

True Names: Afterword by Marvin Minsky

In real life, you often have to deal with things you don't completely understand. You drive a car, not knowing how its engine works. You ride as passenger in someone else's car, not knowing how that driver works. And strangest of all, you sometimes drive yourself to work, not knowing how you work, yourself.

To me, the import of *True Names* is that it is about how we cope with things we don't understand. But, how do we ever understand anything in the first place? Almost always, I think, by using analogies in one way or another—to pretend that each alien thing we see resembles something we already know. When an object's internal workings are too strange, complicated, or unknown to deal with directly, we extract whatever parts of its behavior we can comprehend and represent them by familiar symbol—or the names of familiar things which we think do similar things. That way, we make each novelty at least appear to be like something which we know from the worlds of our own pasts. It is a great idea, that use of symbols; it lets our minds transform the strange into the commonplace. It is the same with names.

Right from the start, *True Names* shows us many forms of this idea, methods which use symbols, names, and images to make a novel world resemble one where we have been before. Remember the doors to Vinge's castle? Imagine that some architect has invented a new way to go from one place to another: a scheme that serves in some respects the normal functions of a door, but one whose form and mechanism is so entirely outside our past experience that, to see it, we'd never think of it as a door, nor guess what purposes to use it for. No matter: just superimpose, on its exterior, some decoration which reminds one of a door. We could clothe it in rectangular shape, or add to it a waist-high knob, or a push-plate with a

sign lettered “EXIT” in red and white, or do whatever else may seem appropriate—and every visitor from Earth will know, without a conscious thought, that pseudo-portal’s purpose, and how to make it do its job.

At first this may seem mere trickery; after all, this new invention, which we decorate to look like a door, is not really a door. It has none of what we normally expect a door to be, to wit: hinged, swinging slab of wood, cut into wall. The inner details are all wrong. Names and symbols, like analogies, are only partial truths; they work by taking many-levelled descriptions of different things and chopping off all of what seem, in the present context, to be their least essential details—that is, the ones which matter least to our intended purposes. But, still, what matters—when it comes to using such a thing—is that whatever symbol or icon, token or sign we choose should remind us of the use we seek which, for that notquite-door, should represent some way to go from one place to another. Who cares how it works, so long as it works! It does not even matter if that “door” leads to anywhere: in *True Names*, nothing ever leads anywhere; instead, the protagonists’ bodies never move at all, but remain plugged-in to the network while programs change their representations of the simulated realities!

Ironically, in the world *True Names* describes, those representations actually do move from place to place—but only because the computer programs which do the work may be sent anywhere within the worldwide network of connections. Still, to the dwellers inside that network, all of this is inessential and imperceptible, since the physical locations of the computers themselves are normally not represented anywhere at all inside the worlds they simulate. It is only in the final acts of the novel, when those partially-simulated beings finally have to protect themselves against their entirely-simulated enemies, that the programs must

keep track of where their mind-computers are; then they resort to using ordinary means, like military maps and geographic charts.

And strangely, this is also the case inside the ordinary brain: it, too, lacks any real sense of where it is. To be sure, most modern, educated people know that thoughts proceed inside the head—but that is something which no brain knows until it's told. In fact, without the help of education, a human brain has no idea that any such things as brains exist. Perhaps we tend to place the seat of thought behind the face, because that's where so many sense-organs are located. And even that impression is somewhat wrong: for example, the brain-centers for vision are far away from the eyes, away in the very back of the head, where no unaided brain would ever expect them to be.

In any case, the point is that the icons in *True Names* are not designed to represent the truth—that is, the truth of how the designated object, or program, works; that just is not an icon's job. An icon's purpose is, instead, to represent a way an object or a program can be used. And, since the idea of a use is in the user's mind—and not connected to the thing it represents—the form and figure of the icon must be suited to the symbols that the users have acquired in their own development. That is, it has to be connected to whatever mental processes are already one's most fluent, expressive, tools for expressing intentions. And that's why Roger represents his watcher the way his mind has learned to represent a frog.

