
David D. Woods

Westinghouse Research & Development Center, 1310 Beulah Road, Pittsburgh, Pennsylvania 15235

Human-Machine Cognitive Systems

Advances in tool building have focused attention on the decision-making and problem-solving activities in work environments. First, through increased control automation, the human role has shifted from an emphasis on the perceptual-motor skills needed for manual control to the cognitive skills (e.g., monitoring, planning, fault management) needed for supervisory activities. Second, developments in computational technologies (i.e., heuristic programming techniques) have greatly increased the potential for automating decisions and have resulted in environments where humans interact with another, artificial, cognitive system. The result is the need for and a growing interest in cognitive technologies—that is, techniques and concepts (a) to identify the decision-making/problem-solving requirements in some domain; (b) to improve decision-making/problem-solving performance; and (c) to develop joint human-machine cognitive systems.

A cognitive system is goal directed; it uses knowledge about itself and its environment to monitor, plan, and modify its actions in the pursuit of goals; it is both data- and concept-driven. People are obviously cognitive systems. Advances in computational technology have greatly expanded the potential for the support of human cognitive activities and for the development of artificial cognitive systems—i.e., systems that perform tasks normally associated with human cognition. However, these developments also create new challenges:

- What is effective decision support?
- How can one allocate decision tasks between human and machine?
- What is useful advice?
- What is an effective combination of human and artificial cognitive systems?

In other words, there can be a third type of cognitive system: A single, integrated system composed of both human and artificial cognitive systems (Hollnagel & Woods, 1983). Implicitly or explicitly, applying computational technology is an exercise in the design of a joint cognitive system, and this article examines the implications of

Abstract

Developments in computational technology have focused on tool building—how to build better performing machines. But tool use involves more. The key to the effective application of computational technology is to conceive, model, design, and evaluate the joint human-machine cognitive system. Like Gestalt principles in perception, a decision system is not merely the sum of its parts, human and machine. The configuration or organization of the human and machine components is a critical determinant of the performance of the system as a whole. Effective decision support then requires that computational technology aid the user in the process of reaching a decision, and not simply make or recommend solutions. As a result, there is need for and a growing interest in cognitive technologies as a necessary complement to computational technologies for research on and the design of decision support systems.

This article explores the implications of one type of cognitive technology, techniques, and concepts to develop joint human-machine cognitive systems, for the application of computational technology by examining the joint cognitive system implicit in a hypothetical computer consultant that outputs some form of problem solution. This analysis reveals some of the problems that can occur in cognitive system design—e.g., machine control of the interaction, the danger of a responsibility-authority double-bind, and the potentially difficult and unsupported task of filtering poor machine solutions. The result is a challenge for applied cognitive psychology to provide models, data, and techniques to help designers build an effective combination between the human and machine elements of a joint cognitive system.
this perspective for research on and the design of decision support systems.

**A Hypothetical Computer Consultant**

To begin our exploration of the joint cognitive system perspective on decision support, consider an alternative paradigm for the use of decision tools. In this paradigm, the primary design focus is to use computational technology to produce a stand-alone machine expert that offers some form of problem solution (*e.g.*, a recommended solution or solution categories ranked by some form of likelihood). The "technical" performance of this system is judged by the criterion of whether the solutions offered are "usually" correct (*e.g.*, Yu et al., 1979). This paradigm emphasizes tool building over tool use and questions about how to interface human to machine are therefore secondary to the main design task of developing a machine that usually produces correct decisions.¹

A typical encounter with a hypothetical system developed in this fashion consists of the following elements: The user initiates a session; the machine controls data gathering; the machine offers a solution; the user may ask for an "explanation" if some capability exists (where explanation usually consists of providing the user with a justification for the machine's solution); and the user accepts (acts on) or overrides the machine's solution.

