

Stored on: {phylum}<stefik>reviews>WF.TEDIT

Review of:

Winograd, T. & Flores, F. *Understanding computers and cognition: a new foundation for design*. Norwood, New Jersey: Ablex Publishing Corporation, 1986, 207 pages, \$xx.xx.

by Mark Stefik & Daniel G. Bobrow,  
Intelligent Systems Laboratory  
Xerox Palo Alto Research Center  
Palo Alto, California 94304

In their new book, Winograd and Flores consider computers, their possible uses, and the ways people think about them. The book is an uneven blend of rhetoric and insight; provocative but disappointing both for those who expect a carefully reasoned position, and those who seek an articulate account of the way things should be. It is strongest in its descriptions of philosophical issues that must be considered as part of the problem of understanding understanding.

This review is in two parts. It begins with a short review, which considers briefly the major themes of the book and how it succeeds with them. That is followed by a more leisurely review, which explores some of the interesting controversies in more depth.

### The Short Review

The position articulated by Winograd and Flores begins with the observation that many of the claims and predictions that we read about computers and AI systems are greatly exaggerated in advertisements and in the popular press. They feel problems evident in the advertisements actually go much deeper. These problems are characteristic of common notions about computers in our society that are both widely perceived and deeply rooted in our scientific ("rationalistic") tradition. To understand computers and the essence of cognition we should look to human communication and consider the ways that people interact. Social activity is the ultimate foundation of cognition. Finally, Winograd and Flores conclude, the appropriate roles for computers are in supporting people in the complex conversational structures generated within human organizations.

These reviewers resonate with much of this argument, for example that the likely near-term capabilities of computers have been misunderstood and overestimated by major institutions [8], that projects for developing AI and computer technology should be focused more on creating knowledge media [9] and technology for collaboration [10] rather than on

creating autonomous, intelligent agents. However, the book fails by confusing fragments of argument and rhetoric for substance. This is discussed further with examples in the next section.

We have problems however, with the description by Winograd and Flores of what they call the rationalistic tradition. This is important to their story, because they hold this tradition responsible for many wrong attitudes and misperceptions about computers. In this, the authors take bits and pieces from much longer unrelated works, and provide such an oversimplified account that it weakens the effect of their criticisms of the elements of this "tradition". Examples from philosophy and linguistics showing real language in action are used to decry the pretensions of early AI works, such as Winograd's own SHRDLU program, and are used to point out difficulties in building "intelligent" computers that can "understand." However, the authors neither review recent work that tries to deal with some of these problems, nor do they elaborate an alternative basis for understanding computers and cognition. They conclude that computer systems should *only* be designed as facilitators of communication, based partly on utility, and partly on what they espouse as the ultimate limits of computer capabilities.

What they have to say about limits is appropriate for current methods and technology, but we feel has little enduring value in understanding the where symbolic computation will eventually fail. Projections of the ultimate scope of a technology should be argued from carefully defined assumptions, and sources of fundamental limits, as found for example in Drexler's excellent discussion of the possibilities for "nano-technology" [2]. Without these, the arguments about why we should *always* treat and understand computers as different from people are not convincing. Certainly this is reasonable now, but that is a much weaker point.

### The Long Review

Winograd and Flores begin their account of computers and cognition with a discussion of what they call the rationalistic tradition. This is an inauspicious beginning, and perhaps the weakest part of the whole book.

### *Burning Straw Men*

Apparently the rationalistic tradition isn't any one thing. It is a strange composite that includes one rendition of the scientific method, some out-of-favor notions about how sentences convey meaning, the physical symbol system hypothesis in AI, and various kinds of reductionism. The authors offer a caveat that these elements do not constitute a tradition that is "uniformly accepted in carefully reasoned work of analytic philosophers", and so it is

not clear to whom they might ascribe this mess. In addition, they make little attempt to present a balanced or representative account of the relevant ideas.

An example of this is their myopic attention to Searle's position that sentences have a literal meaning, derived from the meanings of the words of which they are composed. Such an account of language is incomplete and misleading because it fails to account for the context and purposes of conversation. As the authors show, the whole gamut of "speech act theory" is full of examples of sentences that convey information other than their apparent literal meaning.

Winograd and Flores are secure in their use of Searle as a strawman, whose ideas on literal meaning are criticized within the AI community itself. It would have been more useful if they gave a more balanced and comprehensive account of current research directions. It would have been more interesting to read a critique of Grosz's work on comprehension using the context of a discourse, or Cohen's on interpretation of utterances as elements of a joint plan between speaker and listener. These are better examples to illustrate the richness of theory that is now more the norm in computer models of natural language comprehension. Continuing in this vein, the authors treat the obvious failures of the reductionist account of meaning as proof of inherent weakness in the rest of the rationalistic tradition. While it is apparent that scientists build models that they use for prediction and that technologists build computer simulations for similar purposes, it is not reasonable to claim that they always use them without cognizance of the limitations. When they dismiss the formation and testing of scientific hypotheses as being trivially and methodologically blind, they simply misunderstand the practice [1,7]. We felt an anti-technological bias in their description of science and the major contributions of scientists.

