Marvin Minsky, The Society of Mind, Simon & Schuster (New York, 1985); 339 pages, $10.95 (paperback)

Artificial intelligence is fundamentally an engineering discipline; our collective goal of constructing an intelligent artifact is fundamentally an engineering one. Good engineering builds on good science. Good science, quite frequently, builds on good mathematics.

The Society of Mind does not pretend to be good engineering or good mathematics. My aim in this review is to discuss the question of whether or not it is good science.

I will do this by examining two separate but related issues. The first involves Minsky's abandonment of the usual methods of scientific inquiry and a discussion of the problems to which this leads; the second is an analysis of Minsky's criticisms of formal logic and its role in AI.

1 Falsifiable claims

Consider Newton's observation that the acceleration imparted to objects in a gravitational field is independent of their mass. If I can find two objects for which this rule fails, Newton is wrong and his theory needs to be revised. I do not need to seek him out and discuss the matter; he has no choice. The theory of gravitation allows us to make precise claims about observations to be made in the future, and we evaluate the theory by testing the validity of these claims.

In *The Society of Mind*, Minsky makes no such precise claims (about gravity or anything else). Instead, he builds up a loose framework in which the behavior of a single conscious entity is analyzed in terms of the collective behavior of a host of individual "agents" that comprise it. Minsky goes on to argue that this approach can address a wide variety of longstanding issues in philosophy and psychology – everything from autism to the fact that postpubescents can't learn to speak a new language without an accent.

How are we to evaluate views that are so vaguely formalized? The hallmark of good science is that it forces us to stick our necks out by requiring that we make falsifiable claims, but Minsky has no inclination to do this – indeed, the foundation from which he writes is too vague to allow it. Instead, the book rambles and ends up being a collection of interesting but fundamentally *ad hoc* speculations about how intelligence might work.

In fairness, I should point out that Minsky is aware that he is turning his back on conventional scientific methodology. In Essay 25.6, he writes that a good theory should not be specified in too much detail, because that encourages other scientists to " 'test' it, instead of contributing their own ideas."

Unfortunately this "testing" is the cornerstone of modern science. And what makes AI interesting as a discipline is that it allows us to apply the tools of science, engineering and mathematics to questions that have previously been in the realm of philosophy and psychology. In abandoning these tools, Minsky also abandons exactly those things that make our field unique and support our expectations that AI will make substantial contributions to these far-flung disciplines.

Again, I am saying nothing of which Minsky is unaware. In the postscript, he writes that, "This book should be read less as a text of scientific scholarship and more as an adventure story for the imagination." I suppose that Minsky's stature as a founder of our discipline sanctions such flights of speculation, but he should be more candid throughout about his intentions.

2 Formal logic

At the other end of the scientific spectrum from Minsky's work is the work by some researchers on applying formal mathematics to the understanding of intelligence. Most commonly, some generalization of first-order logic is used; the falsifiable claims made typically revolve around the question of whether some particular knowledge representation scheme can efficiently encode our understanding of some commonsense domain.

Minsky seems at special pains to dismiss these efforts. In Essay 24.5 he tells us, "What appears to be a matter of 'logic' is usually not logical at all and frequently turns out to be wrong." In Essay 27.4, "Logic fails because it tries to find exceptions to [the rule that] exceptions are a fact of life."

The problem is that Minsky is viewing "logic" far too narrowly. Witness his description in Essay 18.4:

> The difference [between logic and ordinary thinking] is that in logic there's no middle ground; a logic link is either there or not. Because of this, a logical argument cannot have any "weakest link."

One way to view the work on nonmonotonic reasoning is as discovering what default assumptions underlie a particular conclusion; as such, this work is concerned *exactly* with the problem of identifying the "weakest link" in a deductive argument. Minsky's 1985 copyright gives him no excuse to ignore this large body of material.

I will return to the role of nonmonotonic reasoning in a moment, but before doing so, let me point out that one of Minsky's other criticisms of logic is even further from the mark. In Essay 18.1, he says that logic "can serve as a test to keep us from coming to invalid conclusions, but it cannot tell us which ideas to generate, or which processes and memories to use. Logic [cannot] explain how we think ... " Underlying this suggestion is the idea that logic can be used to solve classification problems (is the conclusion valid or not?) but not synthesis problems (what plan would achieve the given goal?).