What is the joint cognitive system architecture implicit in this hypothetical machine expert? The primary focus is to apply computational technology to develop a machine expert. In practice, putting the machine to work requires communication with the environment-data must be gathered and decisions implemented. Rather than automate these activities, system designers typically leave them for the human user (Figure 1). Thus, interface design for the hypothetical consultant is not so much how to interface the machine to the user, but rather, how to use the human as an interface between the machine and its environment. This emphasis results in a user interface design process that focuses on features to aid the user's role as data gatherer (*e.g.*, Mulsant & Servan-Schreiber, 1983) and on features to help the user accept the machine's solution. As a result, control of the interaction in this type of joint system resides with the machine. However, human factors research in person-machine systems has established that a machine locus of control can have strong negative effects on user and total system performance (see, for example, Smith, 1981; Hoogovens Report, 1976; Turner, 1984).

A related characteristic of this paradigm is that user acceptance and the machine expert’s "technical" performance (again, in the sense of offering problem solutions that are usually correct) are seen as independent issues (*e.g.*, Shortliffe, 1982). "Excellent decision making performance does not guarantee user acceptance" (Langlotz & Shortliffe, 1983). Thus, lack of user acceptance (where acceptance means the user adopts the machine's solution) is seen as a problem in the user that must be treated by measures outside of the essential characteristics of the machine expert. One proposed technique is to embed other, useful support capabilities in the same computer system that implements the machine expert, *e.g.*, data management functions such as computerized data entry forms or standard report generation (Langlotz & Shortliffe, 1983). In pursuit of user acceptance of the machine's solution, some designers of machine experts will go so far as to suggest that systems "provide the physician (i.e., the user) with the ability to report the facts he considers important (even if they are not used internally) (i.e., by the machine expert)" (Mulsant & Servan-Schreiber, 1983; italics added). The joint cognitive system viewpoint suggests, on
the other hand, that problems with user acceptance are very often symptoms of an underlying deficiency (e.g., machine control) in the “cognitive coupling” (Fitter & Sime, 1980) between the human and machine subsystems.

The emphasis in the hypothetical system is user acceptance of the machine’s solution. However, since these systems are imperfect, output typically consists of some form of confidence or likelihood estimate over a set of possible diagnoses. The user is expected to act on the machine’s solution, but what is the machine’s solution? The highest likelihood category? Likelihood weighed by consequences? By some form of expectation? For real-time domains, are temporal fluctuations important? What if there are several high-likelihood options or no high-likelihood options? Choosing a solution to act on is further complicated because of the non-standard procedures that are typically used to compute likelihood estimates. Due to the method used to generate the confidence values, the likelihood data usually rests on an ordinal measurement scale. However, they are often represented as interval scales to the user. This mismatch creates the potential to mislead the human decision maker and to complicate his decision task. These examples illustrate that computing likelihood over limited categories underspecifies the cognitive activities underlying diagnosis; likelihood is only one element of decision making under conditions of uncertainty and risk. Failure to recognize the nature of the human’s cognitive task can lead to error prone links in the joint system (but see Schum, 1980; Einhorn & Hogarth, 1984; Robinson & Sorkin, 1985 for treatments of how evidence supplied by one system should be used by a subsequent decision maker.)

The hypothetical computer consultant provides some form of problem solution. However, it is the human problem solver who has responsibility for the outcome. Of course, the person has in principle the authority to override the machine—that is, to filter the expert machine’s output. This form of cognitive coupling between subsystems has several strong implications. First, does the user really have the authority to override machine output in practice? Since the user’s only practical options are to accept or reject system output, there is great danger of a responsibility/authority double-bind in which the user either always rejects machine output (perhaps by finding or creating grounds for machine unreliability) or abrogates his or her decision responsibility (the user may not override the computer, regardless of circumstances, if the cost of an error in overriding is too high). When people refer a problem to a human specialist, they generally pass on both authority and responsibility together (e.g., Miller, 1983); thus, a specialist who is called in on a case typically acts as the primary problem solver and not as a consultant to another problem solver. The responsibility/authority double-bind has been observed with non-AI decision aids that provide solutions (e.g., Fitter & Sime, 1980) and with increases in control automation that fail to address the operator’s new role as supervisor of automated resources (e.g., Hoogovens Report, 1976).