### *Symbol Blindness*

At the base of this rhetoric there is a dispute about the limitations of the use of symbols to model reality. Winograd and Flores believe that the difficulties (or impossibilities) of creating intelligent machines turn on these limitations. They quote Heidegger as asserting that "cognition is not based on the systematic manipulation of representations." This assertion seems to be a confusion of levels, a confusion of function with structure, or a confusion of a result with an implementation that achieves the result. It is like saying that computation is not based on the propagation of electrical signals, or that life is not based on organic chemistry. Although they don't say this directly, perhaps the level confusion is their point and that from their perspective, Newell and Simon's physical symbol system hypothesis [5] is also a confusion of levels. Connectionists may or may not agree [11] given

current accounts of symbols distributed among the neurons. The point deserves further discussion, but that won't be found in the book.

Another example of similar argumentation is in the description of Maturana's work. Maturana develops a vocabulary to describe coordinated activity that is independent of the use of any symbols (structural coupling is the key concept). As such, it provides a domain of explanation that is different than the symbolic descriptions of artificial intelligence. They rightly point out that it is unlikely that a baby uses a symbolic representation of its mother's goals and plans in deciding when to turn its head for milk or to cry. However, this does not preclude a symbolic representation at another level of description: one could describe the rooting reflex -- where a baby turns towards a touch on its cheek -- in terms of information flow. Even a phenomenon as complicated, and "biological" as color perception [4], may be usefully described as a symbolic process, even at rather early stages of processing including color recognition [3,6].

### *Ultimate Limits*

The author's attitude about computers seems to be that computers are valuable, even though they will never be intelligent. Symbol blindness (that is, the necessity to interpret the world within an impoverished set of concepts) is essential to their argument that computers can never be intelligent. They say:

"The program is forever limited to working within the world predetermined by the programmer's explicit articulation of objects, properties, and relations among them."

For intelligence, computers would either have to be programmed with a complete set of concepts (which is manifestly impossible), or they must be able to learn and evolve. And evolution just takes too long.

This story is rightfully discouraging, but it isn't air tight. We simply don't understand very much yet about learning programs. Nor is it reasonable to assume that "evolution" would need to go back to the beginning, whatever that might mean. In addition, a computer linked to the world, with the possibility of effective action and feedback, has the opportunity to learn its own new sets of concepts based on its structural coupling with the world. In this sense, computers could "grow up" in a world. Thus, although Winograd's SHRDLU never grew up, newer computer systems connected more intimately to the world might acquire background that we humans bring to bear to understand the world and to interact with each other. Of course, this is beyond the current state of the art.

### *A New Basis?*

The sub-title of the book promises a new basis for design. It does not deliver any formal framework -- that would probably be antithetical to the spirit of book). However, it provides some aphorisms and warnings, worded in the somewhat stilted style of Heidegger. "In creating tools we are designing new conversations and connections." (Computer tools shape the possible communications in an organization. Sometimes they are too rigid.) "Domains of anticipation are incomplete." (You can't predict everything. Design should proceed in a cycle with feedback from users.) "Breakdown is an interpretation -- everything exists as interpretation within a background." (The requirements of success for computer systems are perceived differently by different members of an organization.)

Winograd and Flores provide an example of a new system built with the new premises in mind. A specialized electronic mail system, it categorizes messages in one of a small number of speech acts (request/promise, offer/acceptance, report/acknowledgement), and has required entries for each. Their claim is that using this will help productivity in organizations.

The design is surprising in several ways. First, it uses vocabulary associated with rich and complex human interactions (e.g promises) with very specific technical meaning. Although they acknowledge this explicitly, it seems to violate the spirit of their use of vocabulary. Secondly, although breakdowns are a crucial part of their way of thinking about systems, there is no notion described for the coordinator of how to deal with breakdown. Handling this in the redesign cycle for the system (as suggested later) is not very satisfying. They conclude with an admonishment that we be aware of what our design process is, and how a system can change our way of working. That is of course very good advice.

