

Book Reviews

J. Christopher Maloney, *The Mundane Matter of the Mental Language*, Cambridge University Press (1989), xxvii + 274 pages.

Readers who have followed debates on mental representation through the paths charted by such luminaries as Fodor, Dennett, Stich, Putnam, Searle, and the Churchlands will recognize many familiar themes in J. Christopher Maloney's contribution to the topic. Maloney has produced an admirably clear discussion of some key issues, and, while his own views and arguments generally owe a large debt to his precursors, he nonetheless manages to stake out a position with some distinctive features of its own.

The book's title actually promises less than its contents deliver. Yes, the "matter of the mental language" – in both senses – commands considerable attention, though in neither sense is it "mundane". But Maloney also comprehensively treats a variety of interrelated topics: modularity, rationality, qualia, the frame problem, and the origin of mental contents. Needless to say, any book that advances philosophical discussion of these issues cannot help but provoke substantial controversies. Here I shall highlight just a few of them.

Sententialism

Taking folk psychology as its point of departure, Sententialism maintains that, for beliefs to be causally implicated in the production of rational behavior, they must constitute a series of (internal) physical states that stand to behavior in the relation of premises to conclusion. The causal efficacy of beliefs derives from certain of their physical (syntactic) properties, while their contribution to rationality centrally involves the fact that *syntactic* differences reflect different *contents*.¹ Thus, beliefs are regarded as sentences, physically encoded.

Although this characterization might suggest that beliefs are capable of arising in all sorts of entities, Maloney urges that intentionality is present only when the underlying matter is of the *right sort*.² (He agrees with Searle that instantiating a formal program is not by itself sufficient for mentality.) In humans, and probably many other animals, belief-sentences are hypothesized to be encoded neurally in "Mentalese". It does not follow, however, that *neural* stuff affords the physical basis of intentionality, for neural-based Earthlings, along with non-neural Martians and various artificial devices, could conceivably share (as yet unknown) physical constitutions supportive of intentionality.

One advantage claimed for this viewpoint is that it averts the problem of infinitely embedded homunculi: Mentalese representations ultimately have con-

tent for an agent, not by being interpreted (which would require yet another interpreting agent), but by springing from a particular kind of material. At the same time, as Maloney recognizes, the proposal that intentionality arises naturally *and exclusively* from “the right stuff” courts psychological chauvinism, for it introduces the possibility that various entities, despite whatever behavior they might exhibit, may be deemed bereft of mentality. Maloney does not propose to resolve that issue speculatively, however. What he puts forward instead is the hypothesis that intentionality is a natural kind, the material makeup of which is to be illuminated by future scientific research. The results could be quite surprising:

It is consistent with the prevailing hypothesis to suppose that the study of cognitive processes will someday yield a radically new taxonomy of physical states according to which organic and inorganic devices are comprehended by a natural physical kind, membership in which ensures the possession of contentful mental states. (p. 177)

Of course, it may turn out that no such material basis of cognition is to be found. In any case, although Maloney is rather vague about the details of that envisioned research (the issue is briefly addressed on pp. 73n–74n), he does not appear to think that it will undermine the importance of the neurosciences in studying cognition.

Notation and Content

According to Sententialism (pp. 33–46, 124–139 *passim*), sameness of belief is determined syntactically.³ The argument is that (a) the same beliefs should produce the same behavior, while differences in belief should make a behavioral difference; (b) syntactic sameness and difference is presumably reflected in neural codings, thus yielding a low-level causal account of behavioral sameness and divergence.

But a syntactic criterion for sameness of belief appears to conflict with a behavioral criterion, as the following example suggests (p. 34). Suppose that computer scientists someday create two (notationally distinct) programs – call them ABELARD PRIMUS and ABELARD SECUNDUS – both of which drive devices that perfectly simulate the behavior of Abelard in his affair with Héloïse.⁴ Suppose further that the notation employed by SECUNDUS differs significantly from PRIMUS’s notation (perhaps they involve different higher-level programming languages). PRIMUS in fact may not reflect the beliefs of Abelard, but, even if it did, the existence of SECUNDUS would show that syntactically divergent representations can produce identical behavior. Sameness of behavior argues for sameness of beliefs. Yet this result seems to contradict Sententialism’s insistence upon a syntactic criterion for belief sameness, since PRIMUS and SECUNDUS are different programs.