This principle, of choosing symbols and icons which express the functions of entities—or rather, their users' intended attitudes toward them—was already second nature to the designers of earliest fast-interaction computer systems, namely, the early computer games which were, as Vemur Vinge says, the ancestors of the Other Plane in which the novel's

main activities are set. In the 1970's the meaningful-icon idea was developed for personal computers by Alan Kay's research group at Xerox, but it was only in the early 1980's, after further work by Steven Jobs' research group at Apple Computer, that this concept entered the mainstream of the computer revolution, in the body of the Macintosh computer.

Over the same period, there have also been lesspublicized attempts to develop iconic ways to represent, not what the programs do, but how they work. This would be of great value in the different enterprise of making it easier for programmers to make new programs from old ones. Such attempts have been less successful, on the whole, perhaps because one is forced to delve too far inside the lower-level details of how the programs work. But such difficulties are too transient to interfere with Vinge's vision, for there is evidence that he regards today's ways of programming— which use stiff, formal, inexpressive languages—as but an early stage of how great programs will be made in the future.

Surely the days of programming, as we know it, are numbered. We will not much longer construct large computer systems by using meticulous but conceptually impoverished procedural specifications. Instead, we'll express our intentions about what should be done, in terms, or gestures, or examples, at least as resourceful as our ordinary, everyday methods for expressing our wishes and convictions. Then these expressions will be submitted to immense, intelligent, intention-understanding programs which will themselves construct the actual, new programs. We shall no longer be burdened with the need to understand all the smaller details of how computer codes work. All of that will be left to those great utility programs, which will perform the arduous tasks of applying what we have embodied in them, once and for all, of what we know about the arts of lower-level programming. Then, once we learn better ways to tell computers what we want them to get done, we will be able to

return to the more familiar realm of expressing our own wants and needs. *For, in the end, no user really cares about how a program works, but only about what it does—in the sense of the intelligible effects it has on other things with which the user is concerned.*

In order for that to happen, though, we will have to invent and learn to use new technologies for “expressing intentions”. To do this, we will have to break away from our old, though still evolving, programming languages, which are useful only for describing processes. And this may be much harder than it sounds. For, it is easy enough to say that all we want to do is but to specify *what we want to happen*, using more familiar modes of expression. But this brings with it some very serious risks.

The first risk is that this exposes us to the consequences of self-deception. It is always tempting to say to oneself, when writing a program, or writing an essay, or, for that matter, doing almost anything, that *“I know what I would want, but I can’t quite express it clearly enough”*. However, that concept itself reflects a too-simplistic self-image, which portrays one’s own self as existing, somewhere in the heart of one’s mind (so to speak), in the form of a pure, uncomplicated entity which has pure and unmixed wishes, intentions, and goals. This pre-Freudian image serves to excuse our frequent appearances of ambivalence; we convince ourselves that clarifying our intentions is a mere matter of straightening-out the input-output channels between our inner and outer selves. The trouble is, we simply aren’t made that way, no matter how we may wish we were.

We incur another risk whenever we try to escape the responsibility of understanding how our wishes will be realized. It is always dangerous to leave much choice of means to any servants we may choose—no matter whether we program them or not. For, the larger the range of choice of methods they may use, to gain for us the ends we think we seek, the more

we expose ourselves to possible accidents. We may not realize, perhaps until it is too late to turn back, that our goals were misinterpreted, perhaps even maliciously, as in such classic tales of fate as Faust, the Sorcerer's Apprentice, or The Monkey's Paw (by W.W. Jacobs).

The ultimate risk, though, comes when we greedy, lazy, master-minds are able at last to take that final step: to design goal-achieving programs which are programmed to make themselves grow increasingly powerful, by using learning and self-evolution methods which augment and enhance their own capabilities. It will be tempting to do this, not just for the gain in power, but just to decrease our own human effort in the consideration and formulation of our own desires. If some genie offered you three wishes, would not your first one be, "*Tell me, please, what is it that I want the most!*" The problem is that, with such powerful machines, it would require but the slightest accident of careless design for them to place their goals ahead of ours, perhaps the well-meaning purpose of protecting us from ourselves, as in *With Folded Hands*, by Jack Williamson),—or to protect us from an unsuspected enemy, as in *Colossus* by D.H. Jones, or because, like Arthur C. Clarke's HAL, the machine we have built considers us inadequate to the mission we ourselves have proposed, or, as in the case of Vernor Vinge's own Mailman, who teletypes its messages because it cannot spare the time to don disguises of dissimulated flesh, simply because the new machine has motives of its very own.