Second, how good are people at discriminating correct from incorrect machine solutions, and how does discrimination performance vary with user expertise and over different types and depths of explanation? Very little is known about what factors affect human performance at filtering another decision maker’s solutions. What level of expertise is needed to recognize erroneous machine output or a situation that is beyond the capabilities of the machine? (Machine experts are at best only usually correct.) Can people use syntactic cues (this output looks funny for this type of problem) or experience-driven associations (in this situation the machine usually screws up) to filter erroneous system output?

A related issue is the question of loss of skill. Some degree of expertise would seem to be required to filter machine output; what factors determine if the user of a machine expert can develop or maintain that expertise? Learning by doing applies to cognitive as well as to perceptual-motor skills:

It has to be recognized that in giving up the interplay between knowledge and its regular practical exercise, we are departing from the only conditions we know for the successful development of art and science (Council for Science and Society, 1981).

Issues concerning the loss of cognitive skill are closely related to the loss of skill questions that arise in control automation:4 How will the human acquire or maintain the manual (or cognitive) skill to take over or adjust control when automation breaks down or when disturbances occur that are beyond the capability of the automation (Hoogovens Report, 1976)? Does man-in-the-control-loop (or decision-loop) architecture improve human fault management performance in highly automated systems (e.g., Ephrath & Young, 1981; Wiener, 1985)?

While research developers of machine experts have accepted the need to explain offered solutions in some form

---

4Human performance at filtering machine output can be modeled in a signal detection paradigm (e.g., Sorkin & Woods, 1985) where performance is a function of two parameters: A sensitivity component and a response criterion. The first measures the person’s ability to discriminate correct from incorrect proposed solutions, while the second measures how much evidence it takes to accept or reject a solution given the costs and benefits of the different possible outcomes.

5I am addressing here the person’s ability to filter machine output after the machine expert has been developed and deployed. Typical current methods for constructing a machine expert are extremely dependent on the ability of human domain experts and knowledge engineers to detect and correct erroneous machine output.

4In general, work on human performance in highly automated process control systems (e.g., Wiener & Curry, 1980; Sheridan & Hennessy, 1984; Wiener, 1985) provides a starting point to consider potential challenges and problems in applying tools for automated decision making.
and to some degree, there are pressures in commercial applications which encourage the production of black box systems. For example, I have seen developers of commercial systems express concern about providing explanations, when the “explanation” consists of the rule trace that led to the solution, because a series of these explanations could constitute enough information to aid competitors in developing or marketing comparable systems. Similarly, in diagnostic applications, developers sometimes dismiss efforts to provide good explanation, which might increase the user’s skill at diagnosis and treatment of domain failures, because they fear that the skilled person could then leave and compete with the parent company.

One plausible strategy for filtering machine solutions is for the user to generate his or her own problem solution and then to compare it with the machine’s output. Besides the responsibility/authority double-bind problem, this strategy results in a redundant, as opposed to diverse, joint human-machine cognitive system architecture. Analogous to equipment reliability issues, the design question is whether a redundant or diverse architecture will result in more reliable overall system performance.

The critical question is the criterion for judging an effective system. In the paradigm represented by the hypothetical consultant, the “system” is defined as the machine expert and “effective” means usually correct machine solutions. However, an alternative approach is to define the system as the combination of human and machine (the human-machine cognitive system) and to define effective to mean maximizing joint performance; the performance of the whole should be greater than the possible performance of either element alone.

In sum, what kind of decision tool is the hypothetical consultant?

One of the big problems is the tendency for the machine to dominate the human... consequently an experienced integrated circuit designer is forced to make an unfortunate choice: Let the machine do all the work or do all the work himself. If he lets the machine do it, the machine will tell him to keep out of things, that it is doing the whole job. But when the machine ends up with five wires undone, the engineer is supposed to fix it. He does not know why the program placed what it did or why the remainder could not be handled. He must rethink the entire problem from the beginning (Finegold, 1984).