Suppose now that the programs are compiled into machine language. Maloney asks whether the resulting compilations produce the *same* machine language. If so, he argues, we don’t really have alternative notations after all.⁵ It is not clear

how Maloney is construing “same machine language” here, but it would surely be too strong to insist that two programs are compiled into the same machine language only if their resulting binary codes are identical. (Compare: “*Hamlet* and *King Lear* are translated into Morse code only if the resulting strings of dots and dashes are identical.”) Assume that PRIMUS and SECUNDUS run on identical hardware. The compilations of their higher-level program code will then yield non-identical machine-code – *in the same machine language*. So, *pace* Maloney, it appears that PRIMUS and SECUNDUS might well operate with syntactically divergent representations that cause identical behavior.

But is the example of the three Abelards ultimately consistent? Maloney suggests that it may not be. Beliefs with different contents, he insists, must issue in some behavioral differences as well. To illustrate, he proposes that a “tracer” program be attached to the executing programs of both PRIMUS and SECUNDUS. (The tracer simply reports on what code is executing in each, e.g., “A is now being implemented” or “B is now executing”.) Since PRIMUS and SECUNDUS contain notational differences, their respective tracers will yield different reports at different stages of program execution. The tracers’ reports, Maloney argues, demonstrate behavioral differences “latent” in the operations of the two programs.

It is not clear that this scenario advances the Sententialist’s case, however. The example of PRIMUS and SECUNDUS, after all, was introduced to show that identical behavior of cognitive agents need not stem from identical beliefs. The behavior initially to be simulated – aspects of Abelard’s affair with Héloïse – was far removed from the “latent behavior” revealed by tracer programs. Furthermore, even with the introduction of tracers, the only behavioral differences appearing are something akin to introspective reports concerning Mentalese structures; yet, since Sententialism does not presuppose that cognitive agents have introspective access to those structures (p. 19), the “latent behavior” (revealed by the tracers) in PRIMUS and SECUNDUS seems irrelevant.

Maloney’s example of PRIMUS and SECUNDUS pointed to a conflict between syntactic and behavioral criteria for sameness of Mentalese content. Ultimately, however, Sententialism does not advocate a simple behavioral criterion for content. Rather, in endorsing what it calls “doxastic holism” (pp. 124–139; cf. 202–219), Sententialism maintains that the content of a Mentalese token is fixed, not in isolation, but by taking into account its causal transactions with other mental structures, along with behavior in which it is implicated – in short, a functional specification of content. Even so, a conflict similar to the one drawn above can be generated. Consider two cognitive agents, A and B, both of whom hold the same belief (e.g., that John F. Kennedy was assassinated). Assume that both agents acquired their beliefs under similar circumstances, perhaps by watching the same television news broadcast. Now, it seems perfectly possible that the Mentalese token *T*, which in A records that belief, might also occur somewhere in B – yet, a *functional* specification could well assign *different* content to B’s token

*T.*⁶ But, still supposing **A** and **B** to hold the same belief, not only does the syntactic criterion require that **A** and **B** both token identical Mentalese sentences, Sententialism also insists that both sentences have the *same* content. So, at this juncture, the syntactic criterion for belief sameness and the functional specification of belief content appear to be at loggerheads. Should we seek to reassign content in the mental structures of one agent in order to bring the two criteria into alignment?⁷ If so, which one should we change, and why? Perhaps, instead, if content is to be preserved, the syntactic criterion should be abandoned.

Original Meaning

Since Mentalese is not taken by Sententialism to be purely syntactic, an explanation of its semantic force must be provided. One well-known account of mental content claims that the meanings of Mentalese terms are endogenous, not acquired. Maloney eschews this path, opting instead for a causal theory of content. The first task, then, is to show how content can initially arise without presupposing antecedent contentful states.

