AI & Society 4(3), 214-240, 1990.

User Modeling in Dialog Systems: Potentials and Hazards [1]

Alfred Kobsa
SFB 314: AI - Knowledge-Based Systems
Dept. of Computer Science
University of Saarbrücken, GERMANY
kobsa@cs.uni-sb.de

Abstract

In order to be capable of exhibiting a wide range of cooperative behavior, a computer-based dialog system must have available assumptions about the current user's goals, plans, background knowledge and (false) beliefs, i.e., maintain a so-called "user model". Apart from cooperativity aspects, such a model is also necessary for intelligent coherent dialog behavior in general. This article surveys recent research on the problem of how such a model can be constructed, represented and used by a system during its interaction with the user. Possible applications, as well as potential problems concerning the advisability of application, are then discussed. Finally, a number of guidelines are presented which should be observed in future research to reduce the risk of a potential misuse of user modeling technology.

Keywords: cooperative systems, user modeling, potentials of user modeling, hazards of user modeling

1. What is user modeling?

1.1. Computers as encyclopedias

systems is primarily to help make computers more cooperative and thus more easily accessible, particularly to inexperienced users. One serious drawback to current computer-based information systems is that they make a great number of demands upon their users.

[1] This work was supported by the German Science Foundation (DFG) in its Special Collaborative Research Programme on Artificial Intelligence and Knowledge-Based Systems. Part of the research described herein was carried out while the author was affiliated with the Dept. of Medical Cybernetics and Artificial Intelligence of the University of Vienna, Austria. I would like to thank Ernst Buchberger, Ellen Hays, Otto Neumaier, Gudula Retz-Schmidt, Ingeborg Steinacker, Ulrich Weis, and an anonymous reviewer for various suggestions concerning the content and style of this paper.

In particular, every user must have detailed knowledge about

a) what information is principally relevant for his/her problem (his/her goal) and what information is completely irrelevant;

b) what portion of the relevant information is contained in the dialog system;

c) how to find the relevant information in the system.

(a) means that the dialog system is employed by the user mainly as an "encyclopedia". It is up to the user to know what parts of the information contained in the system are relevant for him/her. The only support the system can provide in this respect is to allow the user to browse through its tables of content or raw data, so that s/he might get some ideas about what information might be of interest for his/her problem. (Typically, this is the same thing the user would do if s/he had a real encyclopedia at hand.) In a system with a wide range of data there is no possibility of additional support, since the system lacks even the basic requirements for such a cooperative behavior, namely, assumptions about the purposes for which the user might need some information.

(b) and (c) mean that the user must possess a "model" of the dialog system. This model need not correspond to the actual architecture of the system. Often the user will only develop a "user perspective" (or is even supposed to only do so, as in data base theory). Nevertheless, even this reduced perspective must comprise a lot of details. In the case of information systems, for instance, it must comprise detailed knowledge about what information is stored in the system and by what means this information can be retrieved.

There are several possibilities for making this cognitive task easier for the user. One traditional means is to provide instruction manuals. Other methods concentrate on developing special system components or on observing specific design principles. Examples include mnemonic abbreviations (or even pictograms) for commands, HELP-options at system's requests, minimization of system-oriented input in comparison to goal-oriented input, window techniques to facilitate maneuvering in the menu space, etc. A number of such proposals have been discussed, particularly in the field of cognitive engineering.

This is not the place for considering the efficiency or realizability of such techniques aimed at facilitating the user's task of building a model of the system. However, it should be noted that, in spite of these methods, considerable demands remain upon the user. There is certainly no problem in expecting a user who frequently communicates with a system to develop a detailed model of the system's functioning. For a casual user of a system, however, this basic cognitive expense is usually far too high.

1.2. Computers as dialog partners

An alternative approach is based on the idea that it should not be left to the user to build a model of the dialog system, but rather the system should build a *model of the user*. This means that in the course of the interaction with the user the system should generate assumptions about all aspects of the user that might be relevant in the task domain at hand. This includes in particular

- the user's goals;
- the plans with which the user wants to reach his/her goals;
- the knowledge (beliefs) of the user about a particular domain.

The set of these assumptions forms the *user model* of the system if they can be separated *by the system* from the rest of the system's knowledge. The ability to automatically draw assumptions about the user's beliefs, goals and plans enables the system to reason about a situation also from the user's point of view. In turn, this is a necessary prerequisite for the system to be able to exhibit cooperative dialog behavior. For instance, it enables the system to

- take into account the goals and plans for which the user needs some requested information, and supply additional relevant information, if necessary;
- take into account what the user probably already knows or does not know about a situation, and thus avoid redundancy and incomprehensibility in its answers and explanations;
- detect wrong beliefs of the user and inform the user about them.