Now, what about the last and finally dangerous question which is asked toward *True Names*' end? Are those final scenes really possible, in which a human user starts to build itself a second, larger Self inside the machine? Is anything like that conceivable? And if it were, then would those simulated computerpeople be in any sense the same as their human models before them; would they be genuine extensions of those real people? Or would they

merely be new, artificial, person-things which resemble their originals only through some sort of structural coincidence? What if the aging Erythrina's simulation, unthinkably enhanced, is permitted to live on inside her new residence, more luxurious than Providence? What if we also suppose that she, once there, will be still inclined to share it with Roger—since no sequel should be devoid of romance—and that those two tremendous entities will love one another? Still, one must inquire, what would those super-beings share with those whom they were based upon? To answer that, we have to think more carefully about what those individuals were before. But, since these aren't real characters, but only figments of an author's mind, we'd better ask, instead, about the nature of our selves.

Now, once we start to ask about our selves, we'll have to ask how these, too, work—and this is what I see as the cream of the jest because, it seems to me, that inside every normal person's mind is, indeed, a certain portion, which we call the Self—but it, too, uses symbols and representations very much like the magic spells used by those players of the Inner World to work their wishes from their terminals. To explain this theory about the working of human consciousness, I'll have to compress some of the arguments from “The Society of Mind“, my forthcoming book. In several ways, my image of what happens in the human mind resembles Vinge's image of how the players of the Other Plane have linked themselves into their networks of computing machines—by using superficial symbol-signs to control of host of systems which we do not fully understand.

Everybody knows that we humans understand far less about the insides of our minds, than what we know about the world outside. We know how ordinary objects work, but nothing of the great computers in our brains. Isn't it amazing we can think, not knowing what it means to think? Isn't it bizarre that we can get ideas, yet not be able to explain what

ideas are. Isn't it strange how often we can better understand our friends than ourselves?

Consider again, how, when you drive, you guide the immense momentum of a car, not knowing how its engine works, or how its steering wheel directs the vehicle toward left or right. Yet, when one comes to think of it, don't we drive our bodies the same way? You simply set yourself to go in a certain direction and, so far as conscious thought is concerned, it's just like turning a mental steering wheel. All you are aware of is some general intention—*It's time to go: where is the door?*—and all the rest takes care of itself. But did you ever consider the complicated processes involved in such an ordinary act as, when you walk, changing the direction you're going in? It is not just a matter of, say, taking a larger or smaller step on one side, the way one changes course when rowing a boat. If that were all you did, when walking, you would tip over and fall toward the outside of the turn.

Try this experiment: watch yourself carefully while turning—and you'll notice that, before you start the turn, you tip yourself in advance; this makes you start to fall toward the inside of the turn; then, when you catch yourself on the next step, you end up moving in a different direction. When we examine that more closely, it all turns out to be dreadfully complicated: hundreds of interconnected muscles, bones, and joints are all controlled simultaneously, by interacting programs which locomotion-scientists still barely comprehend. Yet all your conscious mind need do, or say, or think, is *Go that way!*—assuming that it makes sense to speak of the conscious mind as thinking anything at all. So far as one can see, we guide the vast machines inside ourselves, not by using technical and insightful schemes based on knowing how the underlying mechanisms work, but by tokens, signs, and symbols which are entirely as fanciful as those of Vinge's sorcery. It even makes one wonder if it's fair for us to gain our ends by casting spells upon our helpless hordes of mental under-thralls.

Now, if we take this only one more step, we see that, just as we walk without thinking, we also think without thinking! That is, we just as casually exploit the agencies which carry out our mental work. Suppose you have a hard problem. You think about it for a while; then after a time you find a solution. Perhaps the answer comes to you suddenly; you get an idea and say, “Aha, I’ve got it. I’ll do such and such.” But then, were someone to ask how you did it, how you found the solution, you simply would not know how to reply. People usually are able to say only things like this:

“I suddenly realized...” “I just got this idea...” “It occurred to me that...”