Given limited user participation in the problem solving process, the danger of a responsibility-authority double bind with support systems that offer solutions rather than informative counsel, the potential loss of cognitive skill and the potentially difficult and unsupported task of filtering poor machine solutions, the impoverished joint cognitive system implicit in the hypothetical computer consultant does not represent an effective paradigm for the use of decision tools, i.e., for decision aiding. As a result of the poor cognitive coupling, the performance of the joint system is not likely to exceed or may even be worse than (Robinson & Sorkin, 1985) the performance of the machine alone.

Towards Joint Cognitive Systems

Tool builders have focused, not improperly, on tool building—how to build better performing machines. But tool use involves more. The key to the effective application of computational technology is to conceive, model, design, and evaluate the joint human-machine cognitive system (Hollnagel & Woods, 1983). Like Gestalt principles in perception, a decision system is not merely the sum of its parts, human and machine. The configuration or organization of the human and machine components is a critical determinant of the performance of the system as a whole (Sorkin & Woods, 1985). This means using computational technology to aid the user in the process of reaching a decision, not to make or recommend solutions.

The challenge for applied cognitive psychology is to provide models, data, and techniques to help designers build an effective configuration of human and machine elements within a joint cognitive system.

Effective joint cognitive system design requires, first, a problem-driven, rather than technology-driven, approach. In a problem-driven approach, one tries to learn what makes for competence and/or incompetence in a domain and then to use this knowledge to provide tools that support domain cognitive demands, help the human function more expertly, and mitigate error prone links in the joint cognitive system. If the problem to be solved by the new system is a dangerous environment, then an automated decision system is a viable solution. If the problem is human inconsistency or memory lapses, then a redundant cognitive system architecture may be one appropriate path. It is insufficient to say, “human diagnostic performance (even by experts) is not as good as I would like, therefore I will build a machine for diagnosis.” One must ask what aspect of the diagnostic performance of the current person-machine system is the bottleneck. Studies of cognitive performance in work environments have shown person-machine performance problems such as:

- Fixation or perseveration effects in an operator’s assessment in the state of some process (Woods, 1984; Norman, 1986);
- Weaknesses in counterfactual reasoning: Would y have occurred if x had not?
- Data sampling/information acquisition problems: Can the user find and integrate the “right” data for the current context and task (Woods, 1986)?

Can machine experts that offer problem solutions counteract any of these problems? For example, one characteristic of fixation effects is early termination of the evaluation of alternative hypotheses; therefore, a good joint
cognitive system should support a broader exploration of solution possibilities. Would a machine expert offering its own solution broaden the evaluation of alternatives or narrow evaluation and exacerbate fixation problems? If failures of attention are the underlying problem, then a decision aid that helps the human problem solver to focus on the relevant data set for the current context is the design goal (Woods, 1985); computational technology supplies the means to build real systems that instantiate the decision aiding techniques to achieve this goal. When tools dominate, rather than constrain, the joint cognitive system design, the designer runs a strong risk of committing the error of the third kind: Solving the wrong problem (Mitroff, 1974).

If joint cognitive system design is to be effective, we need models and data that describe the critical factors for overall system performance. Sorkin and Woods (1985) and Robinson and Sorkin (1985) are examples (cf. also, Schum, 1980) of an analysis of joint cognitive systems modelled as two decision makers in series. The first stage consists of automated subsystems that make decisions about the state of the underlying process. When alerted by the first stage, the subsequent human decision maker uses evidence provided by the automated subsystems and his or her own analysis of input data to confirm or disconfirm the decision made by the automated monitors and to decide on further action. These analyses show, first, that the performance of the joint system can be significantly enhanced or degraded relative to the performance of the machine element alone. Overall performance depends on interactions between the characteristics of the subsystems, primarily the response criterion of the automated subsystem and the user's workload and monitoring strategy. Second, the value of the output of the first stage is better thought of as information, in the sense of evidence or testimony, to be used by the person to aid his or her decision, rather than as an offered solution to be accepted or rejected. Third, the inferential value of the information provided to the human decision maker is highly sensitive to the characteristics of the joint system. For example, the value of the evidence provided by the automated subsystem degrades rapidly if it exhibits a bias for or against possible events, even if it is a sensitive detector alone.