As an example, let us consider the following question-answer pair:

(1a) User: When does the Montreal train leave?
(b) System: 3:15 at gate 7.

(Allen 1983)

From the user's question in (1) the system (in this example, a system specifically designed for train information) can conclude (a) that the user has the goal (or better, already has the plan) to take the next train to Montreal; and (b), that s/he does not know the departure time of this train. These assumptions can be entered into the user model. Before the user's plan can be executed, s/he must know both the departure time and the departure gate. It is unlikely that the user knows the latter if s/he does not know the former. Since the system is cooperative and tries to support the execution of the user's plans, it will also supply him/her with the latter information, though it was not explicitly requested.

Similarily in example (2), as in (1), the system can generate assumptions about the user's goals (namely, to get some gas) and supply the user with more information than was originally requested. In (3) and (4), the system can enter the user's wrong conception of the situation into its user model and inform him/her about this fact.

(2a) User: Where is the nearest gas station?
(b) System: The nearest one which is still open is located ...

(3a) User: Which students got a grade of F in CIS500 in Spring '77?

(b) System: CIS500 was not given in Spring '77.

(Kaplan 1979)

(4a) User: I tried to remove a file with the "rm" command. But the file was not removed and the error message was "permission denied". I checked and I own the file. What's wrong?

(b) System: To remove a file, you need to be able to write into the directory containing it. You do not need to own the file.

(Quilici 1989)

Thus user models are a necessary prerequisite for a dialog system to be capable of exhibiting a cooperative dialog behavior as exemplified above. A second reason for the recent emphasis on user modeling, however, is that it has become evident in the last few years that a model of the user is also an important basis for intelligent dialog behavior in general, regardless of whether the dialog is cooperative or not. Such models are required, among other things, for identifying the objects which the dialog partner is talking about, for analyzing a non-literal meaning and/or indirect speech acts in his/her dialog contributions, and for determining what effects a planned dialog contribution will have on the dialog partner (see Allen & Litman 1986, Allen 1987). Thus, a user model does not just "sweeten" a dialog system to render it more cooperative. Instead, user models constitute an indispensable prerequisite for any flexible dialog in a wider domain. They interact extensively with all other components of the system and often cannot easily be separated from them.

Research in the field of user modeling investigates how user models can be represented in a dialog system, how assumptions about the user can be formed from his/her input into the system and from other sources, and how these assumptions can be used by the system (for surveys see Kobsa & Wahlster 1989, particularly Wahlster & Kobsa 1989). User modeling is mostly pursued within the field of natural-language dialog systems (see Kass & Finin 1988) and partly also within the field of human-computer interaction (see Murray 1987). Within computer science, relationships exist with the field of student modeling in intelligent tutoring systems (Self 1974, Dede 1986, Kass 1989), multiple-agent planning systems (Konolige & Nilsson 1980, Georgeff 1983), active databases (Botman et al. 1987), and text comprehension and generation. Other relevant research includes work on mental models in cognitive psychology and the study of human dialog in psychology and linguistics.

2. Sources for assumptions about the user's beliefs, goals and plans

There are various sources from which a dialog system can obtain assumptions about a user's beliefs, goals and plans. Up to now, the following sources have been identified:[2]

[2] Since the information which a computer system can obtain about its users is restricted to information

2.1. Default assumptions

In systems with a restricted area of application, a lot of beliefs and goals can be attributed to *any* user of the system, as long as there is no evidence to the contrary. These assumptions can be used by the system for building up an initial stereotypical user model at the beginning of an interaction with him/her. Such default assumptions may concern, e.g.,

- *The user's general knowledge*: A management information system (MIS) with a user modeling component, for instance, can assume that all of its users possess basic knowledge in business administration.
- *The user's beliefs about a domain*: An MIS installed in a firm can also assume that the user is familiar with the principal facts concerning the firm.
- *The user's goals*: All users of a hotel reservation system (Hoeppner et al. 1984) can be assumed to be searching for accommodations.

Other possible default assumptions concern the user's beliefs about the system's goals (e.g. that the system wants to rent its rooms), assumptions about the user's beliefs about the system's beliefs, etc. More refined strategies take a small number of user inputs into account before a stereotypical initial user model is constructed (see e.g. Rich 1979, Chin 1989). Techniques for overriding or revising incorrect stereotypical assumptions are discussed in Chin (1989) and Rich (1989).

