

A Review of Mental Leaps: Analogy in Creative Thought

Douglas Hofstadter

■ *Mental Leaps: Analogy in Creative Thought*, Keith Holyoak and Paul Thagard, MIT Press, Cambridge, Massachusetts, 1994, 320 pp., ISBN 0-262-08233-0.

One would hardly expect a book on analogy and creativity to devote seven pages to the Vietnam War, but there it is—one of the most informative histories of the period that I have ever read. Of course, the book's authors, psychologist Keith Holyoak and philosopher Paul Thagard, have good reason for this discussion: to focus on the "analogy war" that went on for years in the upper echelons of the U.S. government.

Politicians think by analogy all the time, and the fates of nations hang on their idiosyncratic analogical instincts, wise or not. Military leaders, too, are guided by precedents, and Holyoak and Thagard ironically note that generals often prepare for the war that they last fought. However, they also point out that one can select one's precedents in a deeper manner than that. In fact, they devote three pages to George Ball, undersecretary of state in the Johnson administration, "who history must now credit as the greatest American political analogist of his time" (p. 163), praising him for seeing further into the Vietnam situation than anyone else in a high-level position.

Ball, instead of likening the situation in Vietnam in the early 1960s to that in Korea 10 years earlier, or to the British capitulation to Hitler in Munich, or to the first domino in a chain, picked out a different analogue: the situation facing the French before their defeat in Indochina in

1954. To be sure, Ball saw the appeal of the Korea, Munich, and domino-chain analogies, but in each, he also saw serious weaknesses; more important, he felt he saw deeper similarities to the situation the French had faced, which led him to his gloomy but accurate predictions of what would happen to American troops if they went in.

I found this discussion enlightening, but it left me wondering, Was Ball really so great an analogist? Holyoak and Thagard tell us, "Ball had worked as a lawyer for the French government during its period of grief in Vietnam, and he never forgot that historical source" (p. 161). So wasn't this experience in some sense the last war that Ball had been

I firmly believe that gist extraction, the ability to see to the core of the matter, is the key to analogy making—indeed, to all intelligence. Unfortunately, Holyoak and Thagard do not seem to agree.

through? Was he not simply using his own last war as his precedent, just as Dean Rusk and others used the Korean War as theirs?

If my analogy between George Ball and the Johnny-one-note military minds is reasonable, then Ball might not deserve the label "great analogist"; having predicted the debacle for the United States shows merely that he happened to have had the good luck to pick the closer analog, just as someone who predicts five

heads in a row and turns out right is merely lucky, not insightful. What, then, would prove that someone was a deep analogy maker? Surely, not just one analogy, no matter how successful. It would take a series of fruit-yielding analogies to reveal that someone had the gift of being able to look at diverse situations and see through to their gists.

Having a reliable ability to see to the gist of a situation—now there, I would say, is the gist of analogy-making skill. When we say, "Iraq's kvetching about Iran's threatening troop maneuvers is surely the pot calling the kettle black," we are seeing a situation on an abstract level, homing in on a phrase that highlights this essence. When Maria inadvertently swims an extra seventeenth lap in her first attempt to do a full mile, and Joe says, "Hey, kiddo, did you know you swam a guinea?", he is conveying what he's just witnessed by alluding to the rather la-di-da British monetary unit worth 21 shillings (one pound plus one extra shilling).

These examples involve putting one's finger on a familiar proverb or even the proverbial *mot juste*, but usually there is no *mot juste* or perfect proverb encapsulating a complex situation; in such cases, we are often reminded of some experience in our past. Our storehouse of experiences includes virtual, as well as real, ones. For example, you hear a cynical remark about TV to the effect that no show ever has to be any good, merely better than its mediocre competitors—and seemingly up from nowhere bubbles the punch line of a joke in which Albert Einstein is hurriedly putting on his tennis shoes to escape a bear, and his hiking companion Niels Bohr says, "Dear Albert, you can't outrun the bear!" to which Einstein replies, "Ja, ja, Bohr, but all what I need to outrun is you."

