Human-Computer Interaction

ISSN: 0737-0024 (Print) 1532-7051 (Online) Journal homepage: https://www.tandfonline.com/loi/hhci20

Softening Up Hard Science: Reply to Newell and Card

John M. Carroll & Robert L. Campbell

To cite this article: John M. Carroll & Robert L. Campbell (1986) Softening Up Hard Science: Reply to Newell and Card, Human-Computer Interaction, 2:3, 227-249, DOI: 10.1207/s15327051hci0203_3

To link to this article: https://doi.org/10.1207/s15327051hci0203_3

Published online: 11 Nov 2009.

HUMAN-COMPUTER INTERACTION, 1986, Volume 2, pp. 227-249

Softening Up Hard Science: Reply to Newell and Card

John M. Carroll and **Robert L. Campbell**
IBM Thomas J. Watson Research Center

ABSTRACT

A source of intellectual overhead periodically encountered by scientists is the call to be "hard," to ensure good science by imposing severe methodological strictures. Newell and Card (1985) undertook to impose such strictures on the psychology of human-computer interaction. Although their discussion contributes to theoretical debate in human-computer interaction by setting a reference point, their specific argument fails. Their program is unmotivated, is severely limited, and suffers from these limitations in principle. A top priority for the psychology of human-computer interaction should be the articulation of an alternative explanatory program, one that takes as its starting point the need to understand the real problems involved in providing better computer tools for people to use.

Authors' present addresses: John M. Carroll and Robert L. Campbell, User Interface Institute, IBM Thomas J. Watson Research Center, P.O. Box 704, Yorktown Heights, NY 10598.

CONTENTS

1. NEWELL AND CARD ON BEING HARD

Newell and Card (1985) presented a program for psychological research in human–computer interaction couched as an analysis of how psychology can avoid being "driven out" of human–computer interaction. Their touchstone is an analog to Gresham's law: "Hard science drives out the soft." They argue that only by hardening up the psychology of human–computer interaction can psychology avoid being driven out by "harder" disciplines in computer science. Newell and Card define their program in terms of three key features: task analysis, calculation, and approximation (p. 215). Their objective is to provide fine-grained analyses of user tasks in terms of numerical parameters so as to afford the calculation of specific predictions. Because of the complexity of real human thought and action, these calculations must be approximate.

In this reply, we press the invocation of the original Gresham's law: "Bad money drives out the good." Or more strictly: Overvalued money drives out undervalued money. The appropriate response, of course, is not to learn to live with overvalued money but to see to it that undervalued money is valued at its true worth—or to change the system that produces these distortions of value in the first place. Newell and Card may be correct in worrying that hard science could drive out soft science, meaning that approximative calculation could undermine conceptual and explanatory science. The appropriate course of action, we argue, is to redress distortions in value, not to accommodate them; to protect and develop conceptually deep science, not to abandon it.

For the sake of this discussion, we need to address two potential confusions. The first pertains to the use of the term *hard science*. This is a loaded term. In adopting it, Newell and Card place alternative programs in the position of advocating "soft science." Not only do we find Newell and Card's conception of hard science problematic, but their use of the term clouds questions about theory and method with irrelevant value connotations. The second confusion derives from the contrast between Newell and Card's program and the earlier research monograph of Card, Moran, and Newell (1983). The Newell and Card article is a broadly couched program for psychology in human-computer interaction, but the Card et al. monograph is a tightly focused research project. Whereas Newell and Card focused on implications for interface design, Card et al. focused on research and approached design cautiously and peripherally. Finally, the Newell and Card essay is situated in the mid-1980s, whereas the Card et al. work was undertaken in the mid-1970s.

We are uncomfortable with the "hard science" terminology, but we feel that we have inherited it from Newell and Card (1985) in this debate. We are also uncomfortable making too sharp a distinction between Newell and Card (1985) and Card et al. (1983). Both approaches use a simplified information-processing box diagram of human cognition, called the Model Human Processor, and a family of models for approximative calculation of human performance times, called GOMS (an acronym for the key components of the models: goals, operators, methods, and selection rules). Yet it seems to us very different to speak of GOMS as a specific technical innovation in the human factors of computing systems (circa 1976) and GOMS as a prescriptive program for what can count as scientific psychology in human-computer interaction (circa 1985). The former we recognize as a foundational contribution to human-computer interaction; the latter we see as limiting and mistaken. Moreover, specific features of Card et al.'s research approach are eliminated from Newell and Card's program; chief among these is the detailed, qualitative analysis of behavioral and thinking-aloud protocols. We have respected this contrast by limiting our own use of Card et al.'s terminology (e.g., GOMS) when discussing Newell and Card.

Our argument takes the following course. First, we examine the motivation for Newell and Card's hard science program. We try to show how they misrepresent the current role of human factors in system design, the nature of theory in computer science, and the design process itself. Against a background of progress in human-computer interaction, they dismiss current human factors practice as a failure, without identifying its genuine strengths and weaknesses. They portray computer science theory as hard, when virtually no aspect of computer science theory involves calculation or approximation in their sense. They present no design-scale success stories of their approach, but merely invoke the term *design* and a limited view of what design is like and how it really works. The conclusion we draw is that the need Newell and Card imagine for their hard science simply does not exist.

Second, we examine their responses to existing criticisms of their approach.

We show how these responses fail to answer the criticisms and how they are mutually contradictory. Their reply to the charge that their hard science is too low level is to redefine "psychology" so that it will perfectly coextend with their enterprise, leaving critics to attack psychology itself and not them. Their reply to the charge that their hard science is too limited in scope is to assimilate a variety of current work (much of it not so low level) to their enterprise merely by saying "it fills out our 'vision' " (these two replies contradict each other). Their reply to the charge that hard science is chronically too late to help in the development of new technology is to argue that interface technology is elaborated far more slowly than is commonly thought. Finally, their reply to the criticism that their hard science is too difficult to apply is merely to suggest that we look for better opportunities to apply it. The conclusion we draw is that these criticisms identify true limitations on Newell and Card's approach.

Finally, we examine Newell and Card's research program in terms of its underlying methodological and theoretical commitments. The lack of adequate motivation for their proposals and the serious limitations on their scope and applicability suggest an underlying scientific agenda that goes beyond their official program of task analysis, calculation, and approximation: a view of task analysis as pure description; a conception of tasks as analyzable into highly isolable subtasks; and a conception of hard science as working with interval measurable quantities only. These underlying views are problematic in their own right (even as an understanding of hard science), and would require us to discard most of the information that psychology can render. The stereotype of hard science that underlies their program (a form of positivism) imposes limits on the scope of psychology needlessly and without basis. If adopted, this program would produce a sterile discipline of human–computer interaction that would satisfy neither psychologists nor system designers.