Empirical studies of human-human advisory interactions are a another source of data on what is good advice. Alty and Coombs (1980) and Coombs and Alty (1980) found that unsatisfactory human-human advisory encounters were strongly controlled by the advisor. The advisor asked the user to supply some specific information, mulled over the situation, and offered a solution with little feedback about how the problem was solved. While a problem was usually solved, it was often some proximal form of the user's real problem (i.e., the advisor was guilty of a form of solving the wrong problem: Solving a proximal case of the user's fundamental or distal problem). The advisor provided little help in problem definition. There is a striking parallel between these characteristics of unsatisfactory human-human advisory encounters and the characteristics of the joint cognitive system implicit in the hypothetical computer consultant analyzed earlier.

By contrast, in more successful advisory encounters a partial expert (an experienced computer user with a domain task to be accomplished) consulted a specialist (an expert in the local computer system). Control of the interaction was shared in the process of identifying the important facts and using them to better define the problem. In this process each participant stepped outside of his own domain to help develop a better understanding of the problem and, as a consequence, appropriate solution methods.

These studies (see also, Pollack et al., 1982) reveal that good advice is more than recommending a solution; it helps the user develop or debug a plan of action to achieve his or her goals (Jackson & Lefrere, 1984). Good advisory interactions aid problem formulation, and plan generation (especially with regard to obstacles, side effects, interactions, and trade-offs), help determine the right questions to ask, and help find or evaluate possible answers. A good advisor must be able to do more than provide a solution and some description or justification of the solution process; he or she must be able to participate in the problem solving process, to answer questions like: What would happen if z? Are there side effects to z? How do x and y interact? What produces x? How to prevent z? What are the preconditions (requirements) and postconditions for x (given z, what consequences must be handled)?

Studies of advisory interactions reveal another important characteristic of joint cognitive systems: The relationship between the kinds of skills and knowledge represented in the human and the kinds represented in the machine (as opposed to relative skill levels). The assumption of essential user incompetence inherent in the hypothetical system described earlier is almost always unwarranted. Instead, the human and machine elements contain partial and overlapping expertise that, if integrated, can result in better joint system performance than is possible by either element alone.

Today, no one expert in any field can keep up with the amount or rate of change of information. The result of this fragmentation of knowledge is the generalist-specialist problem. Most real world problems require the integration of different specialists each of whom contributes a unique point of view (Hawkins, 1983; Coombs & Alty, 1984), and one aspect of expertise is the ability to integrate specialist

---

5This assumption is another negative by-product of the emphasis on producing a machine expert that usually outputs correct solutions, where “usually” means that the machine alone outputs a correct solution more often than the expected human practitioners alone, and Hollnagel, 1986 examines the implications of this emphasis for the evaluation of decision aids.
knowledge in some real problem context. As a result, the designer of a joint cognitive system must address the question of what is or what should be the relationship between machine and human expertise in a given domain: Is the person a generalist who manages and integrates input from various machine implementations of specialist knowledge? Is the human one specialist interacting with the knowledge of other specialists to deal with a problem at the junction of two fields? Effective interactions between these kinds of partial, overlapping experts require that knowledge from different viewpoints be integrated in the decision process, including problem formulation and plan evaluation.

