

---

## Preface

There are many reasons to work in the field of artificial intelligence (AI). My reason is a desire to solve the “mind-body” problem, to understand how it is that a purely physical entity, the brain, can have experiences. In spite of this long-range goal, my research has been concerned with seemingly much tinier questions, such as, how might a robot know where it is? How would a computer represent the sort of routine but arbitrary fact or belief that people seem to keep track of effortlessly? (I’m thinking of “facts” such as “If you go swimming too soon after eating, you might get a cramp.”) It may seem misguided to pursue tactical objectives that are so remote from the strategic objective, but scientists have learned that in the end finding precise answers to precise questions is a more reliable means to answering the big questions than simply speculating about what the big answers might be. Indeed, scientists who venture to attempt such speculation are often looked at askance, as if they had run out of useful things to do.

Hence, by writing a book on the mind-body problem from a computational perspective, I am risking raised eyebrows from my colleagues. I take that risk because I think the mind-body problem is important, not just technically, but culturally. There is a large and growing literature on novel approaches to the problem. Much of it is quite insightful, and some is totally wrong (in my opinion, of course). Even authors I agree with often fail to understand the role of computational ideas in explaining the mind. Claims like these are often made with only the flimsiest arguments:

- An intelligent computer program would treat every reasoning problem as a deduction.

- There are two computational paradigms to choose from: symbolic computing and neural networks; they are quite different, and have fundamentally different properties.
- People think serially at a conscious level, but are “massively parallel” inside; so it’s appropriate to model them with a program only when studying conscious problem solving.
- Whether something is a computer depends entirely on whether a person uses it as a computer.
- When a computer program manipulates symbols, the symbols must have a formal semantics, or the program will do nothing interesting.
- Whether the symbols in a computer mean anything depends entirely on whether people treat them as meaning something.
- A computer could be made to behave exactly like a person, but without experiencing anything.

I will show that all these claims are false, meaningless, or at least questionable.

If these misconceptions mean nothing to you, good; by reading this book first, you will avoid them. Unfortunately, if you’ve read one or two books on the subject of computation and the mind, you have probably absorbed some of the nontruisms on the list without even noticing it.

I often assume that the mind-body problem is interesting to everyone, but I have discovered, by watching eyes glaze over, that it isn’t. One reason is that it is surprisingly difficult to convey to people exactly what the problem is. Each of us has little trouble separating mental events from physical ones, and so we gravitate to a theory that there are two realms, the mental and the physical, that are connected somehow. As I explain in chapter 1, this kind of theory, called *dualism*, though it seems at first obviously true, runs into enough difficulties to move it to the “obviously false” column. This should get anyone interested, because many of us have religious beliefs—important religious beliefs—that presuppose dualism. Hence we have a stake in what theory ultimately replaces it.

Some parts of the book are a bit demanding technically. There is a little mathematics in chapter 2, a survey of the state of the art in artificial intelligence. Chapter 5 addresses the knotty technical issues surrounding the notion of symbols and semantics. I was tempted to leave all these hard

bits out, to keep from driving away a large class of intelligent readers who suffer from “mathematics anxiety” or “philosophy narcolepsy.” I decided to leave chapter 2 in to counteract the general tendency in surveys of AI to talk about what’s possible instead of what’s actually been accomplished. The problem with the former approach is that people have an odd series of reactions to the idea of artificial intelligence. Often their first reaction is doubt that such a thing is possible; but then they swing to the opposite extreme, and start believing that anything they can imagine doing can be automated. A description of how it might be possible to program a computer to carry on a conversation encourages this gullibility, by painting a vivid picture of what such a program would be like, without explaining that we are very far from having one. Hence, throughout the book, I try to differentiate what we know how to build from what we can imagine building.

I left chapter 5 in for a different reason. I think the most serious objections to a computational account of mind rest on the issue of the *observer-relativity* of symbols and semantics, the question of whether symbols can mean anything, or can even be symbols in the first place, unless human beings impute meanings to them. This may not seem like the most serious objection for many readers, and they can skip most of chapter 5. Readers who appreciate the objection will want to know how I answer it.

With these caveats in mind, let me invite you to enjoy the book. The puzzles that arise in connection with the mind-body problem are often entertaining, once you’ve wrapped your mind around them. They are also important. If people really can be explained as machines controlled by computational brains, what impact does that have on ethics or religion? Perhaps we can’t answer the question, but we should start asking it soon.