2.2. Initial individual user model

Initial user models need not necessarily contain stereotypical assumptions only. One can imagine systems that store the user model they have constructed during an interaction and retrieve it at the beginning of the next session with the same user. This makes sense only if one can expect (a) that a user will frequently interact with the same system, and (b) that the interaction will mostly be about the same or similar subject areas since rather different assumptions about the user must otherwise be constructed by the system. A variant that eliminates condition (a) (and which could be already realized on the basis of the current state of technology!) would be a pocket-sized "smart" card containing user characteristics, which is inserted into the system at the beginning of an interaction (the use of such a card for emulating a user's personal software environment on different computers is discussed in Joyce & Wingerson 1983). Possible general dangers of such a card with respect of loss of privacy and easier surveillance are obvious. Problems relating more specifically to user modeling will be treated in section 4.

---

received via its input channels (i.e., currently mostly to input via a keyboard), a computer system will in many domains make far fewer assumptions about the user than a human dialog partner, who has far more sensory channels.

2.3. Assumptions based on the user's input

Assumptions gained from the user's input into the system are the most direct ones and hence also the most reliable. There are various techniques for transforming a user's input into assumptions about him/her, which differ in complexity and in the result's reliability. The most simple type occurs if the user directly expresses a belief or goal, as in

(5a) User: I want to buy your new 500 XGL.

(b) Assumption of the system: The user wants to buy one 500 XGL.

Unfortunately, such direct statements of the user about his/her beliefs or goals occur only very rarely. More frequent are cases where the *syntactic form* of the user's input allows assumptions to be made about him/her. The VIE-DPM system, for instance, draws approximately the following assumptions from a user's (German) wh-question, and enters them into its knowledge base (Kobsa 1986):

(6a) User: To whom does Peter give the book?

(b) Assumptions of the system:

- (a) The user believes that Peter and the book exist and that Peter gives the book to somebody whose identity is unknown to the user.
- (b) The user believes that the system knows to whom Peter gives the book.
- (c) The user wants to be in a situation where both the user and the system mutually know to whom Peter gives the book.

These assumptions can be drawn simply because the user has entered a wh-question, without regard to the content of his/her question. In addition, from (6a), a system can also infer:

(6b)(d) Assumption of the system: The user believes that the book is now in the possession of this person whose identity is unknown to him/her.

This inference is based upon the content of the user's question, and upon world knowledge about the meaning of 'to give' (namely that an object is possessed by the recipient after somebody has given it to somebody else). Content-dependent inferences also allow for assumptions about the user's plans and goals, as is shown in examples (1) and (2) above.

Additional assumptions about the user can be drawn from *linguistic clue words* in his/her input, as in

(7a) User (to a text editor system): *Why* is this line *not* centered?

(b) Assumption of the system: The user probably wants the line to be centered.

(8a) User: I do *not* have *enough* disk space.

(b) Assumption of the system: The user wants more disk space.

So far, the analysis of linguistic clue words in the user's input has only been investigated on a theoretical basis (Kobsa 1983). Such an analysis would offer a simple way of obtaining assumptions about the user. It seems, however, that interesting particles are not very frequently used in goal-oriented dialogs. Moreover, this method is only applicable in natural-language dialog systems.

Other techniques for making assumptions about the user based on his/her input into the system are described in Wahlster & Kobsa (1989) and Kass (1990).

2.4. Assumptions from dialog contributions made by the system

Dialog contributions of the system (i.e. answers, questions and commands) also lead to new entries in the user model. If the system had previously planned the dialog contribution by trying to anticipate its probable effect on the user, it can now hypothetically enter these effects into the user model (e.g., after answering a user's question the system can assume that the user now knows the content of the answer). These entries may give rise to the planning of follow-up information and explanation for the user.

3. Some work in the field of user modeling

Apart from a pioneering theoretical paper by Bruce (1975), research in user modeling started with Rich's GRUNDY system and the work of Allen, Cohen and Perrault. GRUNDY (Rich 1979) plays the role of a librarian and recommends books on the basis of its assumptions about the user's personal characteristics. Such characteristics include, for example, his/her educational and intellectual level, his/her preference for thrill, fast-moving plots or romance, his/her tolerance for descriptions of sexuality, violence and suffering, etc. All characteristics are represented by simple linear scale values with associated certainty ratings. The initial values stem from stereotypical assumptions which are activated by a small amount of information requested at the beginning of the interaction (e.g., from the fact that the user has a male first name GRUNDY infers that he has a high sex tolerance and a low one for romance.) To a limited extent, GRUNDY can also interrogate the user about the books it had previously recommended to him/her and thereby revise its initial assumptions. Rich (1989) discusses techniques for resolving contradictions between stereotypes and direct observation, and for automatic stereotype modification.

