Getting started in cognitive science

Sam Gershman
Harvard University

Cognitive science is a great area to work in if (like me) you have a restless mind, always seeking to be at the edge of your comfort zone. It is almost unparalleled as an interdisciplinary nexus. Below, I have assembled some resources that might be useful for people getting started, with a bias towards computational approaches. Obviously, this list is completely subjective, and I wish to make no pretense about the universal usefulness/importance of these choices.

General / big picture

Vision, David Marr (1982). You only really need to read the first chapter, which is foundational.

What is Thought? Eric Baum (2004). A thought-provoking, mostly conceptual and non-technical tour of ideas about computation, cognition and evolution. It doesn't really get into much of the empirical literature.

The Adaptive Character of Thought, John Anderson (1990). A seminal work on "rational analysis" that explores the design principles underlying different aspects of cognition.

The Organization of Learning, Randy Gallistel (1990). Essential reading to understand the nature of animal cognition, particularly the sophisticated algorithms that exist even in apparently very simple insects.

Mind as Machine, Margaret Boden (2006). Massive and comprehensive history. Chomsky hated it. What other recommendation do you need?

Theoretical Neuroscience, Peter Dayan & Larry Abbott (2001). Still the best textbook on the topic.

Close-ups

How Children Learn the Meaning of Words, Paul Bloom (2000). If you think children learn words by simple association, read this book to disabuse yourself of the notion! It shows the sophisticated way in which object perception, social cognition, and learning work together to build language.

The Origin of Concepts, Susan Carey (2011). Wrestles with the big questions of cognitive development. See also *Words, Thoughts, and Theories,* Alison Gopnik & Andrew Meltzoff (1998), for a somewhat different take.

The Big Book of Concepts, Gregory Murphy (2004). Good starting point for learning about the cognitive science of concepts. To go deeper, check out the volume *Concepts: Core Readings* edited by Eric Margolis & Stephen Laurence (1999).

Bayesian Rationality, Mike Oaksford & Nick Chater (2007). Applies the rational analysis approach to the debate over the role of logic in human cognition.

Mental Leaps: Analogy in Creative Thought, Paul Thagard & Keith Holyoak (1994). Great introduction to the study of analogy in cognitive science.

Knowledge Representation, Arthur Markman (1998). Reviews most of the key ideas about representation in cognitive science.

Human Inference: Strategies and Shortcomings of Social Judgment, Richard Nisbett & Lee Ross (1980). An old but still relevant overview of heuristics and biases in judgment. For a more recent treatment oriented towards lay readers, see Daniel Kahneman's (2011) *Thinking, Fast and Slow.*

The Illusion of Conscious Will, Daniel Wegner (2003). An excellent, very readable treatment of why we think we control what we do.

Language and Experience: Evidence from the Blind Child, Barbara Landau & Lila Gleitman (1985). The remarkable thesis of this book is that blind children learn a semantically rich understanding of words (comparable to that of normally developing children) despite their impoverished perceptual input.

The Logic of Perception, Irvin Rock (1983). Argues that perception is more like problem-solving than signal detection or feature extraction. See also *Visual Intelligence*, Donald Hoffman (1998). If you're interested in a more computational and biologically oriented treatment, I recommend *Seeing: The Computational Approach to Biological Vision*, John Frisby & James Stone (2010).

Foundations of Human Memory, Michael Kahana (2012). Theory-oriented survey of human memory.

Math / statistics / computer science

Students often ask what math they need to know. The standard answers are calculus, linear algebra and probability theory. I'm of the opinion that you actually need to know very little about calculus and linear algebra. You need to know what derivatives and integrals are, and you need to know how to manipulate vectors and matrices. There are all sorts of identities and formulas that come in handy, but there's no point in memorizing them before you actually need to use them. There's no point learning how to solve integrals: most interesting integrals are intractable anyway! You'll end up using approximation techniques like Monte Carlo simulation. And there's no point spending a lot of time learning how to calculate derivatives (God knows I wasted years of my life doing this): use Mathematica or WolframAlpha. I find it useful to know the differentiation rules, but again don't memorize this stuff until you need to use it. The same point applies to probability theory. Understand the basics and then you can learn the rest incrementally as you need it. I don't think you need to take a class on any of these topics, because the key ideas can be summarized in a few pages. My sentiment was perfectly expressed in this 1910 textbook (*Calculus Made Easy*, by Silvanus Thompson):

