Psychology and the User Interface:

Science is soft at the frontier

(Abstract of invited talk)

John M. Carroll
User Interface Institute
IBM T.J. Watson Research Center
Yorktown Heights, NY 10598, USA

One source of intellectual overhead that every science inflicts on itself periodically is the clarion call to "be hard", to establish methodological ground rules so severe that they will insure that good science can prevail. This romantic notion would only be that if it were not for the fact that these fits of methodological purification have typically led to conceptual and empirical poverty. The excesses of positivism and its crippling effects on the sciences from physics to psychology are still in recent memory.

Newell and Card, in an invited article in the journal *Human-Computer Interaction*, have undertaken a modern variant of this methodological cleansing. However, in most respects their motivation and arguments are precisely those of the positivists. They urge that the psychology of human-computer interaction needs to be hardened, meaning it must more uniformly subscribe to parameter fitting, calculation, and quantitative approximation. They are explicit in identifying as their motivation the fear that the "harder" disciplines of user interface design and artificial intelligence will not take usability psychologists seriously unless the psychologists have hard methods. They suggest a modified Gresham's Law: "hard science drives out the soft," as if this is both inevitable and a good thing.

My own view is that science is always soft at the frontier. The psychology of human-computer interaction is at a frontier of method and theory in psychology and a frontier of technology and application in computer science. To me, it is fantastic to insist that we start right out on a "hard" psychological theory to guide designs for integrated co-authoring applications on workstations that support multi-media input/output when we can barely couch such a theory for well-worked, toy domains like cryptarithmetic and chess.

Newell and Card are too concerned with the form of science and too little concerned with its content. They urge calculation and quantitative approximation but seem almost blase about what exactly is calculated or approximated. At best, Newell and Card's discussion is very premature; more likely, it threatens to set the psychology of human-computer interaction backward by confusing the project of developing a fundamental understanding of usability and user psychology with the engineering practices we might be able to develop if we had such a science base to begin with.

This talk has four parts. In the first, I consider Newell and Card's clarion call for hard science, reviewing a critique developed jointly with Robert Campbell of IBM Research. Campbell and I argue: (1) that Newell and Card misunderstand and underestimate how psychology currently contributes to interface design and thus set out to solve a nonexistent problem; (2) that they misunderstand and oversimplify the system design process, and that indeed only by doing so can they find a role in it for their clumsy hard science; (3) that their replies to existent criticisms of their hard science are uniformly without serious content.

Their reply to the charge that their hard science is too low level is essentially to redefine "psychology" so that it perfectly coextends with their enterprise, leaving critics to attack psychology and not them. Their reply to the charge that their hard science is too limited in scope is to try 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'." (Notably, these two replies, taken in conjunction, are self-contradictory). Finally, their reply to the charge that hard science takes too long to help at all in the development process is to say that the elaboration of interface technology in fact takes place more slowly than everyone thinks it does!

In the second part of the talk I examine some of the current research work in human-computer interaction that is paradigmatically hard. I argue that the psychology of human-computer interaction, like psychology generally, suffers from a methodological bias for posing elegant, either-or research questions that idealize away variables like task context, e.g., "is mouse driven pointing control better than a velocity control joystick?" Perhaps the question should be: "under what circumstance is a mouse the right design choice, and under what circumstance is a velocity control joystick the right choice?" Hard psychologists seem too willing to trade off ecological scale for laboratory tractability (e.g., a study of command languages that exam-

ines a command set of 3 commands when realistic scale would be 1-2 orders of magnitude larger). The hard science of Newell and Card rests fundamentally on baldly unreasonable idealizations (e.g., assuming errorless performance for purposes of theory when in fact obtained error rates exceed 30 percent).

We need to concentrate on the important facts of user behavior, not ignore them because they lie outside our methodological purview. We must of course strive to make our science harder (in the usual sense of "more systematic"). But we must also guard against too much weight being given to superficial rigor and too little to the practical value of our theories in guiding the design of new technology.

In the third part of the talk, I examine the area of artificial intelligence research specifically directed at the construction of advisory expert systems (intelligent help and training facilities). Newell and Card might find it startling that an domain in the mainstream of AI, which they describe as hard, in fact has no systematic theoretical foundations (no constitutive theories of types of general skill, no principled taxonomy of knowledge domains, no user models that do not obviously violate fundamental facts about human learning and the growth of knowledge).

Indeed, this supposedly hard research area has no comprehensive methodology: experimental systems are routinely designed with the paramount goal of providing advice to users without any systematic consideration of how people give and take advice, what their real problems, goals, or needs are, etc. The field has a large inventory of dialog techniques, for example, but no understanding of the circumstances under which particular techniques are useful or of how to integrate various techniques to capitalize fully on prior work. Finally, there is no effective engineering aspect to this work: no one knows how to develop advisory expert systems with limited resources or on short schedules. It is simply incredible that anyone who understood the state of art in this field could hold it up as a paradigm of hard science.

In the final part of the talk, I consider what human-computer interaction might need in the way of a soft science, a conceptually richer and methodologically less limited science. I urge that we recognize that in a rapidly evolving, technology-driven area hard science can never drive out the soft. Rather it consolidates those areas that have become well-worked. We must learn better how to use soft science to identify concepts and behavioral phenomena that are really worthy of quantification and other "hard" analysis. We must develop an arsenal of realistic empirical methods, methods that efficiently and reliably produce information at the right level to impact the application of new technology, not merely at a convenient level. For example, collecting and taxonomizing users' critical incidents or thinking aloud protocols may generate information more directly pertinent to an iterative design process than a record of individual keystroke times -- but the keystroke times are "hard," more convenient to collect, and more familiar and routine to analyze.

We must develop qualitative theories, and means for expressing such theories. For example, a list of user knowledge states, with associated transition rules, may be more relevant to guiding the design of new technology than an equation describing a fitted curve of millisecond differences between performance means. Finally, we must extend the scope of our theories. For example, if users routinely make many errors, then our theories should incorporate errorful as well as errorless behavior. Empirical taxonomies of error, and even rough theories of action slips, abductive reasoning, and learning via metaphor and analogy are soft science, but perhaps critical if we are to have a serious and effective science of human-computer interaction.

In summary, it is elementary in the history of science that one cannot legislate the quality of the conceptual and empirical *content* of science merely by legislating the methodological *form*. In fact, if history is any gauge, *a priori* limitations on acceptable methods usually have an undermining effect on conceptual and empirical quality. Newell and Card are mistaken in their attempt to confine the psychology of human-computer interaction. Their view of hard science is arbitrary and in particular has been a fairly well-documented failure in providing real leverage in interface design, conceptually and empirically. Their view of AI as hard is similarly inaccurate, as evidenced by the subfield of advisory expert systems. Finally, there are routine alternatives to their hysterical and dismal clarion call.

Gresham's Law states that "bad money drives out the good", but it does not suggest that we accept this as our inescapable fate. Rather, it suggests that we protect good money by responsible fiscal policies. I suggest that we protect soft science by responsible methodological policies. Whenever a scientific program is championed on purely methodological grounds, we should cringe.