Not surprisingly, the story begins with sensation; sensory states are regarded as Mentalese tokens. (Although this point is not argued for explicitly, it seems necessary, granting that we can form beliefs about what we perceive.) Suppose, for example, that an apple's being red causes Abelard to be in physical state *S* (pp. 181ff). *S* turns out to be a complex state, involving subject and predicate. Although in English we might describe the subject as "that apple" and the predicate as "being red", Maloney urges that purely sensory states are much more semantically attenuated than the English expressions suggest. The subject of a *sensory* representation is akin to a demonstrative such as "this", while the associated predicate is equally impoverished – something akin to "that". At this first level of representation, then, neither the apple nor the property of being red is represented as such. Rather, the sensory mechanism isolates a subject and property, and predicates the latter of the former, without having categorized either (e.g., "This is that"). Sensory demonstratives refer and sensory predicates predicate "in the most pristine way without assistance from other lexical items" (p. 184).

Presumably, however, other pristine demonstratives and predicates can render assistance. How else could the referent of "this" be isolated from many other possible referents in a sensory array? Furthermore, since pristine predications occur in a variety of circumstances, in which various properties are assigned to various subjects, each pristine predication must have a unique syntax – otherwise, they would all have the same content (cf. Note 1). Maloney accommodates this point by proposing that Mentalese contains "a unique demonstrative referential type for each object of sensuous reference" (p. 190).⁸ But while this proposal may circumvent the problem just mentioned, it does so by burgeoning the vocabulary of Mentalese beyond plausibility.⁹ (Consider the effect of introducing a new demonstrative in English for each object of reference.)

Turning now to perhaps the most fundamental question, wherein lies the *referential* character of Mentalese demonstratives and predicates? According to Maloney, Mentalese demonstratives

refer simply by virtue of being caused to occur, and they select as their referents the objects that cause them to occur without regard for any other characteristics these objects might possess (p. 184).

Pristinely attributed properties are, yes, *given* in sensation (p. 187n).

Thus, the ultimate semantic basis of Mentalese, it appears, is a kind of referential primitive, simply to be *accepted*. But then is there really any reason to abjure nativism, at least insofar as the most referentially frugal Mentalese sentences are concerned? In the final analysis, both nativism and “the given” appear to leave the source of reference equally unexplained.

The claim that some contents are simply “given”, of course, drives a wedge between sensuous and non-sensuous Mentalese representations, since the contents of the latter were said to be fixed holistically through functional specifications. At the same time, Maloney argues that the contents of the one ultimately give rise to the contents of the other. It is hard to know at this point whether the bifurcation just noted creates a significant difficulty, since he offers just a sketch of how the transformation from sensuous to non-sensuous representation might take place. Further investigation of this topic would be welcome. (Computational models of vision, while they may provide some important guidance in this area, are still very much in their infancy.)

Appealing to “the given” does raise a certifiable objection, however, since a familiar contrary doctrine maintains that observation terms are *not* purely given, but are always theory-laden. Maloney’s reply is as follows:

But pristine attributives are so bereft of semantic force that it is hard to see how they could be subject to the influence of believed theories and spoken, entrenched languages. (p. 187n)

Clearly, there is not much room to maneuver here. On the one hand, these initial Mentalese attributions, semantically impoverished though they may be, must have content sufficient to permit cognitive agents ultimately to “bootstrap” their way up to the full range of their mental representations. On the other hand, those same pristine beginnings are said to be so devoid of content that they are immune to the difficulties just outlined.

Additional Topics

Beyond the themes briefly touched upon in this review, there are many others deserving of attention. Among them are (1) the range of an agent’s actions (Maloney’s account seems rather promiscuous), (2) conceptual development and the question of ambiguity in Mentalese (Maloney’s repeated appeal to *degrees* of Mentalese conceptual mastery appears to concede a fair amount to his objectors), and (3) Nagel’s worries about “objective” knowledge of “subjective” experience (Maloney argues that knowing the syntax of a bat’s Mentalese, as it is realized in a bat’s sensory system, would tell us all there is to know about the qualia of

bats – although this would admittedly not produce in us *experiences* had by bats).

Anyone who has pondered these and related matters from alternative perspectives will find much in Maloney's book to respond to; readers seeking a survey of the current intellectual landscape will find it a stimulating and highly responsible guide

Notes

¹ "Thus, Sententialism abjures strict synonymy in Mentalese. It insists that within a mental language every *notational* difference amounts to a semantic difference" (p. 42n).