If we really knew how our minds work, we wouldn’t so often act on motives which we don’t suspect, nor would we have such varied theories in psychology. Why, when we’re asked how people come upon their good ideas, are we reduced to superficial reproductive metaphors, to talk about “conceiving” or “gestating“, or even “giving birth” to thoughts? We even speak of “ruminating” or “digesting” as though the mind were anywhere but in the head. If we could see inside our minds we’d surely say more useful things than “Wait. I’m thinking.”

People frequently tell me that they’re absolutely certain that no computer could ever be sentient, conscious, self-willed, or in any other way “aware” of itself. They’re often shocked when I ask what makes them sure that they, themselves, possess these admirable qualities. The reply is that, if they’re sure of anything at all, it is that “*I’m aware hence I’m aware.*”

Yet, what do such convictions really mean? Since “Self-awareness” ought to be an awareness of what’s going on within one’s mind, no realist could maintain for long that people really have much insight, in the literal sense of seeing in.

Isn’t it remarkable how certainly we feel that we’re self-aware—that we have such broad

abilities to know what's happening inside ourselves? The evidence for that is weak, indeed. It is true that some people seem to have special excellences, which we sometimes call "insights", for assessing the attitudes and motivations for other people. And certain individuals even sometimes make good evaluations of themselves. But that doesn't justify our using names like insight or self-awareness for such abilities. Why not simply call them "person-sights" or "person-awareness?" Is there really reason to suppose that skills like these are very different from the ways we learn the other kinds of things we learn? Instead of seeing them as "seeing in," we could regard them as quite the opposite: just one more way of "figuring out." Perhaps we learn about ourselves the same ways that we learn about un-self-ish things.

The fact is, the parts of ourselves which we call "self aware" are only a small fraction of the entire mind. They work by building simulated worlds of their own—worlds which are greatly simplified, in comparison with either the real world outside, or with the immense computer systems inside the brain: systems which no one can pretend, today, to understand. And our worlds of simulated awareness are worlds of simple magic, wherein each and every imagined object is invested with meanings and purposes. Consider how one can but scarcely see a hammer except as something to hammer with, or see a ball except as something to throw and catch. Why are we so constrained to perceive things, not as they are, but as they can be used? *Because the highest levels of our minds are goal-directed problem-solvers.* That is to say that all the machines inside our heads evolved, originally, to meet various built-in or acquired needs, for comfort and nutrition, for defense and for reproduction. Later, over the past few million years, we evolved even more powerful sub-machines which, in ways we don't yet understand, seem to correlate and analyze to discover which kinds of actions cause which sorts of effects; in a word, to discover what we call knowledge. And though we often like to

think that knowledge is abstract, and that our search for it is pure and good in itself—still, we ultimately use it for its ability to tell us what to do to gain whichever ends we seek (even when we conclude that in order to do that, we may first need to gain yet more and more knowledge). Thus, because, as we say, “knowledge is power“, our knowledge itself is enmeshed in those webs of ways we reach our goals. And that’s the key: it isn’t any use for us to know, unless our knowledge tells us what to do. This is so wrought into the conscious mind’s machinery that it seems too obvious to state: *no knowledge is of any use unless we have a use for it.*

Now we come to see the point of consciousness: it is the part of the mind most specialized for knowing how to use the other systems which lie hidden in the mind. But it is not a specialist in knowing how those systems actually work, inside themselves. Thus, as we said, one walks without much sense of how it’s done. It’s only when those systems start to fail to work well that consciousness becomes engaged with small details. That way, a person who has sustained an injured leg may start, for the first time, consciously to make theories about how walking works: *To turn to the left, I’ll have to push myself that way—and then one has to figure out, with what?* It is often only when we’re forced to face an unusually hard problem that we become more reflective, and try to understand more about how the rest of the mind ordinarily solves problems; at such times one finds oneself saying such things as, “Now I must get organized. Why can’t I concentrate on the important questions and not get distracted by those other inessential details?”