I firmly believe that *gist extraction*, the ability to see to the core of the matter, is the key to analogy making—indeed, to all intelligence. Unfortunately, Holyoak and Thagard do not seem to agree. At least, they do not focus any attention on this "sense for essence," and it is here that they make their greatest mis-

take, one that permeates their view of analogy and, therefore, seriously mars their book. They indicate no appreciation for either the essential role or the enormous subtlety of this sense for essence. They do not seem to realize the hugeness of the gulf between a full situation as a human perceives it—having no sharp boundaries, woven intricately into the fabric of one's knowledge and life experiences—and a handful of predicate calculus formulas. For example, they constantly refer to “the objects,” “the attributes,” “the relations,” and “the higher-level predicates” in situations, as if real situations (for example, Watergate, a romantic breakup, the abortion debate, the O. J. Simpson trial, John Searle's Chinese room scenario) came given with absolute, unambiguous separations into objects, attributes, and so on—as if distilling *Hamlet* or World War II into two dozen lines of predicate calculus notation were a mechanical task not meriting discussion in a book on analogy. And troublingly, the authors glide back and forth so smoothly between references to real-world situations and references to their tiny, hand-sculpted, frozen caricatures that, I suspect, many readers will be completely unaware of the blur they are witnessing.

I myself realized the subtlety of this effect in their prose when I read their discussion of the analogy between Aesop's sour grapes fable and the story of someone who applies for a job, is turned down, and then says the job would have been boring anyway (p. 31). On first reading it, I was left with a vague twinge of uneasiness, so I went back and tried it again. Suddenly, I saw I had uncritically bought into Holyoak and Thagard's interpretation of these two vignettes, momentarily swallowing their emphasis on causation's central role (“Desiring the job caused the person to apply for it, and being turned down caused the person to say the job was boring” [p. 31]), as well as their plausible-sounding suggestion that “each element in the source maps consistently and uniquely to an element in the target” (p. 31). The critical passage that jolted

me awake was this: “In this analogy there is...only a modest degree of similarity between the corresponding first-order relations, such as *hunger for* and *desire*. The major similarities involve higher-order relations, most notably *cause*” (p. 31).

Of course, being human, I agreed that the sour-grapes-sour-job analogy was strong. The question was why. I was stunned that Holyoak and Thagard were downplaying the role of such vivid, gripping, complex themes as hunger, desire, frustration, disappointment, denial, and self-fooling in favor of a single abstraction so flat and ordinary that it could serve as the core of any event of any sort. Such an overstress on causality—a ubiquitous ingredient of events and, therefore, an essentially useless tool for classifying them—would allow virtually any pair of events to map onto each other. I was reminded of a time a few years ago when, in a demonstration of Falkenhainer, Forbus, and Gentner's analogy-making program SME (a rival to Holyoak and Thagard's ACME, discussed later), I was flabbergasted to realize that the chief reason SME perceived an analogy between two particular situations was that both had been encoded (by people) into predicate-logic trees, each of whose top-level node was the Lisp atom *and*! Although the two situations indeed were analogous, their abstract similarity certainly did not stem from the fact that each could be cast as a pair of subsituations joined by the dime-a-dozen, flavorless Boolean connective *and*. Similarly, the bland, watery, 99-percent-fat-free verb *cause* is not even close to the crux of the sour-grapes-sour-job analogy. To portray it as such is to bend facts in service of an ideology.

Awakened to an ideology lurking backstage, I went back over Holyoak and Thagard's analysis and soon realized there was not, as I had at first docilely accepted, a one-to-one correspondence between the “elements” in the vignettes. What corresponds to the grapevine or gravity, to physical distance, to the fox's many leaps? What to its fatigue or its giving up on the grapes? What to the long snout, reddish fur, bushy tail, brown soil,