² This does not mean that matter determines the content of particular states: "The matter of a state may be what renders it intentional; its functional role may be at least part of what settles its content" (p. 29).

³ This claim is relativized to languages. Maloney suggests that differences in mental languages may distinguish cognitive kinds. Since the same representation, notated differently, might appear in different kinds of agents, functional considerations would presumably be required to individuate beliefs across cognitive kinds (cf. p. 182).

⁴ This thought experiment has a real-life analogue already causing a legal stir: court cases currently abound as software companies complain that competitors have produced programs whose operation is identical to their own, even though the underlying code may be considerably different.

⁵ He also argues that if *different* machine language results from each of the two compilations, we would have something tantamount to different kinds of cognitive agents. In that case, notational differences would not automatically imply differences in belief. (See Note 3.)

⁶ In that case, of course, **B**'s belief about JFK would be carried by some other token, say *U*.

⁷ An alternative ploy would be to claim that, despite all appearances so far, the two Mentalese tokens are nevertheless of the *same* notational type and (consequently) have the same content. Functional specification, then, would severely constrain what counts as notational variation. Maloney nods briefly in that direction at one point (p. 42), but, significantly, his discussion elsewhere (e.g., pp. 124–139) does not seem to presume such a strategy.

⁸ A similar strategy, incidentally, is introduced to handle indexicals (p. 137).

⁹ Actually, the problem is far worse than it first appears. Not only do demonstratives multiply out of control, but pristine predicates may do so even to a greater degree. Each pristine predication is associated with a distinct Mentalese predicate type (p. 191). But, further, pristine predicates at the level of Mentalese may be far more discriminating than predication in English would reveal. For example, it might turn out that no two objects are exactly the same shade of red; if so, and if each different shade causally produced a different type of sensory state, a different pristine predicate would be applied to each shade (p. 187n).

Department of Philosophy,
University of Michigan-Flint

CHARLES E. M. DUNLOP

Philip R. Cohen, Jerry Morgan, and Martha E. Pollack (eds.), *Intentions in Communication*, Cambridge, MIT Press, 1990, x + 508, \$45.00.

Intentions in Communication is the proceedings of an interdisciplinary workshop held in Monterey, California, in March 1987. The collection brings together the

work of prominent researchers in computer science, artificial intelligence, linguistics, philosophy, and psychology whose theories of communication rely on a careful account of the intentions and plans of the communicating agent. Thirteen papers that were presented at the workshop are included in this volume, along with seven commentaries that are used to integrate the work by comparing and contrasting pairs of papers. The one exception is the work of Janet Pierrehumbert and Julia Hirschberg, which is reviewed separately because it represents a departure from the rest. A fourteenth paper, by Herbert H. Clark and Deanna Wilkes-Gibbs was added to the collection because it was relevant to one of the workshop topics.

The editors have written an introductory chapter which includes a brief background on work in planning, action, and communication. This work forms a continuum not only with respect to time, but also approach, stretching from traditional work in philosophy and linguistics on theoretical semantics to relatively recent work in artificial intelligence and psychology on discourse understanding. The editors summarize this research as addressing different subsets of six central questions with respect to communication and discourse. The first question concerns characterizing meaning in a way that is appropriate for all kinds of expressions. The second question, of compositionality, pertains to relating the meaning of a complex expression to the meanings of its parts. The third question, action, is concerned with how the utterance of a sentence succeeds in performing acts in addition to the act of saying. The answer to this is perhaps related to the answer to the fourth question, that is, how it is that the utterance of a sentence can convey something more than, or different from, its literal meaning. The fifth question concerns discourse compositionality, how it is that the meaning of a connected discourse is more than the sum of the meanings of the sentences that compose it. Finally, the answers to these questions are related to answering a sixth question: What is communication?

The editors propose that any viable theory of communication must address all six questions. Moreover, to achieve this, such a theory will have to take into account the role of intention. Speaker intention functions as a guide for planning what to say; for the hearer, it serves as an underlying assumption for interpreting what has been said. The background material helps the reader to focus on the issues this research is concerned with, but it is insufficient to make the reading accessible to someone unfamiliar with the area. The second half of the introduction clarifies the organization of the collection into three sections. This is good, since the organization is not apparent from the physical layout. Within each section, a comprehensive overview of each of the papers is given, emphasizing the relationships of key ideas between successive papers. This helps to create a transitional ordering of papers from the beginning of each section to the end.