### *Computers aren't People, and Vice Versa*

In the last part of the book, the authors come to their fundamental observations about why we should always treat and understand computers as different from people. They assert that computers cannot make commitments or accept responsibility. As in their discussion of intelligence, they make the jump from reasoned arguments to rhetoric. They simply state that computers cannot make commitments, have responsibility, or be intelligent. In this they ignore the lesson they preach in the rest of the book. Terms like promise, commitment and responsibility are mutually defined and used by people in different ways depending on circumstances. The appropriateness of the terms with respect to people and computers must be examined in the particular contexts. We saw previously how the common notion of promise was narrowed through the use of the coordinator system. Let us consider how the term commitment may have application that usefully includes a computer.

Consider the scenario of buying an airline ticket from a vending machine at the airport, only later to discover that the flight was over-booked. The question is, did the ticket machine make a commitment? Presumably Winograd and Flores would say no because only people make commitments. These reviewers suspect that this would not hold up as a defense in a court of law. It is useful to say the ticket machine did make a commitment, just as we would for a human clerk in such a role; the commitment was done for the airline. We wouldn't sue a ticket machine for overbooking, nor the human ticket clerk. We'd sue the airline.

This airline example might make Winograd and Flores uneasy, as an example of the loss of "sense of responsibility in modern society". A more complete consideration of the situation would show that such legal practice balances several different notions of justice. Corporate law may seem to make it possible for the rascals to get away with it in some cases, but it also provides stable accountability through changes of management. The point is that commitment and responsibility are complex notions with social and legal dimensions. The treatment of them in this book dismisses them without any substantial consideration of their fundamental properties.

Given an extended discussion of commitment, is it reasonable to expect that we will ever want to think of computers as making commitments in the same way that people do? Asimov's R. Daniel Olivaw (spelling?) is a robot of the future that must delicately balance his actions with the goals and needs of the people around him (it?). One can ask whether such robots are possible; to be useful such robots would need to be able to give an account of their actions and to act as if they make human-like commitments. Would it still be useful to distinguish between their commitments and those of humans?

## Conclusion

These reviewers agree with the warnings posted by Winograd and Flores that it is important to better understand the uses and limits of computer technology. Their examples from Heidegger point out that computers are embedded in the world and must act as well as reflect. However, the authors fail to give effective prescriptions for the future. They also fail to characterize the ultimate shortcomings of technology in terms of fundamentals.

For example, in their analysis of the futility of artificial intelligence, it is never clear exactly where symbols, search, and sufficient computational power must ultimately fail. Any single example of a computer acting appropriately is not proof of intelligence; but neither is any single example of creative action by humans proof that something is beyond machines. AI has nothing analogous to Hilbert's famous problems. Perhaps we shall one day develop classes of complexity for computational beings that will relate in natural ways

to Piaget's stages of development, or to some other measure of knowledge, familiarity with environment, and maturity.

In conclusion, is the book worth reading? It is odd that a book can be so annoyingly flawed, and yet contain many worthwhile insights. On pondering this, these reviewers recommend that the book be read twice. The first reading is to react to all of the half-arguments and rhetoric, to get over being provoked by them and the somewhat anti-technology stance, and to accept that the book is strictly about the near future. The second reading is to skip or discount those sections and pay attention to the rest.

- 
1. Dobzhansky T., Ayala F.J., Stebbins G.L., Valentine J.W. *Evolution*. W.H. Freeman and Company (especially Philosophical Issues, Chapter 16), 1977.
  2. Drexler, K.E. *Engines of Creation*. New York: Anchor Press/Doubleday 1986 (in press).
  3. Horn B.K.P. *Robot Vision*. Cambridge, Massachusetts: The MIT Press, 1986.
  4. Nathans J., Thomas T., Hogness, D.S. Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science* 232 pp. 193-210, April 1986.
  5. Newell, A. and Simon H.A. Computer science as empirical enquiry: symbols and search, *Communications of the ACM* 19:3, pp. 113-126, March 1976.
  6. Pentland, A. (ed.) *From Pixels to Predicates: inference of world knowledge from visual data*. Norwood, New Jersey: Ablex Publishing Corporation, 1986.
  7. Platt, J. R. Strong inference, *Science* 146:3642, pp. 347-353, 1964.
  8. Stefik, M. Strategic computing at DARPA: an assessment. *Communications of the ACM* 28:7 pp. 690-704, July 1985.
  9. Stefik, M. The next knowledge medium. *AI Magazine* 7:1 pp. 34-46, Spring, 1986.
  10. Stefik, M., Foster, G., Bobrow, D.G., Kahn, K., Lanning, S., Suchman, L. Beyond the chalkboard: using computers to support collaboration and problem-solving in meetings. (in preparation), 1985.
  11. Touretzky T.S., Hinton G.E., Symbols among the neurons: details of a connectionist inference architecture, *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, pp. 238-243, August 1985.