bright sun, blue sky, and on and on? What to the workplace, the boss, the job interview, the outfit worn, the nervousness, the sweat, the wait for a decision, the crushing blow? Of course, these seem to be minor aspects that are not part of the essence of either event. However, this is exactly the point! When Holyoak and Thagard said there was a one-to-one correspondence between these situations, they were blurring the notion of *story* with that of *story's gist*. The one-to-one correspondence is, of course, only at the level of the *story's gist*. Getting from the *story* to the *story's gist*, something that occurs swiftly yet almost invisibly in a human mind, is as central a mystery of cognitive science as any that exists. However, Holyoak and Thagard talk about both levels in the same breath, pointing to one but meaning the other, without even seeming to realize the finesse they are carrying out—and I, on my first reading, had happily gone along with them. Once I realized that this conflation of *gist* with full situation had slipped right by me, a long-time skeptic of the whole idea, I knew it could happen to anyone—even the book's authors. Indeed, when I then saw such conflation reoccur over and over again throughout their book, I realized that they might be the least likely of all to see the illusion. Let us now look at this illusion in the context of their computer model.

To explain the principles of their computer program ACME, which maps formalized situations (or, rather, formalized situation gists!) onto each other, Holyoak and Thagard give, in figure 1a, a stripped-down version of the analogy between the events that launched the 1991 Gulf War and those that led to World War II (pp. 248–250). These lines seem to contain the gist of each situation, but they do so only if you know English and remember both situations, which the program does not. Therefore, a less deceptive picture of what the program is faced with is given by the reencoding in figure 1b.

Now let us examine the basis for the mapping that ACME eventually finds—namely, *S* (for Saddam) onto

a Pre-Gulf War <i>president-of (Saddam, Iraq)</i> <i>invade (Iraq, Kuwait)</i>	Pre-World War II <i>führer-of (Hitler, Germany)</i> <i>occupy (Germany, Austria)</i>
b Pre-Gulf War <i>p (S, I)</i> <i>i (I, K)</i>	Pre-World War II <i>f (H, G)</i> <i>o (G, A)</i>

Figure 1.

a Pre-Gulf War <i>president-of (Saddam, Iraq)</i> <i>invade (Iraq, Kuwait)</i>	Arithmetic <i>greater-than (5, 2)</i> <i>twice (2, 1)</i>
b Pre-Gulf War <i>president-of (Saddam, Iraq)</i> <i>invade (Iraq, Kuwait)</i>	Pre-1812 Overture <i>conductor-of (G-Szell, Cleve-orch)</i> <i>march-onto (Cleve-orch, Symph-Hall-stage)</i>

Figure 2.

H (for Hitler), and so on. The crux of ACME is a parallel-processing connectionistic tussle among all *allowable* mappings, “allowable” here meaning that lowercase letters (predicates) can map only onto other lowercase letters and, similarly, for uppercase letters (objects). Thus, *S* might map onto *H*, *G*, or *A*, as might *I* and *K*. In fact, all these hypotheses, silly or not, will be manufactured and entertained by ACME. However, notice that only one uppercase letter occurs twice inside the Gulf War formulas (*I*), and luckily, there is just one uppercase letter that occurs twice inside the World War II formulas as well (*G*). This result gives a great big boost to the object-object mapping $I \Leftrightarrow G$.

Notice now that *I* occurs as the second argument of the Gulf War predicate *p* and that *G* likewise occurs as the second argument of the World War II predicate *f*. This second clue, combined with the $I \Leftrightarrow G$ hypothesis, tips ACME off to the plausibility of mapping *p* onto *f*. Given these guesses, a mechanical process of elimination hands the rest to ACME:

$$I \Leftrightarrow G$$

$$S \Leftrightarrow H$$

$$K \Leftrightarrow A$$

$$p \Leftrightarrow f$$

$$i \Leftrightarrow o$$

Structural similarities of the sort just shown play the starring role in ACME’s mapping process. However, note how little the mapping found has to do with aggressive invasions or ambitious dictators—all that was needed for ACME to find it was that each “war prelude” consisted of two predicates of order two and that the first argument of one was the same as the second argument of the other. To make the program’s dependence on pure syntactic similarity a little clearer, look at the display in figure 2a. Based on just the same structural similarities, a pre-Gulf War/arithmetic analogy can be discovered, with “2” mapping to “Iraq,” “5” to “Saddam,” “twice” to “invade,” and so on. Although this analogy could be claimed to be a cross-domain mental leap—the Holyoak and Thagard grail—it is less than impressive. Surely, more must be involved in a theory

of creative analogy! Indeed, Holyoak and Thagard say that *semantic* similarity also plays a role in guiding the choice of mapping. To illustrate this concept, figure 2b depicts another pair of situations.