The work of Allen, Cohen and Perrault (e.g. Cohen & Perrault 1979, Allen & Perrault 1980, Allen 1983) concentrates on inferring probable user plans from a formal description of a user's dialog contribution (such as, in English paraphrase, the description 'Who is x such that p(x)?'). On the basis of these assumptions their system tries to detect obstacles in the executability of the hypothesized user plans. Moreover, it generates formal descriptions of its own appropriate dialog contributions which should help the user overcome these obstacles. In Perrault & Allen (1980), the authors also propose ideas for recognizing an indirect meaning in a user's dialog contribution. More elaborate techniques for recognizing user plans have been realized in Carberry (1989), and more complex planning operations for dialog contributions of the system are investigated in Airenti et al. (1984).

The VIE-DPM system (Kobsa 1985a, b) is part of the German-language dialog system VIE-LANG (Trost et al. 1983), and is based on a complex representation system for storing a wide range of beliefs and goals. Processes have been defined that make assumptions about the user's beliefs and goals from his/her natural-language assertions (Kobsa 1984), questions (Kobsa 1986) and commands. Other processes exist which generate appropriate natural-language output if specific belief and goal constellations are detected in the user model. The BGP-MS system (Kobsa 1988, 90), the user modeling component of the XTRA natural-language system (Allgayer et al. 1989), provides a facility for the maintenance of various hierarchical belief and goal spaces in which these assumptions can be represented.

The HAM-ANS system (Hoeppner et al. 1984), when run in the hotel reservation mode, has standard assumptions about what every user knows about hotel rooms, and keeps track of what it has told the user about the specific room it is offering him/her. This information in the user model gives guidance in the selection of appropriate noun phrases (Jameson & Wahlster 1982). Moreover, assumptions concerning the client's preferences with respect to hotel rooms are made and employed for appropriate suggestions (Morik 1985, 89).

The UNIX Consultant (Wilensky et al. 1986) is an online natural-language advisory system for the UNIX operating system. Users can ask the system how to do things in UNIX, request definitions of UNIX terminology, and get help solving problems with command usage. Its user modeling component KNOME (Chin 1989) tries to classify a user as a beginner, novice, intermediate or expert. Stereotypical assumptions are associated with each of these expertise levels concerning the plausibility that a user belonging to a certain level knows UNIX-related information of a certain degree of difficulty. Such assumptions about the current user's expertise are employed, for example, for expressing information in terms of only such concepts which the user probably already knows, or for pruning away those parts of an answer that the user already knows.

Quilici (1989) describes another UNIX help system which should allow the user to report problems s/he has had in interactions with UNIX, and to provide cooperative advice. Quilici concentrates on the recognition of misconceptions inferable from these reports, and on the selection of appropriate responses. In particular, he considers misconceptions involving "plan applicability conditions" (whether a particular plan should be used to achieve a goal), "enablements" (whether a particular state must exist before a plan can achieve a goal), and "effects" (whether a state will exist as a result of a plan execution).

Paris (1989) investigates how assumptions about the user's level of expertise can be employed by a dialog system for tailoring descriptions of complex physical objects: for users who can be expected to possess some background knowledge about the domain, objects or concepts are described in terms of their subparts and the properties of these subparts. Descriptions for "naive" users, on the other hand, essentially describe the processes associated with the operation of the object.

Other work in user modeling is concerned with the prevention or the correction of misconceptions. Joshi et al. (1984) developed strategies for preventing the user from drawing false inferences from the system's dialog contributions. McCoy (1989) deals with the correction of two types of user misconceptions: misclassification of objects (e.g., the misconception that whales are fish), and the user's belief that objects have certain attributes or attribute values they do not actually have (e.g., that whales have gills). For each type of user misconception, McCoy proposes three strategies for generating appropriate responses.

As can be seen, the investigated problems and the research goals of user modeling are rather ambitious (as is the case for most research endeavors in the field of artificial intelligence). The systems described in this section therefore possess only very restricted domains of discourse in order to limit the number of inferences necessary for drawing assumptions about the user as well as the general world knowledge that is employed in these inferences. Even within these restricted domains, only a few kinds of user modeling have been implemented, and for many systems it is doubtful whether the proposed techniques can be generalized in such a way that they are transferable to other domains. Research on user models is still largely an academic discipline, and one can expect that commercial systems in the next few years will perform only a very limited and simple kind of user modeling.

