AI Magazine Volume 19 Number 3 (1998) (© AAAI)

A Review of How the Mind Works

Tapio Elomaa

Steven Pinker's previous book, the widely successful *The Language Instinct* (Pinker 1994) forcefully promoted Noam Chomsky's idea of a universal grammar underlying particular natural languages, the innate language ability of humans, and presented a theory of how a language organ could have evolved through Darwinian natural selection. His provocative monograph is now attracting challenging texts from other cognitive scientists (for example, Deacon [1997]), but Pinker himself is ahead by leaps and bounds. This time, he has decided to tackle an even larger issue; in his latest book, *How the Mind Works,* Pinker presents his hypothesis of the origins and the logical structure of the whole human mind. A book with such a title and content must, surely, interest any AI researcher.

The book is intended for a wider audience than just the professional cognitive scientists' circle. Therefore, it contains many discussions that appear simple and superficial for anyone who has studied the philosophy and foundations of AI a little deeper. However, working within AI-related topics, there are many of us who have not followed closely the involved cognitive science arguments and are not aware of all the recent developments in psychology. For us, Pinker's book serves as a good introduction to the big picture. The less advanced passages of the book should be worth tolerating even for a dedicated cognitionist because of the book's other merits.

It is not an accident that Steven Pinker keeps authoring popular science books that attain great attention and success; he is a talented writer, and together with his background team, he is capable of digging up funny anecdotes from movies, literature, cartoons, and so on, to invigorate the text. All this adds up to a fluent and entertaining reading experience. However, what about the substance? Partly, the research surveyed in this book can already be considered classical; for example, the extensive coverage of human stereo vision is mostly based on Marr's (1982) seminal account of the subject. However, the experimental psychology research that is reviewed in the book is relatively recent. Many of the ideas that Pinker presents have been in the air in evolutionary psychology; particularly influential and much cited in this book are the studies of Cosmides and Tooby (1994). Pinker's own contribution is to boldly combine all these ideas into a united theory of the mind and its origins.

How the Mind Works, Steven Pinker, W. W. Norton & Co., New York, 1997, 660 pp., $29.90, ISBN 0-393-04535-8.

Reverse Engineering the Mind

In Pinker's own words, the book is about "reverse-engineering the neural computer fitted by natural selection with combinatorial algorithms for causal and probabilistic reasoning about plants, animals, objects, and people. …[The mind is] driven by goal states that served biological fitness in ancestral environments.… It is made up of modules for reasoning about how objects, artifacts, living things, animals, and other human minds work" (pp. 524–525).

Pinker chooses to use the word *engineering* on purpose because he wants to convince the reader that the human mind has really been engineered, by evolution, to its present configuration. As evidence of this teleological construction, he takes the universal, complex, well-engineered, and reproduction-promoting design of, for example, language and stereo vision. According to Pinker, we have innate instincts for such abilities because of the many mental modules that have become a part of our genotype.

Pinker's theory of the human mind stands firmly on two legs: The first one is the *computational theory of mind* (CTM), the tenet that the mind in its entirety is a result of the information processing of the brain and nothing else. CTM also encompasses the basic dogma of AI, *computationality:* Intelligence is a form of information processing. The book's introduction to these concepts borrows directly from AI: Pinker briefly goes through production systems and the simplest models of neural computation (perceptron and backpropagation).

The main contests of the CTM, of course, need to be refuted. First, Pinker is an advocate of the view that information in our minds is represented symbolically. Therefore, he reviews some of the inherent problems of the purely connectionistic approaches to intelligence. Second, Pinker discusses the two best-known denials of computationality and AI—Searle's Chinese room and Penrose's fierce and delusive attacks on AI (cf., LaForte, Hayes, and Ford [1997]). Pinker, fortunately, does not dwell on any of these unfruitful topics too long but swiftly moves on to build his case about the general architecture of the mind.

The other cornerstone of Pinker's case is the theory of evolution. Like AI, Darwinian natural selection still raises feelings and voices. Therefore, the book gives an account of natural selection and adaptation. The review

entails refuting backward causation, which, in retrospect, is what natural selection might seem to do. Adaptation of an organ to perform a different task than its original function, however, has been attacked with arguments of the type "half a wing is not useful." Pinker, therefore, meticulously explains the functioning of adaptation. When discussing evolution, one cannot avoid the argument of what is the impact of heredity (evolution produced) and learning (environment imposed) on the human mind. Of course, heredity and learning are not mutually exclusive, but Pinker's position is that little in the human mind's kernel (and personality) is learned. For example, identical-twin studies have clearly demonstrated that much more than we usually think of human behavior and psyche is innate rather than acquired by interaction with the environment.