2. NEWELL AND CARD MISREPRESENT THE STATUS QUO

2.1. Misrepresenting Human Factors

Newell and Card claim that the available roles for human factors professionals are few and unsatisfying: In human factors, one can offer human factors guidelines or maxims, or one can build calculational performance models. Newell and Card dismiss human factors guidelines as vacuous and imply that computer scientists will ignore them. They regard their style of hard science as highly informative and imply that computer scientists will be eager to use it. These are debatable claims. To begin with, there are other roles that human factors professionals can fill. In particular, they can be members of a design team (practicing iterative behavioral testing in the context of rapid prototyping). Newell and Card do not acknowledge published case studies of how the iterative measurement approach, in the context of prototyping, has guided the design of usable interfaces (e.g., Boies, Gould, Levy, Richards & Schoonard, 1985; Good, Whiteside, Wixon, & Jones, 1984; Kelley, 1984). A view of human factors work that overlooks such roles, and the degree of cooperation between psychology and computer science inherent in them, is seriously incomplete.

Newell and Card portray usability guidelines as vacuous, commonsense maxims like "Communicate with metaphors" (p. 213). This maxim, like many others, however, has falsifiable empirical content. The designers of the Apple Lisa followed this maxim. Specifically, they used the metaphor of a desktop. It turns out, however, that professionals learning to use the Lisa are sometimes confused about the referent of "top of the desktop" (the screen, after all, is flat). Users are also confused by the fine distinction the metaphor makes between "stationery pads" and "stationery paper," and about the function of "folders" (Carroll & Mazur, 1986). The guideline that Newell and Card cite has empirical content, and our examples show that the guideline needs to be revised and made more specific. We are accordingly skeptical of Newell and Card's unqualified claim that guidelines derive "from a little common sense, plus placing value on serving the user well" (p. 214). It would be useful to see this claim argued in more detail against a comprehensive corpus of guidelines like Smith and Mosier (1984).

As Newell and Card have underestimated the potential importance of iterative measurement in the context of prototyping and of guidelines, they may have overestimated the potential role of calculational performance models. The considerable human factors tradition of predictive models of human performance on control tasks (e.g., the optimal control model; Pew & Baron, 1978) has been most successful in analyzing tasks that afford little user discretion (e.g., continuous response as opposed to decision), and experience with these models has shown that aggregated predictions at higher levels of behavior organization in fact depend little on low-level modeling structures (Pew & Baron, 1983). Newell and Card are concerned with modeling task environments that incorporate decision making and explicitly assume that low-level modeling can fix aggregated predictions at higher levels (p. 227). In summary, neither the history of human factors nor the nature of current progress in it supports the monolithic program of Newell and Card.

2.2. Misrepresenting Computer Science

Newell and Card portray computer science as hard and therefore unreceptive to a psychology that is not also hard. In some sense, this is true: Systems work in computer science is a design domain; its results are codified in designs and implementations, and to that extent it produces hard results. Psychology is part natural science and part social science, but it is not a design science. Psychologists study minds; they don't build minds (of course, artificial intelligence tries to be a design science in just this sense). However, Newell and Card do not primarily intend the contrast between hard and soft as a stand-in for the design science/natural science contrast (as they make clear in their remarks about rapid prototyping, p. 213). There are two other senses of "hard" to consider: (a) science that seeks to provide formal descriptions whose properties can thereby be analyzed further and (b) science that seeks to provide tools for approximative calculation. Plainly, these need not be the same.

Newell and Card list examples in making the claim that computer science is hard: "compiler construction, parsing, program specification, correctness proofs, denotational semantics. . . ." (p. 213). This work is hard science in the first sense, but not in the second. Indeed, mathematical modeling in computer science rarely traffics in numbers or in interval measurements; it works with other abstract structures in mathematics. Parsing theory, theories of compiler construction, and semantics for programming languages are hard in the ordinary sense, but not in Newell and Card's restricted sense. Algebraic semantics for programming languages does not generate quantitative predictions; the activity that goes on in this area consists of proving theorems and exploring properties of formal systems (Guessarian, 1981). Newell and Card do not produce a single example of computer science research that is hard in their restricted sense (b) in the preceding paragraph.

Other areas of computer science that Newell and Card cite, such as structured programming and artificial intelligence (pp. 213–214), are not hard in any sense. They are conceptual, in fact psychologically conceptual. Structured programming, as well as recent developments in object-oriented languages (e.g., Smalltalk), knowledge engineering languages (e.g., Prolog), and interface toolkits, have emerged in response to usability issues in programming itself. Psychological concerns (at least implicit ones) underlie the current concern in computer science with making programs easier to inspect, more modular, and more reusable. All of these demands stem from what programmers need in order to do their work better. They are not pure computer science, because they are not strictly necessary for programs to get the job done: Automata theory guarantees that if a program can be written at all, a flat, unstructured program (without subroutines!) will always get the job done.

The interdisciplinary thrust of some of these new areas of computer science is problematic for Newell and Card's "hard science" thesis. For example, they allude to work on intelligent interfaces in a strangely exaggerated manner: "if the interface is intelligent, then it is not necessary to know anything about the user, because the interface will be able to interact with the user intelligently" (p. 214). Clearly, the limited, and often toy-scale, examples of intelligence now available in demonstration systems have not eliminated the need for us to know how users think. And much current thinking in AI holds that the interface will not be able to "interact with the user intelligently" unless it also has means of finding out what the user really knows, and this endeavor will rely on psychological theories of knowledge (Carroll & McKendree, 1987; Goldstein, 1982).

In another example, they characterize rapid prototyping as a "panacea" that "bypasses the need to know anything about the human" (p. 213). However, one of the chief impacts of rapid prototyping is that it affords iterative measurement of usability within the software development process (e.g., Richards, Boies, & Gould, 1986). It thus belongs to the same family of usability-motivated developments in computer science as structured programming and object-oriented languages.

We suspect that it is a fundamental mistake to analyze the differences between psychology and computer science in terms of their methodological credentials, in terms of "hard" and "soft." Psychology and computer science are both broad and fragmented disciplines, and simple bases for global comparisons between them should be treated with skepticism.