However, I cannot agree with Schefe (1983), who claims that even in the future it will not be feasible for a dialog system to recognize and to model the intentions of its users, and who suggests that research in the field of cooperative natural-language dialog should be abandoned in favor of the development of "uncooperative" yet easily comprehensible dialog systems with non-linguistic interaction (e.g. via mouse, menus, windows, data dictionaries and inference chains – see section 1.1). On the one hand, even if his scepticism were correct, the deeper understanding of the problems involved in cooperative dialog would certainly justify continued scientific research in this field. On the other hand, many of the alternative interaction techniques he proposes cannot be made available to a lay user without considerable training. These methods are only appropriate for users who already have some computer background. Moreover, a strong interest in a combination of natural language and certain "natural" non-linguistic forms of interaction (particularly pointing gestures) has recently emerged in artificial intelligence research (see Wahlster, forthcoming). Also, a cooperation of research in natural-language interaction and non-linguistic interaction has taken place (see e.g. Endres-Niggemeyer et al. 1990; Sullivan & Tyler, forthcoming).

4. Potentials and hazards of application

In the first part of this section, five different application purposes will be discussed in which a system's ability to model the user's beliefs, goals and plans is of considerable interest. Some of these purposes seem to be acceptable or even desirable; others pose ethical or legal problems. Naturally, in restricted domains all these applications are also possible *without* a user modeling component. However, systems equipped with such a component will certainly do much better.

When discussing possible applications, I will try to keep a “middle-range” perspective, not restricting myself to systems that already seem realizable at the state-of-the-art of the next few years, but refraining from investigating applications that seem too futuristic. For the discussion of some application purposes, prototype systems will be presented which already exhibit some interesting behavior in the respective domain. For other applications, it seems rather unclear which techniques may be used for their realization.

In the second part of this section, four independent issues will be considered which have consequences for the expedience of the application of user modeling. The gravity of these consequences thereby depends largely on the intended application purpose of the system which contains the user modeling component. For some applications the consequences seem rather harmless, for others they might prove serious.

The discussion will not deal with problems of computer application in general (e.g., with the expedience of computerized consultation systems in the Social Security area), but only with the *additional* problems which user modeling in dialog systems might bring about.

4.1. Possible areas of application

4.1.1. Cooperative information

User modeling for the purpose of making information systems more cooperative is certainly the most important, and currently also by far the most intensively investigated area of application. In section 1.2, some examples were given of how a system’s ability to draw assumptions about the user’s beliefs, goals and plans can enhance its cooperativity. As was also noted above, this ability is primarily important if a system is intended to be accessible to casual users, in particular to the general public. If combined with a natural-language interface, user modeling is an important contribution to making man-machine interaction more similar to human communication, and thus to reducing the access barrier to computer systems.

4.1.2. Interest-based information

Knowledge not shared by the system with the user may be used by the system to influence the user’s evaluation and decision-making process. This can be done, for instance, by stressing facts that will probably be positively rated by the user, by leaving out information that will be negatively rated, or by deliberately misguiding the user. Some techniques which would apply to the HAM-ANS system are discussed in Jameson & Wahlster (1982) and Wahlster (1984).

4.1.3. Supervising the user

In cases of supervision, the user's interaction with a system is monitored by a "meta-system". In this context, there are three application purposes for which user modeling can be considered:

a) *Tutoring*

Supervision of the user typically occurs in the field of intelligent computer-aided instruction, when some skill of a student is to be tested and improved. If the system is able to infer and represent the user's beliefs and problem-solving plans, it can make assumptions about what has caused a student's error (Self 1974, Dede 1986, Kass 1989). The WEST system (Burton & Brown 1979), for instance, determines which arithmetic procedure a student probably does not know and provides him/her with tutorial feedback. The DEBUGGY system (Burton 1982), on the other hand, determines what wrong steps the user has probably employed in arriving at his/her incorrect solution.

b) *Unsolicited advice*

An example of such an application is the WIZARD system (Finin 1983), which recognizes correct yet inefficient command sequences in the user's interaction with the VAX/VMS operating system. If such a sequence is detected, the system volunteers advice like

(8a) User: COPY TEST1. TEST2.
(b) User: DELETE TEST1.
(c) Supervising system: If you are trying to change the name of the file TEST1. to TEST2., you could have done this more efficiently with the RENAME command, e.g.,

RENAME TEST1. TEST2.