ACME sees this analogy as stronger than the arithmetic analogy—if informed ahead of time that *president-of* and *conductor-of* are semantically similar, as are *invade* and *march-onto*. Unfortunately, though, the filing on to a stage by the members of an orchestra hardly seems analogous to the invasion of one country by another. Even though the mapping has both structural and semantic underpinnings, it falls flat to a human. What’s missing?

Following Holyoak and Thagard’s lead, let us go back to the theme song of causality. They point out (p. 250) that Iraq invaded Kuwait because Saddam was Iraq’s president and suggest that such flows of causality play a central role in the pre-Gulf War/pre-World War II analogy. If we were to add a new line to this effect on the left side, and on the right side we also added a line to the effect that Szell, as leader, had planned the concert, ACME would glow with contentment at this mapping between higher-level relations, much as SME glowed over finding the two *ands* that supposedly lay at the core of a deep analogy. Yet, even the parallelism of causalities doesn’t make the analogy seem much more insightful than the arithmetic one. The reason is simple but worrisome: ACME discovers this analogy without having an inkling of what orchestras do, what countries are, what people are—or even what an event or a cause is.

An ACME supporter might object: “This is unfair—a crucial part of the difference between the two situations has been left out! Kuwait *didn’t want* to be invaded, whereas the audience was *looking forward* to the orchestra’s entrance!” Quite true, but notice that this objection shifts the discussion from how mapping is done to what ought to have been encoded in formulas and given to ACME. In other words, it points straight at the gaping hole in the entire approach: There is no gist extraction, no sense for essence.

One could, of course, add more lines to both sides to make the pre-Gulf War/pre-1812 Overture mapping appeal less to ACME. Such patch ups can always be done a posteriori to enhance analogies one hopes for and hinder undesirable ones. However, adding lines made of vacuous symbols won't ever impart understanding to ACME because it was made only to map sets of formulas onto each other, not to understand situations. ACME is not—and never was intended as—a model of the “sense for essence.” And that is most unfortunate because a sense for essence is truly the essence of sense.

The entire discussion of analogy in this book rests on an ironclad distinction between “objects,” “attributes,”

dike between “Vietnam” and “vietnam-of,” and the consequent sloshing back and forth among objects, attributes, and so on, practically defines the fluidity of human thought.

In a way, Holyoak and Thagard's own nonchalant sloshing back and forth between full real-world situations and their tiny gists epitomizes what I am talking about: It is the human mind at its most fluid. Unfortunately, our mental fluidity, although responsible for our best insights and most creative findings, often leads us into deep confusions as well. To conflate one's understanding of an Aesop's fable with a few abstractions forming its core, or worse, with a few carefully chosen lines of predicate calculus, is just such a confusion. Another

wants to be accused of falling for or promulgating, although probably all of us do to some extent. In Holyoak and Thagard's case, they have managed to convince themselves, and hope to convince others, that their program makes analogies between Saddam and Hitler, between solar systems and atoms, between Socrates and a midwife, between *Hamlet* and *King Lear*. Given such ostensible successes, they would naturally be unreceptive to any insinuation that their model could have omitted something critical about analogy making and loath to consider the suggestion that the true crux of analogy making—including gist extraction and active concepts—might be studied more fruitfully in a well-designed microworld. Regrettably, Holyoak and Thagard have bought into today's collective wisdom that shuns microworlds as outmoded and irrelevant to modeling cognition. Anyway, whatever modest successes one might have in a microworld, how much grander to be able to say that one's program had, on its own, discovered the analogy between *West Side Story* and *Romeo and Juliet*! And so, real world, ho!