The human mind was fitted for ancestral environments, which is why our senses might sometimes leave us stranded in the modern world. For example, visual illusions have long been used as public amusement, and they cause constant amazement. The human vision is probably the most extensively studied and the best understood of the brain modules (mind organs). It serves as Pinker's main evidence for his case. About vision, we know that it is like Pinker paints the rest of the mind to be: The eyes produce the raw data for the brain to chew on. Allocated in its own specified regions of the brain, there is a mental module that takes care of vision production. Image processing entails an enormous amount of computation. There are inherited constraints that affect our interpretation of the scenery that we are witnessing. We compile the images into internal representations (in Marr's [1982] terminology, a 2-1/2–dimensional (D) sketch is converted into a 3-D model). Moreover, the vision ability is innate but attains its final form only through learning. All these aspects of human visual perception have been convincingly demonstrated in psychology, neuroscience, and other disciplines. The evolutionary development of an eye has a feasible explanation.

Reasons for Emotions

In the second half of the book, Pinker is on his own ground—psychology. This is the most interesting part of the investigation into the human mind—exploring ways in which natural selection could have opted for a humanlike intelligence and reverse-engineering emotions. On the surface, it seems that intelligence of human magnitude—and the large, uneconomical brain that it requires—could hardly have been favored over much simpler survival options. In particular, our capacity for general and abstract reasoning seems, from the evolutionary point of view, to be futile. Such a faculty must be an adaptation, adaptation of the reasoning capabilities that helped our foraging ancestors to outwit prey. What has been the impetus for the unique growth of human brain capacities that has not happened with apes or other, more distant animals? Pinker proposes four main reasons: (1) primates are visual animals, and negotiating the 3-D world requires reasoning about movable objects in there; (2) group living stimulates social competition; (3) bipedal walking has left the hands free for creative tool use; (4), hunting, besides demanding brain, enables it to grow by providing the rich nutrient, meat.

Most of us nurture the idea that emotions come from the nature, inhabit the body, are somewhat irrational, and are mostly uncontrollable, whereas the intellect resides in the mind, derives from civilization, is rational, and is deliberate. Pinker paints another picture: Emotions are as engineered as anything else in our mind. They constitute the mechanisms that determine the brain's highest-level goals. This quite recent view in psychology offers us the possibility to reverse engineer emotions, which is a more yielding approach to understanding human nature than any of the former trends of psychology. As one example, Pinker uses morning sickness of pregnant mothers, a phenomenon for which only limp Freudian or vague hormonal explanations were available before. Recently, this universal emotion was reverse engineered to be a protective mechanism in the state when the embryo's organ system starts developing, and it is most vulnerable to toxins.

This explanation for pregnancy sickness is so natural that it almost seems trivial. Nevertheless, it is only a recent discovery, and no satisfactory explanation existed before it. Morning sickness is a specialized emotion that only affects pregnant women for a limited time. Moreover, it is a physical feeling, not a general emotion. Can general emotions such as disgust, fear, happiness, anger, and romantic love also be traced back in the same manner? The generality of these emotions, of course, makes it harder to pinpoint a sole purpose for them, but Pinker gives it a try. Disgust, naturally, works against digesting poisons; fears prepare us to meet dangers; the pursuit of happiness is a (calibrated) way of striving after Darwinian fitness; anger is one of the moralistic emotions—the strategies in the reciprocity game among humans—the particular function of which is to cause desire to punish the cheaters among us (Trivers 1985); and finally, romantic love is an emotion that ensures the unconditionality of pair bonding.

Pinker also tries to find the cause for some human actions—and the emotions behind them—that are most often written off as works of mental illness or lunacy, with no real need for, or possibility of, a psychological explanation. The ghastly example considered in the book is a mass murderer running amok. Even such horrendously extreme actions have an underlying general psychological phenomenon. Human social life is all about relations, and relations are mainly about balancing between parallel and opposite interests. False emotions and other bluffings are tactics in these everyday negotiations. To deter unscrupulous attempts to take advantage of you and hurt you beyond tolerance, a retaliation mechanism must exist, an uncontrollable doomsday machine—the final seal for the authenticity of the feelings and promises. Homicide streaks get an explanation through this theory: They are results of this retaliation mechanism in action but distorted and blown totally out of proportion. (The articulated reasons for the recent schoolyard killings in

Arkansas fit this theory scarily well—see, for example, Gegax, Adler, and Pedersen [1998]).