The REPLIX system (Hecking et al. 1988) goes beyond the simple detection of user input sequences. This system possesses an inventory of possible user plans and tries to match the user's input with one or several of these plans. The system can detect that the user has temporarily interrupted the pursuit of one plan to pursue another. Certain frequently occurring isolated commands (e.g. 'date') are ignored by the system.

c) *Control*

Modeling the user's (wrong) beliefs and problem solving plans could also be used for pure control of the user. As an example, one could think of an application for testing the user's expertise or reliability in a domain by observing his/her computer-assisted problem solving behavior in this domain. The results of such a test (particularly the speed and the error rate) could be used, for example, in deciding whether or not to hire, promote, or fire a person, or (see Pratt 1987) to report careless or reckless actions.

Such tests can be, and already are, performed with traditional systems. A system equipped with a user modeling component could additionally make assumptions about what wrong beliefs or problem solving plans of the user caused his/her error. User modeling is expensive, however, and the reasons for a user's error are most often only of restricted additional interest if mere verification of expertise and reliability (rather than educational training) is at issue. Therefore it seems unlikely that user modeling will be profitable for this purpose.

A more subtle application of user modeling for controlling the user is the prevention of what is regarded as an unacceptable access to some information in a system. As an example, let us take statistical databases which are usually accessible to a wide number of users. It is well known that, even when only aggregated information may be obtained from such a database, it is often possible for a user to infer data about individuals by generating small overlapping clusters using appropriate search expressions. A system that can monitor the user's operations and make assumptions about the presumable goals and plans behind these operations may possibly prevent attempts to identify individuals in anonymous data.

Such a control ability could also be generalized: even if one agrees that information should be democratized, i.e., be accessible to everybody interested in it, one usually also would like to keep information from being sought for unacceptable purposes. One can imagine that both can be guaranteed by a system equipped with a user modeling component, which provides free access to all its information, while ensuring that it is not possible to obtain information for pursuing "bad" goals. The problem, of course, is who determines what to regard as acceptable and unacceptable use of information. This decision should not be left to the system's designers (or their employers), whose interests may not be identical with the public good.

In general, however, given current knowledge in the field of user modeling, the realizability of such a system for controlling unacceptable retrieval of information is rather doubtful.

4.2. Some issues concerning the expedience of application

4.2.1. Users' awareness, consent and control with regard to user modeling components in dialog systems

Intuitively one would agree (a) that users should be aware that the system with which they are interacting constructs a model of their beliefs, goals and plans, (b) that users must consent to being modeled before this is done, and (c) that users should have some sort of control over the user model the system constructs. This holds true in all application purposes discussed above, particularly of course in sensitive applications such as control systems or interest-based information systems.

If the user is aware that the system with which s/he is interacting is equipped with a user modeling component, then s/he has the chance to apply all the techniques for hiding his/her beliefs, goals and plans which s/he normally uses in everyday person-to-person interaction. This includes, for example, pretending that one knows less or more about a domain, hiding one's goals and plans, being evasive, etc. In some applications (e.g. tutoring or unsolicited advice), the user is necessarily aware of the existence of a modeling component; in other applications its existence might not be apparent.

Control of the user over the user modeling component could be realized in several ways. For instance, the user could inspect the model which the system has made of him/her in the course of a dialog. However, since the number of assumptions made by the system in a normal interaction would be enormous (see section 2.3.), the user has no chance to inspect them all. Another problem is rendering these assumptions comprehensible to him/her. The contents of current knowledge representation systems are often very difficult to translate into ordinary English.

The user's control over the user modeling component could also mean that the user is entitled to alter the assumptions that s/he has inspected. In this case, too, the problem arises of how to explain the consequences of such alterations to him/her. One might argue that such an option might seduce the user to change actually correct assumptions in order to make them correspond to his/her (possibly more positive) self concept. In the case of cooperative systems and computer-based tutoring, however, it is the user's problem if the system's cooperativity declines as a result of such manipulations.

Another interpretation of control of the user over a system's user modeling component is that the user has the possibility to "switch off" this component and to communicate only with the "unintelligent" remaining system, if s/he does not consent to being modeled by the system. In some cases (e.g. tutoring) this is not possible due to the nature of the application. In others (e.g. cooperative information) it is certainly not advisable to do so. However, there are also applications (e.g. unsolicited advice) where such a demand of the user may be justified. The technical question remains, however, whether it is possible to separate the user modeling component from the remaining system in such a way that the user can in fact "switch it off". In the last few years it has become increasingly apparent that a user model is a prerequisite not only for a system's cooperativity, but even for its ability to conduct a coherent intelligent interaction at all.

4.2.2. Modeling short-term goals and beliefs vs. modeling personality traits

Rich's GRUNDY system, which was discussed briefly in Section 3, primarily models personality traits of a user. In the train information system of Allen, Cohen and Perrault, as well as in VIE-DPM, short-term goals and beliefs are modeled which will certainly be abandoned by the user in the near future. A misuse of the former assumptions is therefore obviously far more serious than a misuse of the latter.