2.3. Misrepresenting System Design

Newell and Card argue that the psychology of human–computer interaction must become harder in order to establish and maintain a role for itself in interface design. However, they provide almost no analysis of either the specific nature of current barriers to psychology in interface design work or of specific benefits that would accrue from adopting their program. To establish that problems exist they cite one newsletter (*Human Factors Society Bulletin*) article complaining that things should be better. They fail to discuss any of several analytical case studies of new roles and new successes for psychological applications in interface design (e.g., Boies et al., 1985; Good et al., 1984). They also fail to show how their hard science approach could guide the development of advanced workstations and new applications.

In fact, Newell and Card's optimism about the utility of hard science for system design seems based on an oversimplified model of the system design process. They model design as a process with fixed, antecedently clear goals, a process consisting of top-down refinement only. "The designer searches [the design] space by successive refinements of the initial partial representation" (p. 223). Their diagram (p. 224) does not provide for iteration, although that is a defining characteristic of design; it completely ignores the redefinition or discovery of goals that often occurs in the design process (Carroll & Rosson, 1985; Dreyfus, 1953). Even Jeffries, Turner, Polson, and Atwood (1981), working within Simon's (1973) top-down approach, found extensive trial-and-error activity in the design process. Newell and Card have abandoned the caution displayed by Card et al. (1983, p. 406), who stated that "design is an open process, in the sense that the design problem is constantly being redefined."

A specific assumption that Newell and Card make is that design trade-offs can be computed. They assert that only when design trade-offs can be computed will usability trade-offs be considered in system design, giving as an example the trade-off "between the effort to learn a complex interface and the power of having it" (p. 222). Unfortunately, they give no indication of how such a trade-off would be calculated, and indeed it would appear to lie beyond the purview of performance time models (requiring models of the process of problem solving, and of learning and motivation, as well as longitudinal studies of learning and performance patterns in a naturalistic setting). Even if Newell and Card could provide a real worked-out example, they would still need to show why it is important to quantify given trade-offs; these are frequently not quantified in branches of engineering like civil engineering that draw on hard science bases far more well developed than those of either psychology or computer science.

A key limitation on an approach that focuses on quantitative trade-offs are the many discontinuities in usability effects. Small differences in interfaces often cause large differences in usability. For example, altering semantic and structural relations between command names may have no impact whatsoever on keystrokes or other quantitative measurements of errorless expert performance but may cause specific and recurrent errors (in one case of this type, Carroll, 1982, reported 7-fold differences in error rate, persisting through repeated learning trials; see Buxton, 1986, for discussion of related phenomena pertaining to input devices). The many case studies Newell and Card do not cite or discuss also show how details matter (Boies et al., 1985; Good et al., 1984).

In criticizing Newell and Card's representation of the status quo in human factors, computer science, and system design, we do not wish to imply that there are no problems. We agree with Newell and Card that system designers do not always pay enough attention to usability issues. However, as we have noted above, Newell and Card have *not demonstrated that the reason for this* lies in presenting concepts and qualitative relations rather than numbers or that the remedy lies in approximate calculations of expert performance times. In fact, Newell and Card themselves raise and discuss four important criticisms of their program.

3. NEWELL AND CARD FAIL TO ANSWER THEIR CRITICS

Newell and Card devote more than half of their article to defending their program against four fundamental criticisms; namely, that it is too low level, too limited in scope, too late, and too difficult to apply to provide any real science base or leverage in system design. We claim they fail to answer any of the four criticisms in a satisfactory manner. Indeed, the very manner in which they try to address these criticisms shows how well founded they are.

3.1. Being Too Low Level

Newell and Card respond to the criticism that their approach is too low level (concerned only with keystrokes and unit tasks) by claiming that psychology itself is very low level. Indeed, they seek to redefine psychology so that it becomes coextensive with the range of phenomena that their approach is able to handle. The basis for this redefinition can be found at the beginning of the article: "the hard science will tend to be used and the soft science ignored, regardless of whether all of the important issues are within the scope of the hard science" (p. 211). The problem with their response is not only that many important issues in psychology lie outside the scope of their hard science, but that the way that they have chosen to define the scope of hard science is not coherent.

Newell and Card (p. 225) propose a classification of mental processes in terms of time bands or time scales of human action. From short duration to long, these are the domain of natural law, the domain of psychology, the domain of bounded rationality, and the social/organizational domain. (Historically, this scheme seems derived from Newell & Simon's, 1972, notion of "cycle time," a notion that succeeding work in cognitive psychology has made little use of.) Newell and Card define *psychology* in terms of the temporal durations of the events that it studies. We will call their low-level psychology *psychology** so as not to presume any connection with the scientific domain of psychology as normally understood. The scope of psychology* is processes ranging from .1 sec to 10 sec (see p. 225). This happens to be the same range of low-level processes as those Newell and Card seek to account for: The basic operators of GOMS vary from .5 sec to 2 sec in duration (p. 217).

Newell and Card's redefinition of psychology raises many new questions. To begin with, duration seems to be an incoherent basis for individuating mental processes. It does not enable us to divide mental processes into natural kinds. Learning, for instance, may happen quickly or slowly. If we are to understand learning, we had better examine longer- as well as shorter-term examples of it. In Newell and Card's time band of bounded rationality, "human activity is described by the goals or ends being attempted" (p. 225). This conflates goals, which on many analyses are an inherent part of all cognitive processes, with conscious deliberation, which their examples of bounded rationality all seem to involve. Indeed, most of the processes studied by cognitive psychologists lie outside of psychology* in the bounded rationality and social/organizational time bands. The larger tasks that people do with computers (tasks that they define for themselves, or plan in terms of, like writing letters) cover much longer spans of time than psychology*. Learning and exploring a computing system or an application, and developing serious expertise, can take weeks or months. Perception is also important to human-computer interaction, but much of it lies outside psychology* too, in the natural law time band.

After redefining the scope of psychoogy to suit the narrow horizons of their style of modeling, Newell and Card have little to say about the consequences of rethinking the scope of the science. They cast aspersions on the quality of all undertakings outside of psychology*: "Scientific psychology seems to traffic heavily at the low end of things" (p. 225), whereas everything else is "pop psychology" and unfounded speculation. In a single dictum, they implicitly dismiss organizational, social, and developmental psychology along with a substantial part of cognitive psychology, ignoring the genuine research successes in all of those areas. The only argument they raise for confining attention to psychology* is a reductionist one:

> New psychological laws of information processing do not arise at longer durations. There are, instead, the accumulated effects of long-term memory and skill acquisition. . . . New psychological phenomena occur as time increases, but their theoretical explanation is to be found in the interplay of the limited processing mechanisms of the psychological band [that is, psychology*] and the user's intendedly rational endeavors. (p. 227)

As a programmatic claim, this is highly dubious (see Section 2.1. here and Pew & Baron, 1983). Even if true, it is virtually irrelevant to applied work. If the derived phenomena are important for understanding human-computer interaction, then modeling efforts will have to focus on them and not on the supposedly basic phenomena studied by psychology*.