Researchers are beginning to develop decision support systems that embody these characteristics of useful, informative advice. One path is to build machine advisors that critique the human problem solver’s plan (Coombs & Alty, 1984; Langlotz & Shortliffe, 1983; Miller, 1983). A second path towards joint cognitive systems is to build direct manipulation (Hutchins, Hollan, & Norman, 1985) or graphic knowledge (Woods, 1986) systems. Both paths use explicit knowledge about the cognitive demand characteristics of a domain, about the state of the problem solving encounter, and about the user’s plans/goals to provide useful advice. The former incorporates this knowledge in an explicit machine advisor; the latter embeds this knowledge in a graphic, conceptual looking glass through which the user views the domain. Rather than offering solutions, systems built from either of these approaches support user problem formulation and plan evaluation by providing informative counsel such as warnings of prerequisite violations, reminders of potentially relevant information, reports of potential consequences or side effects, and reminders of a plans’ postconditions.

Implications

Fundamentally, the difference between the paradigm for decision support represented by the hypothetical consultant and the joint cognitive system paradigm is a difference in the answer to the question “what is a good consultant.” One operational definition of a consultant (operational in the sense that systems purported to be consultants are built in this fashion) is some one (thing) called in to solve a problem for another, on the assumption that the problem was beyond the skill of the original person. Given this definition, the important issues for building decision aids is to build better automated problem solvers and to get people to call on these automated problem solvers (the acceptance problem). The joint cognitive system perspective, on the other hand, defines a consultant as a resource or source of information for the problem solver. The human problem solver is in charge; the consultant functions more as a staff member. As a result, the joint cognitive system viewpoint stresses the need to use computational technology to aid the user in the process of solving his or her problem. The human’s role is to achieve total system performance objectives as a manager of knowledge resources that can vary in “intelligence” or power (Sheridan & Hennessy, 1984).

To build effective human-machine cognitive systems, we need techniques and concepts to identify the decision-making/problem-solving requirements in some domain and to improve cognitive performance. The analysis of a hypothetical computer consultant from the joint cognitive system viewpoint reveals some of the kinds of problem-solving errors that can occur when one attempts to build a problem-solving system (e.g., pseudo-consultants, overuse of redundant as opposed to diverse human-machine cognitive system architectures, design for novice human/expert machine interaction when the interaction is actually between overlapping partial experts). Research to provide designers with a cognitive technology is underway (e.g., Card, Moran & Newell, 1983; Rasmussen, Leplat & Duncan, 1986; Norman & Draper, 1985; Hollnagel, Mancini & Woods, 1986), but this cognitive technology is more than one ingredient for the development of “intelligent” machines. It is a valuable entity in its own right and can support many avenues for performance improvements including decision training, interface design, human reliability assessment, non-AI decision aids, and information management.

References


AAAI-86
CALL FOR PARTICIPATION

Tutorial Proposals
Due by: January 7, 1986

Mark Stefk
Tutorial Chairman for AAAI 86
Xerox Palo Alto Research Center
3333 Coyote Hill Rd
Palo Alto, California 94304
(415) 494-4366
Stefik@XEROX.ARPA

Scientific Workshop Proposals
Due by: March 1, 1986

Stanley Rosenschein
Program Co-Chairman for AAAI-86
SRI International
333 Ravenswood Avenue
Menlo Park, California 94025
(415) 859-4167
Stan@SRI-ALARPA

Application Workshop Proposals
Due by: March 1, 1986

Tom Kehler
Program Co-Chairman for AAAI-86
IntelliCorp
1975 El Camino Real West
Menlo Park, California 94040
(415) 965-5500
Kehler@USC-ECL

Panel Session Proposals
Due by: March 1, 1986

Stanley Rosenschein or Tom Kehler
Program Co-Chairs for AAAI-86
AAAI
445 Burgess Drive
Menlo Park, California 94025-3406
(415) 328-3123
AAAI-OFFICE@SUMEX-AIM.ARPA

General Questions

Claudia Mazzeotti
AAAI Office
445 Burgess Drive
Menlo Park, California 94025-3496
(415) 328-3123
AAAI-OFFICE@SUMEX-AIM.ARPA