Fortunately, there are only a few applications where it is of interest to model a user's personality traits (and in fact, GRUNDY is the only system to date in which assumptions of this kind have been modeled). Such applications primarily include systems aimed at supporting the user in decisions based on subjective preferences, such as the selection of books, films, holiday trips, forms of investment, a spouse, a job, etc. In more customary application domains, only concrete (i.e., in general, short-term) beliefs, goals and plans are relevant.

Generally it is probably true that the modeling of short-term beliefs and goals is rather harmless since they are outdated after a short period of time. However, there exists the subclass of *periodically recurring* short-term beliefs, goals and plans. Because of their short-term nature, it seems rather harmless that a system draw assumptions about them, which might induce an uninformed user to be more frank than usual. A misuse of such information, however, can at least cause a lot of annoyance. (Think of a counseling system that sells you a car based on selection criteria it has attributed to you. Then, after allowing time for the first probable repairs, it unsolicitedly advertises this year's model to you, which it claims fits your criteria much better.)

4.2.3. Reliability of assumptions

The inferences that a system equipped with a user modeling component can draw about the beliefs, goals and plans of the user are associated with varying degrees of certainty. Some assumptions made by the system, and hence also the system's replies based on them, may be wrong or incomplete. One might argue that a good solution to this problem would be to have the user verify the assumptions the system automatically draws about him/her. This, however, is not feasible.[3] Even from a simple question or assertion of the user (like those in examples (1)-(4)) the system must draw so many inferences that the user has no chance to check them all.

User modeling therefore involves the risk of misunderstandings (as is also the case in normal human communication). It can be expected, however, that the gain in cooperativity will greatly exceed this risk. Often it is only this cooperativity that allows the inexperienced user to gain access to the system at all. Moreover, in traditional dialog systems misconceptions may also occur, though only on the part of the user (e.g. about the meaning of some command, which, when undetected, may lead to an incorrect interpretation of the results.) User modeling adds the danger that misconceptions may also occur on the side of the system. On the other hand, it contributes to detecting the user's misconceptions.

In any case, a necessary characteristic of a user model must be that the assumptions stored in it be revisable if counterevidence is observed. Recent research on this problem can be found in Bonarini (1987), Finin (1989) and Huang (1990).

4.2.4. Legal restrictions

Inferences that a system draws about the beliefs, goals and plans of a user are, like any personal data, liable to unauthorized use. Such misuse may concern, for example, the

[3] Such an option, though, is offered by Kay's (1991) user modeling took kit UM. However, this system only contains a very shallow user model.

disclosure or transfer of data, their conservation for longer than necessary, their processing for unauthorized purposes, etc. For instance, one could imagine a case in which a travel booking system assumes that the user is very family-minded, or that s/he is planning a business trip to Beirut. When this assumption is disclosed to an insurance company, it might try to sell him/her a life insurance policy.

National privacy laws usually determine what is to be considered an unauthorized use of personal data. Other legislation, such as the regulation of shop steward's participation and consultation in matters concerning personnel, sets boundaries to the introduction of certain technical innovations such as supervision systems.

A check should certainly be made, however, as to whether this legislation also corresponds to the needs and possible dangers of the application of user models in dialog systems. The German Privacy Act, for instance, requires that personal data to be processed by a computer must be pointed out to the respective individuals. (In Austria, alternatively, a detailed specification of the nature of the data may be given to a register that is accessible to the general public.) In the case of traditional dialog systems this does not cause problems, since the only personal data the system can process during the dialog are those that the user has him-/herself entered into it.

In dialog systems with a user modeling component, however, assumptions about the user's beliefs, goals and plans (which are personal data in the sense of the German Privacy Act) are drawn from practically every input of the user into the system. Even a simple question of the user about some fact which has nothing to do with him/her (e.g. about whether George Bush likes broccoli) will lead to personal data about the *user* (e.g. about what the user does not know, wants to know, etc.). And in most cases, the user cannot foresee which assumptions a system will draw from his/her input.

How can one then specify which personal data a system with a user modeling component may process, in order to meet the legal stipulations? It is nearly impossible to prepare a list of all the assumptions which can *in principle* be inferred about a user in a concrete application (and thus to deliver such a list to the user or the public register). The only feasible way seems to be to present the knowledge representation scheme used by the system, together with an explanation of its semantics. Since such an account would be totally opaque to non-specialists, it obviously would not meet the legal stipulations. A reconsideration of the privacy regulations in the light of user modeling, or knowledge representation in general, seems in order, so that these regulations cover not only traditional forms of data processing and collections of clearly separable data, but also modern forms of knowledge processing and highly interconnected bodies of knowledge and vague assumptions.