The papers in the first section are concerned with characterizing intention. The first paper, "What is Intention?" by philosopher Michael Bratman, examines the role that intentions play in focusing and constraining the reasoning of an agent. In

the following paper, "Persistence, Intention, and Commitment", Philip R. Cohen and Hector J. Levesque use many of Bratman's notions in a model presented as a modal, first-order logic. The third and fourth papers change the focus from planning to plan recognition. In the third paper, "Plans as Complex Mental Attitudes", Martha Pollack develops her theory of plan recognition around a model for the complex mental attitude of "having a plan". She suggests that an adequate theory of cooperative communication must concern itself with the structure of this attitude. The final paper of this section, "A Circumscriptive Theory of Plan Recognition" by Henry Kautz, presents a formal theory of plan recognition using circumscription. The analysis is aimed at unintended plan recognition, but Kautz argues that his approach applies to intended plan recognition as well, the kind found in speech-act recognition.

The traditional view of sentence meaning holds that the semantic content of a sentence is its truth value (Tarski, 1956). In contrast, speech act theory suggests that speech is best viewed as purposeful action used by speakers to change the mental state of another agent (Austin, 1962). The second section comprises seven papers (half the book) spanning the full range of research approaches to communication. The common thread among the papers is that they all move in the direction of resolving the two interpretations of sentence meaning. I will mention four of these papers here. In "An Application of Default Logic to Speech Act Theory", C. Raymond Perrault presents a model of discourse in which the effects of speech acts are modeled as defaults in a nonmonotonic logic. For example, that a speaker believes what he asserts is a default assumption. "On the Unification of Speech Act Theory and Formal Semantics", by philosopher Daniel Vanderverken, attempts to relate the speech-act properties of sentences to their truth-conditional ones by distinguishing two sets of semantic values, those relating to the success of the spoken sentence and those relating to the truth conditions of the propositional content of the sentence. In their second paper of this volume, "Rational Interaction as the Basis for Communication", Cohen and Levesque argue that what is communicated in an utterance is a propositional attitude expressing the speaker's mental state. In a slight departure from the rest, "The Meaning of Intonational Contours in Discourse", by Janet Pierrehumbert and Julia Hirschberg, investigates the role that intonation contour plays in analyzing the beliefs and attitudes of a speaker.

To underscore the volume's theme, a short third section includes three papers that return to the issue of the nature of intention in communication. Unlike the first section, these three papers emphasize that discourse is a collective act involving both the speaker and hearer. In the first paper, "Collective Intentions and Actions", philosopher John Searle theorizes that collective behavior is not the summation of the individual behaviors of the participants. Instead, he argues that collective behavior is guided by collective intentions. Furthermore, he suggests that due to this collective intentionality, each agent intends to achieve the collective goal by having an individual intention to do his part. In the next paper, "Plans for Discourse", computer scientists Barbara Grosz and Candice

Sidner develop a formal analysis for joint plans, which differs from Searle's in that two agents mutually believe that each agent intends to do his part and that each agent will do his part if and only if the other agent does his. Finally, in "Referring as a Collaborative Process", psychologists Herbert H. Clark and Deanna Wilkes-Gibbs supply data from a referential communication task. They argue that their results support the claim that referring is a collaborative action involving the hearer as well as the speaker.

Some excellent commentaries help the reader to focus on the main ideas in these papers by carefully comparing approaches, highlighting relative strengths and weaknesses, and thus identifying issues for further research. A good example is James Allen's discussion of Michael Bratman's process account for the intended and unintended consequences of plans. Allen compares it to the model of intention proposed by Cohen and Levesque in which intention is a modal operator. He suggests how Bratman's theory can be made more precise, and he raises issues for Cohen and Levesque to address so that their model provides better coverage. A few commentators use the papers they critique as a springboard for presenting their own work. This results in some interesting papers in their own right. For example, noting some difficulties with the paper on speech acts by Vanderverken, and the second paper by Cohen and Levesque on the same, Jerrold Sadock presents a new scheme for understanding the structure of, and the relationships among, illocutionary acts.

