Computing the Mind

How the Mind Really Works*

Shimon Edelman
Cornell University

August 2007

*The only book explaining how the mind works that also contains a usable map of the London Tube (see p. 274). You’ll need a magnifying glass, though.
CONTENTS

9.1 Attention ........................................... 397
9.2 Modes of awareness ............................... 409
9.3 The phenomenal Self .............................. 412
9.4 Access consciousness ............................ 427
9.5 The narrative Self ................................. 438
9.6 Narratives of others ............................... 443
9.7 Altered states .................................... 447
9.8 Gone beyond ..................................... 455

10 Imagine ............................................... 463
  10.1 Selfhood and freedom .......................... 463
  10.2 Ethics ............................................ 472
  10.3 Whosoever is wise? ............................. 480

11 The Mind in the Act of Finding .................... 497
  11.1 On the nature of the quest ..................... 497
  11.2 What we have learned .......................... 499
  11.3 Where we are headed ............................ 503
  11.4 Mind games forever ............................. 504

A Bayesian inference and decision making .......... 511
  A.1 The Bayes Theorem ............................... 512
  A.2 Using the Bayes Theorem ........................ 513

B Mind the Gap ......................................... 515
  B.1 Some of the things not included in this book .... 515
  B.2 How this book really ends ........................ 519

Bibliography ............................................ 521

Index .................................................. 587

Credits ................................................ 609
2.4 The heart of the mind: hierarchical abstraction

It is difficult to imagine how computation that is constrained to be effective can amount to anything interesting. How can a sequence of operations that are simple enough to be carried out by mechanisms such as the Turing Machine (Figure 2.6 on page 27) or the dynamical system (Figure 2.7) ever be powerful enough to explain cognition? To understand that, we must realize that the requirement that any effective procedure consist of a series of simple and exact steps does not imply that effective computation is good only for doing arithmetic, or only for supporting the kind of behavior that we have developed a bad habit of calling “robotic”: inflexible and uninventive (it is a bad habit because conscious robots, who are due to appear on the scene before this century is out, will not appreciate it). Although each step on its own must be simple (to allow execution by mindless entities, such as neurons, or electronic circuits controlling state transitions, memory access and input/output functions of a computer), representational systems have at their disposal an excellent method for generating complexity: HIERARCHICAL ABSTRACTION.

Let me show what hierarchical abstraction is by pointing out what it is not. First, hierarchical abstraction is not a luxury. Invoking it to explain the stunning complexity of so much of cognition is really not a far-reaching leap of the imagination, but rather a dictate of the reality. Simplicity is what complexity must be made of, because there isn’t anything else to make it
2.4. THE HEART OF THE MIND: HIERARCHICAL ABSTRACTION

out of, and hierarchical abstraction is the only way in which sufficiently interesting complex stuff can be built out of simple building blocks.

On a related note, **hierarchical abstraction is not a miracle**, but rather a natural by-product of the idea of using symbols as vehicles of representation. The basic requirement of representations is that they make explicit certain properties of the world. This insight, which elaborates on John Locke’s (1690) observation of the indispensability of representations to minds (as stated in the epigraph to the present chapter), has been formulated by David Marr. In an example that appears in his book *Vision*, Marr (1982) compares the suitability of the Roman numerical notation and the modern one for judging the parity of a number. In the Roman notation, it is impossible simply to read parity off the symbolic representation of the number. In the decimal positional notation, it is easier: you have to look just at the rightmost digit, and remember that 0, 2, 4, 6 and 8 are even. The ultimate representation for telling parity at a glance is binary: the number is even if the rightmost digit is 0, and that’s it; the binary notation makes parity explicit. We can imagine a notation that would include an extra bit indicating whether the number represented is prime; although primality can be computed when needed (using one of the many existing precise or approximate algorithms), such a representation would save this work on future occasions by carrying it out once and making the result explicit.

Finally, **hierarchical abstraction is not merely a tool** that the cognitive scientists use to study the mind, but rather, first and foremost, the mind’s tool for the study of the world. True, a complex phenomenon such as a mind is seen as fundamentally simpler precisely when it is represented in terms of hierarchically structured components (rather than, say, in terms of a long list of uniformly simple elementary components). Indeed, it is only when each successive layer of representation allows computations to be carried out by simpler means that an infinite explanatory regress and the need for a homunculus are forestalled. Let us, however, look at the situation from the point of view of the mind itself: without appropriately structured mediating states (representations), a cognitive system would be incapable of dealing with the complex real world. Thus, in a very real sense, there would be no cognitive (or any other) scientists on this (or any other) planet, were it not for the mind’s capacity for hierachical abstraction.