It is mainly at such moments—the times when we get into trouble—that we come closer than usual to comprehending how our minds work, by engaging the little knowledge we have about those mechanisms, in order to alter or repair them. It is paradoxical that these are

just the times when we say we are “confused“, because it is very intelligent to know so much about oneself that one can say that—in contrast merely to being confused and not even knowing it. Still, we disparage and dislike awareness of confusion, not realizing what a high degree of self-representation it must involve. Perhaps that only means that consciousness is getting out of its depth, and isn’t really suited to knowing that much about how things work. In any case, even our most “conscious” attempts at selfinspection still remain confined mainly to the pragmatic, magic world of symbol-signs, for no human being seems ever to have succeeded in using selfanalysis to find out very much about the programs working underneath.

So this is the irony of *True Names*. Though Vinge tells the tale as though it were a science-fiction fantasy—it is in fact a realistic portrait of our own, real-life predicament! I say again that we work our minds in the same unknowing ways we drive our cars and our bodies, as the players of those futuristic games control and guide what happens in their great machines: by using symbols, spells and images— as well as secret, private names. The parts of us which we call “consciousness” sit, as it were, in front of cognitive computer-terminals, trying to steer and guide the great unknown engines of the mind, not by understanding how those mechanisms work, but simply by selecting names from menu-lists of symbols which appear, from time to time, upon our mental screen-displays.

But really, when one thinks of it, it scarcely could be otherwise! Consider what would happen if our minds indeed could really see inside themselves. What could possibly be worse than to be presented with a clear view of the trillion-wire networks of our nervecell connections? Our scientists have peered at fragments of those structures for years with powerful microscopes, yet failed to come up with comprehensive theories of what those

networks do and how. How much more devastating it would be to have to see it all at once!

What about the claims of mystical thinkers that there are other, better ways to see the mind. One recommended way is learning how to train the conscious mind to stop its usual sorts of thoughts and then attempt (by holding very still) to see and hear the fine details of mental life. Would that be any different, or better, than seeing them through instruments? Perhaps—except that it doesn't face the fundamental problem of how to understand a complicated thing! For, if we suspend our usual ways of thinking, we'll be bereft of all the parts of mind already trained to interpret complicated phenomena. Anyway, even if one could observe and detect the signals which emerge from other, normally inaccessible portions of the mind, these probably would make no sense to the systems involved with consciousness, because they represent unusually low level details. To see why this is so, let's return once more to understanding such simple things as how we walk.

Suppose that, when you walk about, you were indeed able to see and hear the signals in your spinal cord and lower brain. Would you be able to make any sense of them? Perhaps, but not easily. Indeed, it is easy to do such experiments, using simple bio-feedback devices to make those signals audible and visible; the result is that one may indeed more quickly learn to perform a new skill, such as better using an injured limb. However, just as before, this does not appear to work through gaining a conscious understanding of how those circuits work; instead the experience is very much like business as usual; we gain control by acquiring just one more form of semi-conscious symbol-magic. Presumably, what happens is that a new control system is assembled somewhere in the nervous system, and interfaced with superficial signals we can know about. However, bio-feedback does not appear to provide any different insights into how learning works than do our ordinary, built-in senses. In any

case, our locomotion-scientists have been tapping such signals for decades, using electronic instruments. Using those data, they have been able to develop various partial theories about the kinds of interactions and regulation-systems which are involved. However, these theories have not emerged from relaxed meditation about, or passive observation of those complicated biological signals; what little we have learned has come from deliberate and intense exploitation of the accumulated discoveries of three centuries of our scientists' and mathematicians' study of analytical mechanics and a century of newer theories about servo-control engineering. It is generally true in science that just observing things carefully rarely leads to new "insights" and understandings. One must first have at least the glimmerings of the form of a new theory, or of a novel way to describe: one needs a "new idea". For the "causes" and the "purposes" of what we observe are not themselves things that can be observed; to represent *them*, we need some other mental source to invent new magic tokens.