Newell and Card's redefinition of psychology is irrelevant to the question of whether the range of phenomena they address are too low level to help analyze the interesting problems of human-computer interaction that motivated the enterprise in the first place. Changing the name of the domain of endeavor cannot save it from being too low level. What we need instead is a demonstration that the reductionist analysis can be carried out.

3.2. Being Too Limited

Newell and Card's approach has often been criticized for having too limited a scope (Shneiderman, 1984). Its prime application has been to predict the performance times of highly skilled users on line editors (Card et al., 1983). Providing a conceptual account of the structure and genesis of user error (i.e., beyond error times and frequencies) is beyond its capabilities (e.g., Robertson, 1983). Providing an analysis of people's problem-solving activities is beyond its capabilities (Card et al., 1983; Polson, 1987). Newell and Card's response to this criticism is to claim that a vast amount of current research in cognitive science—none of it explicitly related to their program—is, in fact, filling out their "vision." For example, Newell and Card assimilate Anderson's (1983) ACT* model as filling out their vision in the areas of problem solving and memory. Newell and Card's vision is so expandable that any area in cognitive psychology can be annexed by fiat (p. 230). Whereas elsewhere they implicitly dismiss most psychology as soft, and insist that models must be constructed to meet their criteria of hardness, to bolster the vision they appropriate theories that have never yielded calculational models, that have no apparent prospect of doing so, and that lie outside of psychology*.

What exactly is Newell and Card's vision and how is it related to their main program? They do not provide specific criteria and arguments for the assimilation of a variety of work into their vision. This makes it impossible to assess the claim that such work plays a role in the vision, or indeed to discriminate the vision from an arbitrary and vacuous device masking an equivocation about what the program advocated actually encompasses. Further, it must be noted that Newell and Card's replies to the first two criticisms contradict each other. They defend against the charge that their approach is too low level by saying that this focus is appropriate because it is coextensive with that of psychology*. They defend against the charge that their approach is too limited in scope by claiming that research in areas like learning and mental models that lie outside of psychology* are filling out their vision. On grounds of consistency alone, at least one of the two criticisms is true.

Perhaps sensing these problems with their notion of vision, Newell and Card present a second response to the criticism of limited scope in asserting that many psychological issues just don't come up in the area of interface design.

> Our topic is the psychology of the human–computer interface, not all of psychology. This distinction is important, because the proposal is not to make all of psychology better. The human–computer interface is, in fact, a psychologically limited microworld. Many issues of the wider world of psychology do not arise. (p. 222)

This is an untenable claim, made without argument. The user interface, after all, can be used to present "microworlds" in many different task domains. Although there are many human activities not encompassed by the user interface, the range of problems that get solved, or could be solved, using the interface, is far broader than the range of problems actually studied in the laboratory by academic cognitive psychology (Carroll, 1987; Norman, 1987). It is hard to name an area of psychology that is not involved in understanding the cognitive interface: perceptual, personality, educational, developmental, social, organizational, among others.

3.3. Being Too Late

Newell and Card respond to the criticism that their approach is too late in application to make a difference in system design by arguing that computer technology develops far more slowly than commonly thought. They represent interface technology development as a process in which a canonical interface design is introduced every 5 years, then quickly stabilizes for a 15-to-20-year life cycle. This pacing, they suggest, can provide the time for a psychology of human–

computer interaction to have impact (p. 235). This seems gravely over-optimistic. Indeed, the view of interface development as the stately progress of canonical technologies is just irrelevant to the problem of being too late. Whether a canonical interface last 15 years or 1 year, the place to have real influence on its design is at the birth of the new technology. After that, constraints are imposed by the initial choices that were made, wise or unwise, as Newell and Card themselves acknowledge: "throughout the life cycle . . . the gestalt of the interface does not change" (p. 234). If Newell and Card need even 1 year after the technology is introduced to calculate the right design, they will be one year too late.

A secondary problem with the appeal to canonical interface technology is exposed by the depiction of the development process in a bar chart with unlabeled axes (Figure 7, p. 235). What purports to be a taxonomy of interfaces merely classifies visual display devices, listing as "correlated features" such aspects as amount of intelligence and rates of interaction. These features may indeed be correlated, because the chart provides a scale of historical time. Such correlations are only interesting if they point to causal relations or relations of constraint between different aspects of the interface. But it is clear that they do not (progress in knowledge engineering has been independent of display devices, and interaction rates, at least early on, *slowed* as character-based CRTs were replaced by bitmapped displays). Mere correlations between features give us no insight into the evolutionary development of interfaces.

3.4. Being Too Difficult to Apply

Newell and Card respond to the charge that their approach is difficult to apply by suggesting that "ripe application domains" (p. 236) be sought. They recommend that their approach be extended to intelligent tutoring systems. Clearly, this is an important area and one in which models of human knowledge and skill play a central role. (This suggestion also enables them to assimilate the work of Anderson once again—in this case, the research on LISP tutoring—Anderson & Reiser, 1985.) However, it is clear that Newell and Card are again being inconsistent here. Intelligent tutoring, as they acknowledge (p. 225), is not in the purview of psychology*: Tutoring is a "high-level task" in the bounded rationality time band. Moreover, intelligent tutoring is an extremely soft area in which foundational questions are still being addressed: Theories of knowledge domains and knowledge construction have not been developed or applied in this area (Carroll & McKendree, 1987). None of the work in this area seems suited to approximative calculational modeling, and Newell and Card do not indicate how such modeling could be carried out.

Beyond the problem of identifying domains to which Newell and Card's approach might be extended, we need to ask about the applications that have already been made (the approach, after all, has been around for nearly 10 years).

Newell and Card are disappointingly cursory; they present a list of "studies that have applied, tested, improved, or extended the approximative models" (p. 236). Demonstrating how these examples worked would have been much more valuable than just listing them. In particular, it would have been valuable to show which of these studies, if any, applied calculational models in design rather than strict research settings. Newell and Card themselves insist that "design is where the action is, not evaluation" (p. 214) and that "the payoff for design must be demonstrated" (p. 223). Yet they provide no evidence that their calculational models have been used in any real designs, much less that they were useful.