However, Holyoak and Thagard's way of computationally modeling real-world analogy making is, before the computer is even plugged in, to hand shrink each real-world situation into a tiny, frozen caricature of itself, containing precisely its core and little else. Only then does the computer enter the picture, mechanically converting the two frozen representations into a highly overpopulated connectionist network, most of whose nodes represent mappings so silly that a human mind would never consider them, and then performing a kind of relaxation process in which the network sighs and settles down into an equilibrium. During this slow sagging process, no fresh new concepts or ideas can possibly enter the characterization of either situation; in fact, nothing changes at all except for numeric values on nodes, the biggest of which tells what seems to match what. What emerges at the end of this dull, unconscious sagging is an analogy—or, rather, a match up between certain Lisp atoms.

Many AI researchers, in their haste to reach what passes for real-world performance, settle for surface-level appearance and sweep deep issues under the rug.

“relations,” and so on. AI researchers are, of course, free to erect rigid walls anywhere they want in their computer code, but in a human mind, such distinctions are anything but ironclad. Hitler, for example, was the one and only Führer of Germany. For this reason, we humans experience a profound blur between the role “Führer” and the filler “Hitler.” As Gilles Fauconnier has documented so extensively in his book *Mental Spaces* (MIT Press, 1985), role-filler blurs are ubiquitous in human thought and pervade human language. But the ACME model would forbid, every bit as rigidly as arithmetic forbids multiplying an equals sign by the integer 7, the mapping of the predicate *fürher-of* onto the object *Saddam*. However, the analogy “Saddam is the Hitler of Iraq” slides the object *Hitler* into a predicate *hitler-of*—a role that the object *Saddam* can fill. In a human mind, a role-filler blur will ensue, much as Vietnam turned into a category about which one could cry “No more Vietnams!” or even assert “Cambodia is the Vietnam of Vietnam.” A mind erects no watertight object-relation

er such confusion is the complete omission from a computer model of analogy making of any model of *active concepts*—that is, a model of how concepts in long-term memory can tentatively be recognized in a situation, activated to various degrees, and selectively promoted from long-term memory into working memory to serve as constituents of fluid, hierarchical representational structures. To my mind, trying to develop a theory of analogy making while bypassing both gist extraction and the nature of concepts is as utterly misguided as trying to develop a theory of musical esthetics while omitting all mention of both melody and harmony.

Probably the reason for these major gaps in Holyoak and Thagard's approach is that like so many others in AI, they are eager to tackle the real world—too eager. Many AI researchers, in their haste to reach what passes for real-world performance, settle for surface-level appearance and sweep deep issues under the rug. In so doing, the researchers themselves fall victim to the ELIZA effect, something that nobody in AI

At this point, the tiny, inert predicate calculus cores are conflated with the original full-blown situations, subtly leading many intelligent people to such happy conclusions as that the program has insightfully leaped to a cross-domain analogy, that it has used semantics and purpose in settling on its mappings, that it has understood how Socrates coaxes ideas from students much as a midwife helps bring babies into the world, and so on. However, as the mappings of “Iraq” onto “2” and “Cleveland-orchestra” reveal, ACME, despite its creators’ wishes and claims, functions in a microworld; it’s just a microworld whose sparseness is disguised by a paper-thin facade of real-world terms. Alas, for numerous reasons, anemic microworlds posing as rich macro-worlds are par for the course in most of AI today. Deep and facadeless microworld projects such as the ancient SHRDLU seem rare as hen’s teeth.

A key premise of *Mental Leaps* is that purpose is indispensable to analogy making. In Holyoak and Thagard’s theory, purpose seems to play the dubious role that spin plays in politics and the law, serving merely to distort various aspects of a situation, trying to get people to see things in a particular way. As I see it, the authors waffle about purpose, sometimes doing their best to show how central a role it plays, other times implying that purpose-free analogies are the best of all.