But where do we get the new ideas we need? For any single individual, of course, most concepts come from the societies and cultures that one grows up in. As for the rest of our ideas, the ones we "get" all by ourselves, these, too, come from societies—but, now, the ones inside our individual minds. For, a human mind is not in any real sense a single entity, nor does a brain have a single, central way to work. Brains do not secrete thought the way livers secrete bile; a brain consists of a huge assembly of sub-machines which each do different kinds of jobs—each useful to some other parts. For example, we use distinct sections of the brain for hearing the sounds of words, as opposed to recognizing other kinds of natural sounds or musical pitches. There is even solid evidence that there is a special part of the brain which is specialized for seeing and recognizing faces, as opposed to visual perception of other, ordinary things. I suspect that there are, inside the cranium, perhaps as many

as a hundred kinds of computers, each with its own somewhat different architecture; these have been accumulating over the past four hundred million years of our evolution. They are wired together into a great multi-resource network of specialists, which each knows how to call on certain other specialists to get things done which serve its purposes. And each of these sub-brains uses its own styles of programming and its own forms of representations; there is no standard, universal language-code.

Accordingly, if one part of that Society of Mind were to inquire about another part, this probably would not work because they have such different languages and architectures. How could they understand one another, with so little in common? Communication is difficult enough between two different human tongues. But the signals used by the different portions of the human mind are even less likely to be even remotely as similar as two human dialects with sometimes corresponding roots. More likely, they are simply too different to communicate at all—except through symbols which initiate their use.

Now, one might ask, “*Then, how do people doing different jobs communicate, when they have different backgrounds, thoughts, and purposes?*” The answer is that this problem is easier, because a person knows so much more than do the smaller fragments of that person’s mind. And, besides, we all are raised in similar ways, and this provides a solid base of common knowledge. Even so, we overestimate how well we actually communicate. The many jobs that people do may seem different on the surface, but they are all very much the same, to the extent that they all have a common base in what we like to call “common sense”—that is, the knowledge shared by all of us. This means that we do not really need to tell each other as much as we suppose. Often, when we “explain” something, we scarcely explain anything new at all; instead, we merely show some examples of what we mean, and

some non-examples; these indicate to the listener how to link up various structures already known. In short, we often just tell “which” instead of “how”.

Consider how hard we find it to explain so many seemingly simple things. We can't say how to balance on a bicycle, or distinguish a picture from a real thing, or, even how to fetch a fact from memory. Again, one might complain, *It isn't fair to expect us to be able to put in words such things as seeing or balancing or remembering. Those are things we learned before we even learned to speak!* But, though that criticism is fair in some respects, it also illustrates how hard communication must be for all the subparts of the mind which never learned to talk at all—and these are most of what we are. The idea of “meaning” itself is really a matter of size and scale: it only makes sense to ask what something means in a system which is large enough to have many meanings. In very small systems, the idea of something having a meaning becomes as vacuous as saying that a brick is a very small house.

Now it is easy enough to say that the mind is a society, but that idea by itself is useless unless we can say more about how it is organized. If all those specialized parts were equally competitive, there would be only anarchy, and the more we learned, the less we'd be able to do. So there must be some kind of administration, perhaps organized roughly in hierarchies, like the divisions and subdivisions of an industry or of a human political society. What would those levels do? In all the large societies we know which work efficiently, the lower levels exercise the more specialized working skills, while the higher levels are concerned with longer-range plans and goals. And this is another fundamental reason why it is so hard to translate between our conscious and unconscious thoughts! The kinds of terms and symbols we use on the conscious level are primarily for expressing our goals and plans for using what we believe we can do—while the workings of those lower level resources are represented in

unknown languages of process and mechanism. So when our conscious probes try to descend into the myriads of smaller and smaller sub-machines which make the mind, they encounter alien representations, used for increasingly specialized purposes.

The trouble is, these tiny inner “languages” soon become incomprehensible, for a reason which is sunpie and inescapable. This is not the same as the familiar difficulty of translating between two different human languages; we understand the nature of *that* problem: it is that human languages are so huge and rich that it is hard to narrow meanings down: we call that “ambiguity”. But, when we try to understand the tiny languages at the lowest levels of the mind, we have the opposite problem—because *the smaller be two languages, the harder it will be to translate between them, not because there are too many meanings but too few. The fewer things two systems do, the less likely that something one of them can do will correspond to anything at all the other one can do.* And then, no translation is possible. Why is this worse than when there is much ambiguity? Because, although that problem seems very hard, still, even when a problem seems hopelessly complicated, there always can be hope. But, when a problem is hopelessly simple, there can’t be any hope at all!

