to try to concoct some philosophical argument to show why "in principle" Tabletop couldn't be right. The exciting ways are to be found down the paths leading from the other questions.

My raising of these challenging questions in the foreword is the most unignorable way I can think of to demonstrate my confidence in the strength and value of this book. Go ahead; give it your best shot.

Daniel Dennett