This is simply wrong; it is well known that logic can be used to solve planning problems (for example) by proving that there exists a plan to achieve a particular goal. If the proof is constructive (and most proofs found in logic programming are), then the details will contain a specification of the plan that should be used. This idea dates back to Green's work in 1969 [2]; more recently, Finger [1] and others have shown that synthesis problems can be solved using truth maintenance techniques as well.

But let me return to nonmonotonic reasoning and issues arising from it. When I look at the progress that has been made since the realization around 1980 that commonsense reasoning had nonmonotonic properties, I am struck not by how *much* progress has been made, but by how *little*. Nonmonotonic reasoning has turned out to be far more difficult than any of us expected. The fundamental advantage of working with formal methods – the advantage of making falsifiable claims – is that we are forced to address these difficulties, to confront the nasty surprises that lie between speculation and practice.

As an example, consider inheritance reasoning, which deals with things such as the fact that birds fly but ostriches don't. This example is due to Minsky himself, and he discusses it at various points in the book. The suggestions made aren't any different from the basic idea, well known in the nonmonotonic community, that subclasses should inherit by default the properties possessed by the superclasses that contain them. Thus canaries can fly because birds do, and ostriches are an exception to this sort of a rule.

The surprise is that inheritance reasoning is hard; this simple idea is not enough. In fact, inheritance reasoning is *very* hard – the nonmonotonic community has seen a flurry of recent papers concerned with finding a satisfactory formalization of it. The problem is that for any particular property (e.g., "rich") and object (e.g., a particular public-defence attorney), there is likely to be conflicting default information (lawyers tend to be rich but public servants tend not to be). The fact that Minsky has no answer to this problem doesn't trouble me (I don't have one either); what troubles me is that he manages to discuss inheritance without even mentioning the issue.

The need to select among competing default rules appears to be completely ubiquitous in commonsense reasoning. Most of the recent formal work on nonmonotonic methods, whether concerned with inheritance or not, has been aimed at addressing this issue. If this is where all the "action" is from a scientific point of view, how is it that Minsky leaves the topic completely unaddressed?

Just as the nonmonotonic work on conflicting defaults is not limited to inheritance reasoning, the problem also has a general analog in Minsky's setting. When different agents disagree about a particular issue, how is their difference to be settled? Perhaps Minsky would introduce new agents to mediate the conflict, but that completely misses the point. How, *exactly*, is this mediation to be accomplished? Without an answer to this, the concept of an agent is an idea with no practical value; it is so general that it might encompass anything. The science, the engineering, the *value* of ideas lie in specifications and the associated falsifiable claims that Minsky chooses not to make.

It is all well and good to invent interesting theories and speculate about their consequences, but such an approach almost inevitably leads to an examination of the strengths of these theories and not of their weaknesses – and Minsky's inability to choose between conflicting agents is indeed a weakness of his theory. A formal approach would have forced him to focus on this weakness until some resolution was obtained; his informality allows him to deal with it by sticking his head in the sand.

3 Concluding remarks

The Society of Mind is a fascinating and provocative book. But in abandoning the standard methods of scientific inquiry, Minsky has allowed himself to fall into the classic trap of speculative science, highlighting the strengths of his approach and ignoring its shortcomings. His failure to either discuss difficult issues or to make sharp claims concerning the applicability and limitations of his approach makes it difficult to evaluate the work; it is a sign of AI's maturation as a discipline that scientific validation is now expected.

Where Minsky does make sharp claims regarding the inapplicability of first-order logic to AI, he has taken too myopic a view. The shortcomings of the formal work on AI are apparent, but they are not as dire as Minsky suggests. In fact, the focus they provide to the formal community can be viewed as much as a strength as as a weakness.

All told, *The Society of Mind* is as Minsky describes it: an adventure in speculation, and not substantive science. It is well worth reading, but only in that light.

References

[1] J. J. Finger. *Exploiting Constraints in Design Synthesis.* PhD thesis, Stanford University, Stanford, CA, 1987.

[2] C. C. Green. Theorem proving by resolution as a basis for question-answering systems. In B. Meltzer and D. Mitchie, editors, *Machine Intelligence 4*, pages 183–205. American Elsevier, New York, 1969.