Given the dearth of successful applications of Newell and Card's approach, we might well ask what has impeded its application. An approach can be difficult to apply because it introduces unfamiliar concepts whose implications are hard to work out, or because it requires the construction of auxiliary theories before it can be tested empirically or used practically. On the other hand, an approach can be difficult to apply because its content is too thin, not too rich. It is difficult to generalize a conceptually impoverished program beyond the narrowly conceived set of problems that it was originally designed to address. We will argue that the limitations that Newell and Card have imposed on themselves in the name of hard science intrinsically prevent the approach from being extended beyond predicting ideal expert performance times on line editors.

4. NEWELL AND CARD'S PROGRAM IS INADEQUATE

To this point in our discussion, we have argued that the motivation for hard science, as cited by Newell and Card, is grounded in misunderstandings of the status quo (Section 2) and that their replies to specific criticisms of their approach are unsuccessful (Section 3). It is possible that these deficiencies apply only to current implementations of their research program, and not to the program itself. The way to resolve this question is to examine the underlying program, to isolate its presuppositions, its methodological and theoretical commitments, and the constraints they impose on the program (Campbell & Bickhard, 1986). The other reason for examining the program, of course, is to better understand just what Newell and Card mean by "hard science," in order to determine whether a hard cognitive psychology of the interface is either possible or desirable.

We have referred many times to Newell and Card's focus on calculation and approximation, and to their view of task analysis. Each of these requires further explication, however, if we are to understand how the program has been applied and how Newell and Card respond to the alternatives to it. For instance, is the task analysis that they advocate descriptive or explanatory? What kinds of values will be calculated? What will be retained in the approximations they urge and what will be left out? Is there any causal connection between tenets of the program for hard science and the fact that the particular research projects do not address the conceptual nature of user errors? These questions can be answered, but only if certain implicit features of the program are brought to light. Newell and Card's approach rests on three important commitments: (a) Their program of task analysis is descriptive instead of explanatory; (b) their task analyses are atomistic—tasks are built up out of independent low-level units; and (c) their approximative calculations produce predictions of interval measurements only.

4.1. Descriptivism

The distinction between descriptive and explanatory models is fundamental (Campbell & Bickhard, 1986). An explanatory psychological model attempts to account for the mental processes by which the user does a task. A descriptive model predicts performance on tasks, but without modeling the actual means by which the user does the tasks. Newell and Card clearly distinguish between tasks as understood by people performing them and tasks as objectively analyzed (p. 227). Their concern is with the latter, which they do not view as being a psychological endeavor: "the cognitive psychologist has no unique capabilities for . . . investigations of specific task domains" (p. 227; see also Card et al., 1983, p. 10). Thus, when Newell and Card envision the extension of their approach to intelligent tutoring systems (p. 237), they aim at "objective calculation and task analysis." By contrast, those who tried to build such systems have been interested in the nature of learning and its constraints on teaching systems (Brown & Burton, 1978), and in making detailed investigations of learner goals and learning paths (Goldstein, 1982). But which task analysis, the learner's or the observer's, actually affects the learner's behavior, helps to account for what is learned, makes the learner enthusiastic or inattentive? Newell and Card use task descriptions to predict selected measures of user performance; an explanatory approach would model the user's mental processes and representations in order to account for skills, learning, and motivation.

Newell and Card's commitment to descriptivism is not limited to task analysis (which belongs to the time band of bounded rationality) or to imagined extensions of the model to new domains, like intelligent tutoring. It is evident in their discussion of psychology* itself. They denigrate "the theory game" in psychology—the interest in generating and contrasting explanations of phenomena. They explicitly attack the view that theories are for explaining, not predicting (p. 219). Explanatory psychology "will never beat Gresham's law" (p. 219). The attack on the theory game by Card et al. (1983) can be understood as a reaction to academic cognitive psychology in the mid-1970s, which often consisted of trivial information-processing models (e.g., the models of sentence verification that prevailed at the time), and a proliferation of experimental paradigms and measurements specialized for distinguishing one small-scale model from another. Against this background the desire for an alternative that could address practical questions, even if it were only descriptive, is understandable. Serious explanatory psychology, however, has always been concerned about more fundamental issues and a broader range of phenomena than the miniature, paradigm-bound approaches that Card et al. seem to have reacted against. Insofar as foundational questions about perception, learning, and communication are raised by the study of human–computer interaction, we can ill afford to dispense with the resources and concerns of explanatory psychology.

Once the descriptivism that underlies Newell and Card's project has been recognized, other aspects of the project become more understandable. The lack of attention to understanding errors in their program can be derived from the fact that task descriptions tend to be a priori models of perfect performance. Descriptive task models usually derive from a formal theory of the task domain. Thus, if they model knowledge of that domain, they model perfect knowledge. Newell and Card allow that their analysis is easier to apply to experts than to novice users (p. 233). But this is not the real point. Their model is a model of ideal experts, not of real experts or real novices. Real experts make errors, use nonoptimal methods, and don't bother to learn or use some aspects of system function (Draper, 1985; Rosson, 1983).

Errors have no place in Newell and Card's predictive models of ideally skilled performance. Newell and Card's suggestion that errors be split off, and that their "GOMS-like" regularity be captured in theories of error (p. 230), underscores their descriptivism. In explanatory psychology, errors are of interest because they test boundary conditions of knowledge and skill and thereby provide valuable insights into the underlying structure of knowledge and skill. The analysis and interpretation of errors is a crucial part of approaches to cognition and development from Piaget to Anderson. Of course, using errors for this purpose presupposes that one is interested in understanding how human beings act and the purposes for which they act, and not merely in describing the distribution of behavior tokens in the stream of events. The limitations of descriptivism may explain why Newell and Card appeal to a vision as a device for saving their approach from being too limited.

4.2. Atomism

Accompanying the predilection for descriptive task analysis, though not entailed by it, is the commitment to analyzing performance into isolable units or atoms. This has led to extreme disregard for context effects or interactions. For example, the simplified instantiation of the GOMS model known as the Keystroke-Level model (Card et al., 1983) assumes only one operator at a time and the independent execution of contiguous plans and actions. These assumptions are incorrect empirically, as shown by coarticulation phenomena in speech (Denes & Pinson, 1973) and nonindependent components of human movement (Smith & Smith, 1962). Proposals now being made to model the full range of cognition in terms of massively parallel processing (Rumelhart & McClelland, 1986) suggest that atomistic analyses are unacceptable in general.

Although the Keystroke-Level model is readily understandable as an instance of atomism, Newell and Card seek to justify it in terms of approximation. Just what sense of approxmation do Newell and Card have in mind? They claim that approximation is characteristic of engineering; this suggests selective and simplified applications of a hard basic science, as in the applications of Newtonian mechanics made in civil engineering. Such applications are virtually impossible in human-computer interaction. How could a "hard" applied psychology be built out of a basic psychology that is mostly "soft"? Newell and Card specifically claim that they want to reform the psychology of human-computer interaction without reforming all of psychology (p. 222). This suggests that they will not be applying basic psychology, approximately or otherwise.