> **CHAPTER I.**
>
> **TO DELIVER YOU FROM THE PRELIMINARY TERRORS.**
>
> THE preliminary terror, which chokes off most fifth-form boys from even attempting to learn how to calculate, can be abolished once for all by simply stating what is the meaning—in common-sense terms—of the two principal symbols that are used in calculating.
>
> These dreadful symbols are:
>
> (1) d which merely means "a little bit of."
>
> Thus dx means a little bit of x; or du means a little bit of u. Ordinary mathematicians think it more polite to say "an element of," instead of "a little bit of." Just as you please. But you will find that these little bits (or elements) may be considered to be indefinitely small.
>
> (2) $\int$ which is merely a long S, and may be called (if you like) "the sum of."
>
> Thus $\int dx$ means the sum of all the little bits of x; or $\int dt$ means the sum of all the little bits of t. Ordinary mathematicians call this symbol "the
>
> C.M.E. A
>
> 2 CALCULUS MADE EASY
>
> integral of." Now any fool can see that if x is considered as made up of a lot of little bits, each of which is called dx, if you add them all up together you get the sum of all the dx's, (which is the same thing as the whole of x). The word "integral" simply means "the whole." If you think of the duration of time for one hour, you may (if you like) think of it as cut up into 3600 little bits called seconds. The whole of the 3600 little bits added up together make one hour.
>
> When you see an expression that begins with this terrifying symbol, you will henceforth know that it is put there merely to give you instructions that you are now to perform the operation (if you can) of totalling up all the little bits that are indicated by the symbols that follow.
>
> That's all.

Okay, with my sententious preface out of the way, here are some book recommendations.

Probability Theory: The Logic of Science, E.T. Jaynes (2003). This is the place to start if you want to understand why probability is so fundamental to the sciences. Jaynes walks through how probability can be understood as an extension of logic. For an excellent historical introduction to this topic, see Ian Hacking's (1975) *The Emergence of Probability*.

Pattern Recognition and Machine Learning, Christopher Bishop (2006). I learned machine learning from this textbook. There are many other good textbooks, but this is the reference I've returned to many times.

Elements of Information Theory, Thomas Cover & Joy Thomas (1991). The standard textbook on information theory, wonderfully clear.

Information Theory, Inference, and Learning Algorithms, David MacKay (2003). Covers a lot of the same material as Bishop and Cover & Thomas, but also some other material not either of those books. MacKay also has a great fund of vivacity not found in the other books.

All of Statistics, Larry Wasserman (2004). Rigorous and concise introduction. Also check out the companion volume, *All of Nonparametric Statistics* if you're interested in that topic.

Reinforcement Learning: An Introduction, Richard Sutton & Andrew Barto (2018). If you want to learn about reinforcement learning, start here.

If you're in high school or college

Gödel, Escher, Bach, Douglas Hofstadter (1979). This is mind candy to enjoy before you know too much. Don't read it in grad school! I think this point applies to basically all of Hofstadter's books.

The Origin of Consciousness in the Breakdown of the Bicameral Mind, Julian Jaynes (1976). A completely insane, totally fascinating, and almost certainly wrong book. Don't read it in grad school!

Thinking broadly

As Sherlock Holmes quipped, "One's ideas must be as broad as Nature if they are to interpret Nature." I think this is true of cognitive science. You need to expand your mind by reading things that aren't obviously cognitive science.

The Structure of Scientific Revolutions, Thomas Kuhn (1962). This book totally reshaped my understanding of science as a whole. Kuhn was also very interested in Gestalt psychology, and likened paradigm shifts to multistable perception.

Against Method, Paul Feyerabend (1975). One wild ride of a book!

The Power of Images, David Freedberg (1989). A fascinating study of how works of art elicit emotional responses that seem to precede and overflow our rational/aesthetic responses.

Foraging Theory, David Stephens & John Krebs (1987). Classic introduction to optimal foraging theory, which has seen applications to risky choice, visual search, and other areas of cognition.

The Art of Memory, Frances Yates (1966). History of mnemonic devices and their wide-ranging implications (some historians think the Globe Theatre was designed as a memory palace!). A fun personal memoir on this topic is Joshua Foer's (2011) *Moonwalking with Einstein,* in which he recounts how he became a memory champion. Another related and absolutely fascinating book is Aleksander Luria's (1968) *The Mind of a Mnemonist*, which describes Luria's decades-long study of an individual with extraordinary memory abilities.

In the Land of Invented Languages, Arika Okrent (2010). Why is it so hard to invent new languages, and what does that tell us about human thought? And why are people who invent languages so weird?