Life is all about gene spreading—maybe not a person's life but the life on Earth in general. Such is the prerequisite of natural selection. Therefore, human relations and interactions should be, to a high degree, determined by the expected profits and losses in the competition of forwarding one's own genes. That, naturally, pretty much boils down to sexual relationships—to their desirability and undesirability, different benefits and costs of sex for men and women, competition over the most promising mating partners, and so on. Of course, that is not the whole truth; there are other aspects to life than just gene spreading. For example, the social relations and pressures affect our behavior and feelings.

It is inevitable that a book entitled *How the Mind Works* should close with a chapter entitled "The Meaning of Life"—the title of a Monty Python film. The final chapter of the book reflects on the higher values of life. What purpose do the arts, music, fiction, humor and laughter, and religion serve in our life, and what has been their role in the human evolution? The speculative explanations for these aspects of life are not worked through as thoroughly as those for emotions in the preceding chapters of the book.

Pinker is not able to clarify everything by taking the evolutionary viewpoint. An emotion that evades explanation is grief. A short period of sorrow could be explained away by it being a pause that helps to adapt to the changed situation, but what purpose can excessive grief, which can even lead to the death of the griever, serve? What is music's function? Many theories about music have been proposed, but none of them seem convincing enough. Of course, Pinker is not able to solve the most profound philosophical problems, which also tease AI researchers, such as subjective experience, self, free will, meaning, knowledge, and morality. He circumvents even trying to explain these topics by accepting the view of cognitive closure—these matters are simply beyond the human cognitive faculties.

Taking the Geek Out of AI

Many of the ponderings in this book might seem to be out of the scope of an AI researcher's interest, humanistic jabber with no relevance to building intelligent machines—for example, the argument of innate abilities and free will. This book also does not probe deeply into the foundational issues of the CTM, and at times, Pinker skates on thin ice. Why then should an AI researcher spend his/her valuable time with this voluminous book? If you are well acquainted with the theories of evolutionary psychology, Pinker's book will probably not add much to your knowledge. Otherwise, there are many good reasons for reading this book. The approach of reverse engineering the human mind, including the emotions, can widen the scope of AI and make us better see the true potential of our discipline.

Can the CTM model that is developing in evolutionary psychology answer Yeap's (1997) recent cry for an AI theory of the mind—not yet, but taking the evolutionary aspects of the human mind and intellect into account opens new doors for AI research as well. *How the Mind Works* gives us an appreciation of the fact that the human neural computer comes fitted with many basic algorithms, which only need to be tuned by learning for full operation. It is too often that *mind* and *intelligence* are searched from systems that start from a tabula rasa configuration.

Quite often, the AI curriculum has, in addition to computer science, only philosophy and mathematics as other reference disciplines. It is refreshing to get psychology and even biology as other new coordinates. It takes the worst geek out of the idea of AI. In teaching a basic course on AI, I would use this book as a source of additional, motivating material, and I would suggest it as background reading for my students.

Steven Pinker's book takes you on an enjoyable ride within topics of both professional and personal interest. It forces your mind to think about the mind, and the pondering does not stop after you lay the book down; as pointed out in the book, we all are intuitive psychologists. I find this book to be excellent promotion for AI research and Steven Pinker to be a welcome spokesman for our common cause.

References

Cosmides, L., and Tooby, J. 1994. Beyond Intuition and Instinct Blindness: Toward an Evolutionarily Rigorous Cognitive Science. *Cognition* 50:41–77.

Deacon, T. W. 1997. *The Symbolic Species: The Co-Evolution of Language and the Brain.* New York: Norton.

Gegax, T. T.; Adler, J.; and Pedersen, D. 1998. The Boys behind the Ambush. *Newsweek* 131(14): 10–14.

LaForte, G.; Hayes, P. J.; and Ford, K. M. 1997. Why Gödel's Theorem Cannot Refute Computationalism: A Reply to Penrose. Pensacola, Fla.: University of West Florida, Institute for Human and Machine Cognition. Unpublished manuscript.

Marr, D. 1982. *Vision.* New York: Freeman.

Pinker, S. 1994. *The Language Instinct.* New York: Harper Collins.

Trivers R. 1985. *Social Evolution.* Reading, Mass.: Benjamin/Cummings.

Yeap, W. K. 1997. "Emperor AI, Where Is Your New Mind?" *AI Magazine* 18(4): 137–144.

Tapio Elomaa is an assistant professor in the Department of Computer Science at the University of Helsinki, Finland. He currently works as a Marie Curie fellow in the Joint Research Centre of the European Commission in Ispra, Italy. His main field of research is machine learning. His other research interests include knowledge discovery and data mining, robotics, and the philosophy and foundations of AI. His e-mail address is elomaa@cs.helsinki.fi.