More generally in science, approximation is carried out against a background of surveying the problems of interest and the complexity of their solution. A model is not to be called an approximation simply because it is incomplete and imprecise; all models are incomplete and imprecise. Approximation is also not the same as arbitrary simplification, or idealizations adopted for expediency. Some thought has to go into the real nature of the problem, what is left out by the approximation, whether anything essential to the problem has been left out, and how much of a practical difference it makes. Newell and Card neither carry out nor draw on any such survey of the important problems in human-computer interaction and the means available to solve them. They never consider what the relevant trade-offs are; they do not show, for instance, that in return for specific simplifying assumptions, it becomes easier to make a useful prediction. Without these details, we cannot distinguish Newell and Card's advocacy of approximation from a commitment to atomism.

The commitment to atomism entails a reductionist ontology, which Newell and Card clearly accept in their argument that higher level psychological time bands of human behavior and experience can be reduced to atomic mechanisms that live within the psychology* time band. Their consideration of goals is limited to very low-level goals, which function in their analysis as unanalyzed atoms. Newell and Card's cursory discussion of error amounts to the assertion that error be split off into a separate atomic theory. The kind of theory that they want for dealing with errors is a general theory of error (or at least, of error-recovery times), not tied to skill or understanding in any specific domain. They want "a way of accumulating data on errors that has cross-situational validity" (p. 231). We interpret this as a call for a general theory of errors, not dependent on the knowledge or learning that is being assessed. Indeed, Newell and Card reinterpret (p. 230) Norman's (1981) account of action slips, which was in-tended as part of a model of skilled action, as a general-purpose theory of error. Do errors ever have context-free information value?

4.3. Interval Measurements

Newell and Card advocate calculation as a touchstone of hard science. Their discussion of partly linear models (pp. 232–233) illustrates what they mean by calculation. The linear part of a model is "linear in time, errors, solution opportunities, or whatever" (p. 233): It requires something that is continuously measurable or countable. The nonlinear part or "difficulty component" is to be estimated and plugged into the equation so predicted times (or similar measures) can be obtained—its qualitative features or underlying explanation are not of interest. Performance can thus be resolved into number of operations performed, number of chunks learned, and so on, with a nonlinear, grab bag difficulty component tacked on, in order to facilitate the generation of quantitative predictions. Not only do partly linear models waste qualitative information, but such models are inherently unsuited to situations in which small changes in the interface, which have a major influence on understanding or motivation, lead to big changes in performance. Partly linear models presuppose that small changes in the interface will produce small increments or decrements in performance, but this is wrong (Carroll, 1982; see also such design case studies as Boies et al., 1985; Good et al., 1984).

All of the examples Newell and Card cite are consistent with this bias for interval measurement. They dismiss "qualitative factors" as soft (p. 211). The GOMS technical work that their program rests on predicts performance times. The best-known extensions of the GOMS model (Polson, 1987; Roberts & Moran, 1983) predict learning times. Newell and Card seek "a simplified model of the human in terms of memories, processors, and *a few quantitative parameters of each*" (p. 215, italics added).

This orientation leads them to place higher priority on identifying "a few quantitative parameters" than on developing analyses that fit the target phenomena. It deprives the hard science approach of most of what psychology has to offer. Outside of psychophysics and motor control, there are few interval measurements available in psychology; performance times and percent correct are the most widespread. Because nominal categories and ordinal measurements are not hard, they can play no role in this program (at most, they can be included in the vision, or packed into a nonlinear model component)—regardless of the information they provide. Errors, interview and protocol statements, and explanations do not provide the right kinds of data and so are to be banished from the study of human–computer interaction. In psychology, we labor under the restriction of not being able to observe mental processes, and having only a few kinds of behavioral data to provide empirical constraints on our accounts of process. Under these circumstances, we had better use all of the information that we can get. Newell and Card's approach is not low level and limited in scope by accident; rather, it has these limitations as a consequence of its commitment to interval measurement.

Perhaps in recognition of these problems, Newell and Card state that "The prominence of performance-time measures . . . occurred because of the locus of the initial research successes" (p. 228), implying an interest in a wider range of phenomena. This may have been true of the original GOMS research program (Card et al., 1983). However, Newell and Card have sought to turn the limitations of that research into prescriptions for hard science. They have moved from selecting performance times for one program of research, to promulgating norms of calculation and approximation that compel the use of measures like performance times in all research. Performance times (along with learning times and error-recovery times) are among the few things in cognitive psychology that can be measured on an interval scale. Errors are essentially qualitative. Indeed, the nonrole of errors in Newell and Card's hard science is overdetermined by its three underlying commitments: Errors as a means of exposing boundary conditions belong to explanatory and not descriptive psychology; errors, because they derive their meaning from the contexts in which they occur, cannot be usefully examined atomistically; errors are qualitative and so cannot be measured as interval quantities.

5. PROSPECTS FOR THE NEW POSITIVISM

Although Newell and Card's discussion contributes to theoretical debate in human–computer interaction by setting a reference point, their specific argument fails. As we have seen, the program is unmotivated (Section 2), suffers from serious limitations (Section 3), and in fact suffers from these limitations in principle (Section 4). Newell and Card's program does not proceed from a consideration of the problems to be solved in the domain of human–computer interaction and the methods that might be useful to address those problems. It races ahead of such fundamental questions as "What is our science about?" and "What should our science do?" What Newell and Card have focused on is the *form* of their theory rather than the *content*. They have in effect undertaken a scare campaign, arguing that unless our science takes a particular form, computer scientists will disdain and ignore us. Although Newell and Card claim to be concerned about what will produce good designs (pp. 210–211), their arguments appeal much more strongly to the demand for scientific credentials than to any benefits for interface design. This emphasis on scientific credentials recalls positions that have prevailed in the past; for instance, the doctrine that talk of mental states is unscientific.