Collections of this kind are useful because they provide a broad snapshot of the work in an area. The book purports to be interdisciplinary, addressing the issue of intention as it arises in different areas of research in communication and discourse. In reality, the collection seems to concentrate on approaches involving formal semantic theories. This is clear from the list of contributors, which includes eight philosophers and linguists and three researchers in computer science and artificial intelligence who rely heavily on the philosophical literature. In fact, only two papers directly use empirical research, and these are not integrated with the other papers (the papers by Pierrehumbert and Hirschberg and by Clark and Wilkes-Gibbs). However, this volume is essential reading for anyone in any discipline who is interested in speech act planning, speech act recognition, and discourse interpretation.

References

- Austin, J. L. (1962), *How to Do Things with Words*, NY: Oxford University Press.
 Tarski, A. (1956), *Logic, Semantics, and Meta-mathematics*, Oxford: Clarendon Press.

*Department of Computer Science
 and Center for Cognitive Science,
 State University of New York at Buffalo*

SUSAN M. HALLER

Tom Forester and Perry Morrison, *Computer Ethics*, Cambridge, Massachusetts: The MIT Press, 1990, 193 pages.

Computer Ethics is a welcome addition to the growing number of books that address ethical issues in computing. The subtitle of the book, "Cautionary Tales and Ethical Dilemmas in Computing", summarizes the contents of the book well. The book is largely a collection of up-to-date stories of malfunctions and misuses of computers. *Computer Ethics* is in the category of consciousness-raising books. The primary purpose of the book is to send a wake-up message to those, particularly computer science students, who think there are no ethical issues in computing. As the authors state, the book has "... the simple aim of sensitizing undergraduate computer science students to ethical issues." (p. viii)

The book is sensibly organized around standard themes. Following an introductory chapter, the book has chapters on computer crime, software theft, hacking and viruses, unreliable computers, invasion of privacy, AI and expert systems, and computerizing the workplace. An appendix contains a discussion of "Star Wars", the Strategic Defense Initiative. Each of the chapters contains a bevy of relevant examples and some discussion of the issues. The chapter on hacking and viruses gives a nice list of definitions of terms, such as "Trojan horse", "logic bomb", "virus", "vaccine", "worm", and "tempest", which might be familiar to, but not clearly understood by, some students. The book is easy to read and contains numerous interesting tidbits of information. Did you know that, at least according to one survey, the most common passwords in Britain are "Fred", "God", "Pass", and "Genius", as well as names of spouses and family pets, whereas in the United States the favorite password is "Love", closely followed by "Sex"? (p. 21)

The teaching strength of the book lies in its presentation and discussion of numerous contemporary examples that most students should find relevant and engaging. The book is more current than other recent books such as *Silicon Shock* (Simons, 1985) and *The High Cost of High Tech* (Siegel and Markoff, 1985). However, for number of examples and depth of analysis on just the issues of privacy and surveillance, *The Rise of the Computer State* (Burnham, 1983) is still the best choice.

Another teaching strength of the book is the presentation of a hypothetical case study at the end of each chapter. These cases are well-written and ought to generate good class discussions. The authors clearly have this in mind and suggest that members of the class might do some role-playing in some of the cases. *Ethical Conflicts in Information and Computer Science, Technology, and Business* (Parker, Swope, and Baker, 1990) has more hypothetical case studies, but the hypothetical cases in *Computer Ethics* are better developed and are designed, of course, for discussion of the issues raised earlier in each chapter.

One weakness in the book is the absence of a coherent ethical theory and its application to the ethical problems. As the authors readily admit, "Readers will

notice that we have not adopted an explicit theoretical framework and have avoided philosophical discussion of ethical theory.” (p. viii) It is somewhat strange that a book on computer ethics would omit a discussion of something as central as ethical theory, but, in fairness, this omission seems to be more the rule than the exception in books on computer ethics. Certainly, the book should be supplemented with some discussion of ethical theory and perhaps used with a book like *Computer Ethics* (Johnson, 1985), which does integrate ethical theory with ethical issues in computing.