More precisely, hierarchical representations are indispensable for cognitive systems that aspire to *scale up* their understanding of the world. The mind of an *Aplysia californica* may do without hierarchical abstraction; the
mind of Marcel Proust — definitely not.¹⁹

### 2.5 Computation \textit{ad astra}

The power of hierarchical abstraction opens up endless possibilities for the evolution of brains that can acquire, maintain and improve sophisticated systems of representations, and thus serve as hosts to advanced minds. For such entities, even the sky is not a limit. When scientists control an interplanetary probe on a flyby of Saturn, they can afford to think about the task in terms of the specific impulse and the vectoring provided by the rocket engine, abstracting away the details of its function. The probe’s engine, in turn, was built by engineers who could abstract away the chemical composition of nitrogen tetroxide ($N_2O_4$) and monomethylhydrazine ($N_2H_3CH_3$) and focus on channeling the energy made available by their reaction. All this rocket science is made possible by one and the same overarching principle: the capacity for dealing with any information-processing task on a number of levels of abstraction. Consider this example: in each engineer’s brain, there must exist some level of representation capable of treating other members of the team as persons, rather than focusing mindlessly on, say, what they look like (which is the only thing any person’s visual cortex really cares about), or what precise words they use to express a given thought (say, “hydrazine running low,” or “fuel almost exhausted”).

Many of us humans, and certainly most of the human-invented deities, see such unfettered capacity for acquiring and using representations — knowledge — of the world as a grave menace. The mythical torments to which Zeus condemned Prometheus and the curse with which Jehovah cursed the biblical Serpent both express the fears that people’s predilection for occasionally using their brains to the full extent evokes among the faint of heart.²⁰

A contrasting view, found in the Mishna (Avot 2, 13), sees a good brain, along with a good heart, as something to be cherished:

Rabban Jochanan ben Zakkai said to them: Go and see which is the good a man shall cherish most. Rabbi Eliezer said, a good eye. Rabbi Joshua said, a good companion. Rabbi Yosi said, a good neighbor. Rabbi Shimon said, foresight.²¹ Rabbi Elazar said, a good heart. He said to them: I prefer the words of Elazar ben Arach to your words, for in his words yours are included.
To better appreciate this passage, it helps to know that at the time of its writing the Hebrew word for “heart” (lēḇ) stood also for “mind” in all its manifestations: cognitive, affective, and conative (according to Gier (2007), “none of the ancients that we know attributed thinking to the brain, so lēḇ expresses all the functions that we would now ascribe to the brain; Wolff (1974, p.51) lists them as ‘the power of perception, reason, understanding, insight, consciousness, memory, knowledge, reflection, judgment, sense of direction, discernment’”). Thus, the conclusion offered by Rabban (“Great Rabbi”) ben Zakkai — that a “good heart” is to be cherished above other desiderata because it contains them — resonates surprisingly well with the cognitively informed humanistic ethics that I shall describe in Chapter 10.22

Meanwhile, observe that to have a truly good heart, you must know good from evil. How can you do that? Use your brain, of course. In the next chapter, you’ll see what that means.

Notes

1 Because NOTHING IS PERMANENT, objects are merely slow events (Hurford, 2003).

2 BARE-MINIMUM REPRESENTATIONS. Single neurons are not the simplest entities imaginable that are capable of representing things: even a lowly thermostat can represent (the temperature of the room, for the benefit of the furnace). The simplest thermostat consists of a bimetallic strip, which bends this way or that, depending on the ambient temperature, because the two metals have different thermal expansion properties:

In doing so, the strip closes or opens the circuit that fires the furnace.

3 PHYSICAL SYMBOLS. Consider for a moment a representational system that harbors symbols that are not physical. Such a system would be totally ineffectual, because anything non-physical is incapable, as a matter of principle, from interacting in any way with the physical world (if it does interact, it becomes by definition physical). This argument has been very effective in keeping cognitive scientists from succumbing to Cartesian dualism.