In several different contexts in Chapters 6 and 7, they reveal admiration for the impartial, wise figure who objectively chooses between rival analogies proposed by purpose-driven, hence biased, parties. One context they describe is law, where a judge evaluates the merits of different precedents proposed by opposing lawyers. Another context is baseball, where a neutral arbitrator sets players’ salaries by choosing between rival analogies proposed by the player and the team. A third such context is philosophical argumentation, where analogies called *intuition pumps* are proposed by philosophers to get people to see abstract issues in a particular, biased way. Here, Holyoak and Thagard themselves play the role of

the objective, metalevel judge, using John Searle’s famous Chinese-room scenario as their prime example of a flawed intuition pump, which it certainly is, and pointing out its flaws. (Ironically, on the very next page, they voice pessimism about the possibility of there ever being an objective way to evaluate intuition pumps. This is what I meant by “waffling.”) A fourth arena where they show respect for detachment and unbiasedness in analogy making is politics, illustrated by their high praise for George Ball, who saw more deeply than others into the Vietnam situation.

I would certainly agree that most human analogy making is highly biased. For this reason, anyone who wants to study analogy making in its full glory and squalor will have to consider all its flaws, a major one of which is bias, whether conscious or unconscious. But many of us strive, as does an arbitrator, a judge, or a scientist, after unbiased, purpose-independent insights, and it is this quest for objectivity, this attempt to banish purpose from one’s analogies, that Holyoak and Thagard seem to have a hard time reconciling with their stress on the indispensability of purpose, with all the subjectivity and intellectual relativism that it implies.

If I have been harsh on this book, it is because it is so pervaded by the authors’ computer-modeling approach, with which I obviously disagree on many levels. However, the book was extremely engrossing and gave me many new insights about how the mechanisms for analogical thinking evolved over eons, develop in each of us as we grow, and function in real time as we strive to deal with life’s imponderables.

Chapter 3, for example, about analogy making by apes, was most rewarding. The crucial idea of a ladder of levels of understanding of the abstract notion of sameness, although I could quibble with some of the distinctions, is well charted. In particular, there is a long discussion of the degree to which the chimpanzee Sarah’s ability to make certain fairly abstract analogies was a consequence of her having been trained linguistically and, in particular, of her having learned a symbol for

the very concept of sameness!

Chapter 4, on analogy making in children, reports on many ingenious and astonishing experiments. My favorites, done by psychologist Judy DeLoache and colleagues, revealed a deep mental breach between 2-1/2-year olds and 3-year olds. Her first experiment showed that 3-year olds, if they saw an adult hiding “Little Snoopy” behind or under toy furniture in a small-scale model of a room and then were taken to the full-sized room and asked to find “Big Snoopy,” would generally make a beeline for “the same place” and find it. The younger children, by contrast, could almost never do such a mapping. They couldn’t go from one domain to the other. However, DeLoache’s next experiment showed something amazing: If these younger children were tricked into thinking that the experimenters had a “size-changing machine” that could make a puppet or a dollhouse room double in size, they would have no trouble finding “the same place” in a supposedly enlarged room that was magically made by the machine, right before their innocent eyes, from a small room. In other words, as long as 2-1/2-year olds thought they were dealing with a single room at different sizes and, thus, didn’t need to make a mapping from one room to another, they effortlessly found the toy, but when the search task seemed to involve two different rooms, thus requiring the mental building of a room–room mapping, they failed. These and other wonderful experiments on children’s thinking are excellently described, and even if one balks at some of the notions and notations used in the discussions, there is no doubt that deep revelations about the human mind are being presented, and presented in a lively, engaging way.