In the history of science, soft conceptual science lays the necessary groundwork for hard quantitative science. The appropriate ontology, the right problems and the right ways of looking at them, and the mathematical techniques appropriate to the subject matter have to be in place for hard science to develop.

n the past, philosophy of science tended to ignore the actual reasoning of scientists, in favor of legislating methodological criteria a priori (Laudan, 1977; Shapere, 1977). A particularly disruptive instance of this approach was logical positivism, a philosophy of science founded on doubts about the legitimacy of inferred entities in scientific explanation (Carnap, 1928/1974; Hempel, 1965). Theoretical entities were to be tolerated only as part of a deductive system for yielding empirical predictions (metaphysical claims were forbidden) and strict rules of correspondence between theoretical and observational statements were to be followed. Logical positivism did considerable and lasting damage to the social sciences which, unlike the natural sciences, have been chronically vulnerable to worries about appearing "scientific" enough. Social scientists have often rushed to adopt the trappings of hard science without the substance; for instance, by using mathematical modeling techniques inappropriate to their subject matter (ranging from structural models of cognitive operations, Campbell & Bickhard, 1986, to models of static equilibrium in economics, von Mises, 1966, to frankly bizarre applications of mathematics in sociology, Andreski, 1973).

The worst distortion of all was the ascendancy of behaviorism, which was accepted because it imposed extrinsic requirements on what could count as scientific, not because it could answer many of the questions understood as psychological (e.g., Chomsky, 1959). The effect was to detour psychology from asking or dealing with most of the interesting questions in its domain for 40 years. Eventually these questions were addressed anyway, but at the price of starting over almost from scratch (e.g., Miller, Galanter, & Pribram, 1960). Although the behaviorist prohibition against mental entities is no longer accepted, the view of science that underlay behaviorism—logical positivism—still exerts a strong influence on thought and method in psychology 20 years after philosophers of science rejected it (Bickhard, Cooper, & Mace, 1985). Clearly, Newell and Card are not behaviorists, but in other respects they are renewing the methodological strictures of logical positivism. Their view of theory as predicting instead of explaining, and of quantification as the hallmark of good science, are vintage positivism (Campbell & Bickhard, 1986; Suppe, 1977). The legacy of behaviorism and positivism should make psychologists deeply suspicious of any attempt to restrict the range of their science to suit an extrinsic criterion of correct method. Methodology should fit the subject matter, not the other way around.

History suggests how Newell and Card's approach might develop. Newell and Card have clearly articulated their position: They will ignore many of the most important empirical problems in the area, to obtain analyses with the correct form. They will not seriously address perception, problem solving, or learning, although they may blur this by conflating their approach with its limitless "vision" and by invoking a grab bag of nonlinear model components. They will probably not impress computer scientists as they hope to. Computer scientists already have a richer view of theory (as evidenced by their current practice) and in any case look to psychologists for answers to important problems, not limited, low-level, late and inapplicable theories that are methodologically pure. Positivist strictures are easier to impose on pure laboratory research than on applied design work, precisely because the insignia of methodological purity matter less in applied areas. In the meantime, the institutional barriers to human factors influence on the design process (the ones that motivated Newell and Card in the first place) will not be affected at all.

History suggests some possible reactions against the new positivism in human–computer interaction. Behaviorism strengthened the appeal of phenomenology and hermeneutics by showing that "science" was incapable of dealing with meaning in human thought and action. Should Newell and Card's approach become an establishment view in human–computer interaction, it will stimulate a backlash in the form of approaches that reject all information-processing analysis, not just Newell and Card's overconstrained analysis. The elimination of meaning, even in the context of simple problem solving, makes Newell and Card's approach a natural target for advocates of hermeneutics. Whiteside and Wixon (1987) urge the rejection of information-processing cognitive models; they recommend a hermeneutic approach that aims at interpretation of particular cases rather than generalization or explanatory theory (cf. Winograd & Flores, 1986).

We are concerned to distinguish the baby from the bathwater. We favor an explanatory psychology of human–computer interaction, one that starts from an understanding of the real problems that have to be solved in order to provide better computer tools for people to use. Indeed, such work is going on in many places right now, and could only seem extraordinary in the context of Newell and Card's revival of positivism. What this work lacks is a clearly articulated statement of its program, and we suggest this as a top-priority work item for the psychology of human–computer interaction.

Newell and Card have supposed that the problems facing psychology in human–computer interaction are due to a conflict between hard computer science and soft psychology. We see the problem more simply: Interfaces can be built without the contributions of psychologists, but they cannot be built without the contributions of programmers. We also see the solution more simply: Psychologists need to make the case for usability by providing concepts, methods, and demonstrations of impact. An alternative to Newell and Card's monolithic view of conflict between hard computer science and soft psychology is the view that an interdisciplinary field of human–computer interaction is taking form, and that psychology and computer science must work together to develop research areas like artificial intelligence and rapid prototyping. This interactive relationship will thrive if psychologists can play an effective role within the design loop for new interface technology. To do this they will have to address the questions that designers really need answered, at a useful grain of analysis, and at the pace of the design process.

Real interface design problems arise when real users use real applications to achieve their real goals. Psychologists working in human-computer interaction need to examine rich slices of behavior in realistic settings. Laboratory studies of undergraduates performing simple and repetitive tasks with toy-scale mock-ups will often be irrelevant. Users make many errors, so explanatory proto-theories of error must be developed that will guide the redesign of interfaces to minimize errors and support error recovery. Analyzing only errorless performances, or merely timing error recovery, may not be useful. Understanding such difficult areas as sustained motivation and long-term learning (including perceptual learning), in which the psychological theory base is deficient, is critical; human-computer interaction may have to drive explanatory theory development in these areas. Psychologists in human-computer interaction will have to pursue these issues if they are to be effective.

Human-computer computer interaction is a frontier science: It is at a frontier of method and theory in psychology and of technology and application in computer science. It is evolving faster perhaps than any area in psychology ever has. As such, it will probably be impervious to monolithic and methodologically narrow paradigms. Rather, it will favor interdisciplinary cooperation between psychology and computer science. It will require of its participants a thorough appreciation of what system design is really like. It will not sustain approaches that are too low level, too limited in scope, too late, and too difficult to apply in real design. Technological change will continually impose new problems on the area, and methods will have to developed to address and resolve these problems whether some limited view deems them scientifically legitimate or not. It is our hope that good science cannot be driven out by science that merely tries to look good.

Acknowledgments. This article is fully collaborative. We would like to thank Mark Bickhard, Stu Card, Don Foss, Wendy Kellogg, Clayton Lewis, Tom Moran, Dick Pew, Peter Polson, Michael Rosenbloom, Eric Wagner, and John Whiteside for comments, criticisms, and discussion.

REFERENCES

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge: Harvard University Press.

Anderson, J. R., & Reiser, B. J. (1985). The LISP tutor. *Byte, 10(4),* 159–175.

Andreski, S. (1973). *Social sciences as sorcery.* New York: St. Martin's.