Now, finally, let’s return to the question of how much a simulated life inside a world inside a machine could be like our ordinary, real life, “out here“? My answer, as you know by now, is that it could be very much the same—since we, ourselves, as we’ve seen, already exist as processes imprisoned in machines inside machines. Our mental worlds are already filled with wondrous, magical, symbol-signs, which add to everything we “see” a meaning and significance.

All educated people already know how different is our mental world from the “real world” our scientists know. For, consider the table in your dining room; your conscious mind sees

it as having a familiar function, form, and purpose: a table is “a thing to put things on”. However, our science tells us that this is only in the mind; all that’s “really there” is a society of countless molecules; the table seems to hold its shape, only because some of those molecules are constrained to vibrate near one another, because of certain properties of the force-fields which keep them from pursuing independent paths. Similarly, when you hear a spoken word, your mind attributes sense and meaning to that sound whereas, in physics, the word is merely a fluctuating pressure on your ear, caused by the collisions of myriads of molecules of air—that is, of particles whose distances, this time are less constrained.

And so—let’s face it now, once and for all: each one of us already has experienced what it is like to be simulated by a computer!

“Ridiculous,” most people say, at first: “I certainly don’t feel like a machine!”

But what makes us so sure of that? How could one claim to know how something feels, until one has experienced it? Consider that either you are a machine or you’re not. Then, if, as you say, you aren’t a machine, you are scarcely in any position of authority to say how it feels to be a machine.

“Very well, but, surely then, if I were a machine, then at least I would be in a position to know that!”

No. That is only an innocently grandiose presumption, which amounts to claiming that, “I think, therefore I know how thinking works.” But as we’ve seen, there are so many levels of machinery between our conscious thoughts and how they’re made that saying such a thing is as absurd as to say, “I drive, therefore I know how engines work!”

“Still, even if the brain is a kind of computer, you must admit that its scale is unimaginably large. A human brain contains many billions of brain cells—and, probably, each cell

is extremely complicated by itself. Then, each cell is interlinked in complicated ways to thousands or millions of other cells. You can use the word “machine” for that but, surely, no one could ever build anything of that magnitude!”

I am entirely sympathetic with the spirit of this objection. When one is compared to a machine, one feels belittled, as though one is being regarded as trivial. And, indeed, such a comparison is truly insulting—so long as the name “machine” still carries the same meaning it had in times gone by. For thousands of years, we have used such words to arouse images of pulleys, levers, locomotives, typewriters, and other simple sorts of things; similarly, in modern times, the word “computer” has evoked thoughts about adding and subtracting digits, and storing them unchanged in tiny so-called “memories”. However those words no longer serve our new purposes, to describe machines that think like us; for such uses, those old terms have become false names for what we want to say. Just as “house” may stand for either more, or nothing more, than wood and stone, our minds may be described as nothing more, and, yet far more, than just machines.

As to the question of scale itself, those objections are almost wholly out-of-date. They made sense in 1950, before any computer could store even a mere million bits. They still made sense in 1960, when a million bits costs a million dollars. But, today, that same amount of money costs but a hundred dollars (and our governments have even made the dollars smaller, too)—and there already exist computers with billions of bits.

The only thing missing is most of the knowledge we’ll need to make such machines intelligent. Indeed, as you might guess from all this, the focus of research in Artificial Intelligence should be to find good ways, as Vinge’s fantasy suggests, to connect structures with functions through the use of symbols. When, if ever, will that get done? Never say

“Never”.

VERNOR VINGE

A Hugo and Nebula Award finalist for *True Names*, he is also the author of *The Peace War*, *Grimm's World*, and a number of short stories. A mathematician and computer scientist, he has published articles in magazines such as *Omni*. He teaches at San Diego State University.

BOB WALTERS

His illustrations have graced the pages of SF magazines such as *Analog* and *Isaac Asimov's SF Magazine*. He has also done a great deal of scientific illustration for college texts, as well as general advertising illustration. He lives in Philadelphia, Pennsylvania.

MARVIN MINSKY

Considered by many to be the father of Artificial Intelligence, he has written especially for this book an essay on the nature of intelligence, natural and artificial. He is the director of the Artificial Intelligence laboratory at the Massachusetts Institute of Technology.