Chapter 8 is a stimulating one—mostly on analogy in scientific discovery, technological innovation, and teaching—but even better is Chapter 9, much of which concerns metaphor. Perhaps my enthusiasm is predictable because metaphor is an area so cloudy and complex that Holyoak and Thagard don’t try to

shoehorn everything into the terms of their own theory. Rather, they are fully cognizant of the magnitude of the barrier and welcome it as a challenge. Most surprising to me was that in discussing the metaphor “Socrates as a midwife of ideas” (pp. 219–220), they back away from their usual pre-fabricated-gist approach (although they offer no alternative to it). Indeed, in their excellent discussion of analogy, metonymy, and metaphor, they verge on the emerging notions of frame blending and blended spaces, which I think are promising and important and will offer a new view of analogy as part of a much more comprehensive theory of human understanding. In this, I am happy to say, I find exciting convergences between the thinking of Holyoak and Thagard, the ideas of linguists Gilles Fauconnier, George Lakoff, and Mark Turner, and my own thoughts, among others. Slowly but surely, the centrality and ubiquity of analogy and metaphor in the human mind are being revealed, thanks to such work.

Although most of this book is committed to an impossibly rigid and implausible theory of analogy making, it manages to survey analogical thinking in a remarkably thorough way, make clear its place and power in cognition, and pinpoint many of the most perplexing issues about analogy. The authors are to be saluted for their impressive bringing together of so many ideas from so many sources, and it is to be hoped that in their next book on analogy, they will have moved beyond the seriously flawed tacit assumptions that kept this book from becoming all it could and truly should have been.

Douglas Hofstadter is on the faculty at Indiana University, where he is College of Arts and Sciences professor of cognitive science and computer science. He directs the Center for Research on Concepts and Cognition and is the author of numerous books and papers, including *Gödel, Escher, Bach: an Eternal Golden Braid* (Basic Books, 1979) and, most recently, *Fluid Concepts and Creative Analogies* (Basic Books, 1995).

Android Epistemology

Edited by Kenneth M. Ford, Clark Glymour, & Patrick J. Hayes

Published by the AAAI Press / The MIT Press
336 pp., \$25.00 cloth, ISBN 0-262-06184-8

Epistemology has traditionally been the study of human knowledge and rational change of human belief. Android epistemology is the exploration of the space of possible machines and their capacities for knowledge, beliefs, attitudes, desires, and action in accord with their mental states. From the perspective of android epistemology, artificial intelligence and computational cognitive psychology form a unified endeavor: artificial intelligence explores any possible way of engineering machines with intelligent features, while cognitive psychology focuses on reverse engineering the most intelligent system we know, us. The editors argue that contemporary android epistemology is the fruition of a long tradition in philosophical theories of knowledge and mind.

The sixteen essays by computer scientists and philosophers collected in this volume include substantial contributions to android epistemology, as well as examinations, defenses, elaborations, and challenges to the very idea.

Contents

- The Prehistory of Android Epistemology / *Clark Glymour, Kenneth Ford, & Patrick Hayes*
- Machine as Mind / *Herbert A. Simon*
- The Vitalists' Last Stand / *Anatol Rapoport*
- Could a Robot Be Creative--And Would We Know? / *Margaret A. Boden*
- From Cognitive Systems to Persons / *Antoni Gomila*
- Could, How Could We Tell if, and Why Should, Androids Have Inner Lives? / *Selmer Bringsjord*
- Android Epistemology: An Essay on Interpretation and Intentionality / *Kalyan Shankar Basu*
- Taking Embodiment Seriously: Nonconceptual Content and Robotics / *Ronald L. Chrisley*
- Imagination and Situated Cognition / *Lynn Andrea Stein*
- Towards Constructivist Unification of Machine Learning and Parallel Distributed Processing / *Chris Stary and Markus F. Peschl*
- Towards A Sentential 'Reality' for the Android / *Cary G. deBessonet*
- Towards the Ethical Robot / *James Gips*
- The Ethics of Autonomous Learning Systems / *A. F. Umar Khan*
- How to Settle an Argument / *Henry E. Kyburg, Jr.*
- Machine Stereopsis: A Feedforward Network for Fast Stereo Vision with Movable Fusion Plane / *Paul M. Churchland*
- Alienable Rights / *Marvin Minsky*

To order, call toll free 1-800-356-0343 (US & Canada) or (617) 625-8569, MasterCard & VISA accepted. Prices will be higher outside the US and are subject to change without notice. Visit the AAAI Press website! <http://www.aaai.org/Publications/Press/press.html>