Bickhard, M. H., Cooper, R. G., Jr., & Mace, P. G. (1985). Vestiges of logical positivism: Critiques of stage explanations. *Human Development, 28,* 240–258.

Boies, S. J., Gould, J. D., Levy, S., Richards, J. T., & Schoonard, J. (1985). *The 1984 Olympic Message System—A case study in system design* (Research Rep. No. RC 11138). Yorktown Heights, NY: IBM.

Brown, J. S., & Burton, R. R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science, 2,* 155-192.

Buxton, W. (1986). There's more to interaction than meets the eye: Some issues in manual input. In D. A. Norman & S. W. Draper (Eds.), *User centered system design* (pp. 319-337). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Campbell, R. L., & Bickhard, M. H. (1986). *Knowing levels and developmental stages.* Basel: S. Karger.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Carnap, R. (1974). *The logical construction of the world and pseudoproblems of philosophy.* Berkeley: University of California Press. (Original work published 1928)

Carroll, J. M. (1982). Learning, using and designing command paradigms. *Human Learning, 1,* 31-62.

Carroll, J. M. (Ed.). (1987). *Interfacing thought: Cognitive aspects of human-computer interaction.* Cambridge, MA: Bradford/MIT Press.

Carroll, J. M., & Mazur, S. A. (1986). LisaLearning. *IEEE Computer, 91*(11), 35-49.

Carroll, J. M., & McKendree, J. E. (1987). Interface design issues for advice-giving expert systems. *Communications of the ACM, 30*(1), 14-31.

Carroll, J. M., & Rosson, M. B. (1985). Usability specifications as a tool in iterative development. In H. R. Hartson (Ed.), *Advances in human-computer interaction* (Vol. 1, pp. 1-28). Norwood, NJ: Ablex.

Chomsky, N. A. (1959). Review of B. F. Skinner's *Verbal behavior. Language, 35,* 26-58.

Denes, P., & Pinson, E. N. (1973). *The speech chain.* Garden City, NY: Anchor Books.

Draper, S. W. (1985). The nature of expertise in UNIX. In B. Schackel (Ed.), *Human-computer interaction — Proceedings of the Interact '84 Conference* (pp. 465-472). New York: North Holland.

Dreyfus, H. (1953). *Designing for people.* New York: Simon & Schuster.

Goldstein, I. P. (1982). The genetic graph: A representation for the evolution of procedural knowledge. In D. Sleeman & J. S. Brown (Eds.), *Intelligent tutoring systems* (pp. 51-77). New York: Academic.

Good, M. D., Whiteside, J. A., Wixon, D. R., & Jones, S. A. (1984). Building a user-derived interface. *Communications of the ACM, 27,* 1032-1043.

Guessarian, I. (1981). *Algebraic semantics.* New York: Springer-Verlag.

Hempel, C. G. (1965). *Aspects of scientific explanation.* New York: Free Press.

Jeffries, R., Turner, A., Polson, P., & Atwood, M. (1981). The processes involved in designing software. In J. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 255-283). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Kelley, J. F. (1984). An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Office Information Systems, 2,* 26-41.

Laudan, L. (1977). *Progress and its problems.* Berkeley: University of California Press.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior.* New York: Holt, Rinehart & Winston.

Newell, A., & Card, S. (1985). The prospects for psychological science in human-computer interaction. *Human-Computer Interaction, 1,* 209-242.

Newell, A., & Simon, H. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Norman, D. A. (1981). Categorization of action slips. *Psychological Review, 88,* 1–15.

Norman, D. A. (1987). Cognitive science—Cognitive engineering. In J. M. Carroll (Ed.), *Interfacing thought: Cognitive aspects of human-computer interactions.* Cambridge, MA: Bradford/MIT Press.

Pew, R. W., & Baron, S. (1978). The components of an information processing theory of skilled performance based on an optimal control perspective. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 71–78). New York: Academic.

Pew, R. W., & Baron, S. (1983). Perspectives on human performance modeling. *Automatica, 19,* 663–676.

Polson, P. (1987). A quantitative theory of human-computer interaction. In J. M. Carroll (Ed.), *Interfacing thought: Cognitive aspects of human-computer interaction* (pp. 184–235). Cambridge, MA: Bradford/MIT Press.

Richards, J. T., Boies, S. J., & Gould, J. D. (1986). Rapid prototyping and system design: Examination of a toolkit for voice and telephony applications. *Proceedings of CHI '86 Conference on Human Factors of Computer Systems,* 216–220. New York: ACM.

Robertson, S. P. (1983). *Goal, plan, and outcome tracking in computer text-editing performance* (Tech. Rep. No. 25). New Haven: Yale University.

Roberts, T. L., & Moran, T. P. (1983). The evaluation of text editors: Methodology and empirical results. *Communications of the ACM, 26,* 265–283.

Rosson, M. B. (1983). Patterns of experience in text editing. *Proceedings of the CHI '83 Conference on Human Factors of Computer Systems,* 171–175. New York: ACM.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations.* Cambridge, MA: MIT Press.

Shapere, D. (1977). Scientific theories and their domains. In F. Suppe (Ed.), *The structure of scientific theories* (2nd ed., pp. 518–565). Urbana: University of Illinois Press.

Shneiderman, B. (1984, January). Review: *The Psychology of Human-Computer Interaction. Datamation, 30,* 236–237.

Simon, H. A. (1973). The structure of ill-structured problems. *Artificial Intelligence, 4,* 181–201.

Smith, K. U., & Smith, W. M. (1962). *Perception and motion: An analysis of space structured behavior.* Philadelphia: W. B. Saunders.

Smith, S. L., & Mosier, J. W. (1984). *Design guidelines for user-system interface software* (Tech. Rep. No. MTR-9420). Bedford, MA: MITRE Corporation.

Suppe, F. (1977). The search for philosophic understanding of scientific theories. In F. Suppe (Ed.), *The structure of scientific theories* (2nd ed., pp. 3–254). Urbana: University of Illinois Press.

von Mises, L. (1966). *Human action* (3rd ed.). Chicago: Regnery.

Whiteside, J. A., & Wixon, D. R. (1987). Improving human-computer interaction: A quest for cognitive science. In J. M. Carroll (Ed.), *Interfacing thought: Cognitive aspects of human-computer interaction* (pp. 353–365). Cambridge, MA: Bradford/MIT Press.

Winograd, T., & Flores, F. (1986). *Understanding computers and cognition: A new foundation for design.* Norwood, NJ: Ablex.

HCI Editorial Record. First manuscript received April 7, 1986. Revision received May 30, 1986. Accepted by Donald Norman. —*Editor*