Computer ethics is not simply the application of ethics to computing but requires conceptual analyses of computing situations. Different conceptions can lead to different ethical judgments. To its credit, *Computer Ethics* emphasizes some of the differences created by computers. For example, as the authors discuss, breaking into a computer system is not completely analogous to a typical physical break-in (p. 60). Normally, nothing is broken as one breaks into a computer system; nothing like a physical lock need be damaged. One can steal information without removing anything; all of the original information is still there. One can browse through files without disturbing the contents. If nothing is altered, who is harmed by such activity? Of course, understandings of computer break-ins that stress similarities to typical physical break-ins and harm generated are equally possible. It is wise to consider the merits of different conceptualizations, for differences in ethical judgments may depend as much upon differences in how situations are conceived as upon differences in ethical values. Getting clear about how to conceptualize a computing situation is half the job of computer ethics.

Computer Ethics has a casual, almost journalistic, style that makes it easy to read. However, in some places, the style seems too loose for careful analysis. For example, in a discussion of the nature of intelligence and AI, the authors say, “. . . it has been claimed by some that both Joseph Weizenbaum’s program, ELIZA, which simulates a Rogerian psychotherapist’s conversation, and Kenneth Colby’s PARRY program, which simulates a paranoid schizophrenic, are able to satisfy the Turing test.” (p. 116). As far as I know, nobody claims that both ELIZA and PARRY can pass the Turing test or that either one can. Joseph Weizenbaum was impressed with how some people got taken in by ELIZA, but that’s not passing the Turing test (cf. Weizenbaum, 1976). Kenneth Colby states very clearly that his tests for PARRY are modifications of the Turing test restricted to psychiatric parameters, and he doesn’t believe that PARRY would pass a regular Turing test (cf. Colby and Hilf, 1973). These programs are not a basis for any ambiguity about the Turing test, contrary to what the authors suggest.

In another place, the authors say “. . . there may be at least one way in which AI could be unambiguously (in our minds anyway) considered an improper goal for society – that is because of its funding base and clear links with the military establishments of both the US and the UK.” (p. 130) Unless the authors wish to

defend pacifism, which they don't do, why is the funding base and links with the military a sufficient ground for classifying AI as an improper goal for society? What really bothers the authors, as I think it should, is the proposed use of AI to operate extremely complex systems, like those of the Strategic Defense Initiative. Putting an unproven control system in charge of nuclear defenses could easily lead to disastrous consequences. But this kind of argument is most forceful when it is given as an empirical argument about the probability of malfunction rather than as a purely antimilitary argument. I certainly share the authors' concern about the introduction of AI on the battlefield. Another worthwhile, recent book that critically examines the military use of computers is *Computers in Battle* (Bellin and Chapman, 1987).

Although I have some philosophical quibbles with *Computer Ethics*, I have no hesitation in recommending it as a good source of examples for a course or part of a course on computer ethics. Consciousness-raising is its stated aim, and it succeeds admirably.

References

- Bellin, David and Chapman, Gary (eds.) (1987), *Computers in Battle: Will They Work?* (Boston: Harcourt Brace Jovanovich).
- Burnham, David (1983), *The Rise of the Computer State* (New York: Random House).
- Colby, Kenneth and Hilf, Franklin (1973). "Multidimensional Analysis in Evaluating a Simulation of Paranoid Thought", *MEMO AIM-194* (Menlo Park, CA: Stanford University Dept. of Computer Science). [Editor's note: Cf. also Colby, Kenneth Mark (1981), "Modeling a Paranoid Mind", *Behavioral and Brain Sciences* 4: pp. 515-560.]
- Johnson, Deborah G. (1985), *Computer Ethics* (Englewood Cliffs, NJ: Prentice-Hall).
- Parker, Donn B., Swope, Susan, and Baker, Bruce N. (1990), *Ethical Conflicts in Information and Computer Science, Technology, and Business* (Wellesley, MA: QED Information Sciences).
- Siegel, Lenny and Markoff, John (1985), *The High Cost of High Tech: The Dark Side of the Chip* (New York: Harper & Row).
- Simons, Geoff (1985), *Silicon Shock: The Menace of the Computer Invasion* (New York: Blackwell).
- Weizenbaum, Joseph (1976), *Computer Power and Human Reason* (San Francisco: W. H. Freeman).

*Department of Philosophy,
Dartmouth College*

JAMES MOOR