5. Consequences

User modeling is a necessary prerequisite if (natural-language) dialog systems are to be capable of exhibiting a wide range of cooperative dialog behavior, and possibly even for exhibiting merely coherent dialog behavior. It is certainly an important contribution to making computers accessible to casual users, particularly to the general public, and thus to democratizing information.

However, as is the case with many other new technologies, user modeling is subject to several more or less serious kinds of potential misuse, as outlined in the previous section. In order to resist these dangers right from the beginning, it seems worth while to condense the discussion of the last section into a number of guidelines which should be observed both in research on and in the practical implementation of dialog systems equipped with a user modeling component:

1. *The user should be aware of a user modeling component*
   This requirement guarantees that the user can decide how frank to be to the dialog system, and thus indirectly control what assumptions may be drawn about him/her.
2. *The user should be aware that the dialog system can make errors and might even pursue non-cooperative interests*
   This requirement guarantees that the user regards information and suggestions of the system with care. The user should bear in mind that computers are not infallible and that their information is as "subjective" as any human expertise (though on average possibly more reliable than that of any human expert). Moreover, s/he should be aware that, as is the case with human experts, the opinions of dialog systems may be shaped by non-cooperative interests (e.g., commercial interests of those who commissioned the system).
3. *The user should be able to "switch off" the user modeling component*
   This requirement guarantees a large degree of user control over whether or not s/he will be modeled by the dialog system s/he is interacting with. Of course, if the user declines to be modeled, s/he must bear the consequences (which may consist in a drastically reduced cooperativity and even the inability of the system to carry out a coherent dialog). In some applications (e.g. intelligent tutoring), moreover, the deactivation of the user modeling component makes absolutely no sense, and in others it is questionable whether the user modeling component can be separated from the remaining system to such a degree that it is possible to disconnect it.
4. *The modeling of long-term user characteristics should be avoided*
   Any kind of misuse of information gained from user modeling will be more serious for long-term than for short-term information, since the latter usually become outdated after a short period of time. Moreover, long-term user characteristics are often regarded as more "personal" than short-term beliefs and goals.
5. *Prevention of unauthorized access should not be realized by "intention control"*
   Though it might be tempting to employ user modeling, particularly intention recognition, for preventing unacceptable exploitation of information in generally accessible data bases, it seems prudent to refrain from research in this

direction. Although it is extremely doubtful that such a system could be implemented within the foreseeable future, the risk seems to high that in certain political situations such a system could be used for political control of access to information.

User modeling in dialog systems is doubtless still in its infancy. Research in this field is just now spreading from university labs to industry. No commercial system embodying a user model has been developed to date, and one can expect that in the next few years commercial systems will perform only a very restricted and simple kind of user modeling. Nevertheless, as should be the case with any new technology, it is necessary to assess the potentials and possible hazards of user models even at the current state-of-the-art, so that decisions can be made as to whether research in this field (or in specific subfields) is justified, and restrictions be imposed to prevent certain possible dangers from the outset. For this purpose it is necessary that knowledge about this new technology should not remain within the small group of scientists now working in the field of user modeling or related areas. Instead, the knowledge should be disseminated to the general public so that additional opinions on these decisions can be sought from persons who are less subject to possible scientific bias, and from persons who will eventually be affected by the implementation of this technology. This paper should be seen as a contribution to this endeavor.

6. References

Airenti, G., B. G. Bara and M. Colombetti (1984): Plan Formation and Failure Recovery in Communicative Acts. Proceedings of the 1984 European Conference on Artificial Intelligence, Pisa, Italy, 259-268.

Allen, J. F. and C. R. Perrault (1980): Analyzing Intention in Utterances. Artificial Intelligence 15 (3), 143-178.

Allen, J. F. (1983): Recognizing Intentions from Natural Language Utterances. In: M. Brady and R. C. Berwick, eds.: Computational Models of Discourse. Cambridge, MA: MIT Press.

Allen, J. F. and D. J. Litman (1986): Plans, Goals, and Natural Language. Proceedings of the IEEE 74, 939-947.

Allen, J. F. (1987): Understanding Natural Language. Menlo Park, CA: Benjamin/ Cummings.

Allgayer, J., K. Harbusch, A. Kobsa, C. Reddig, N. Reithinger and D. Schmauks (1989): XTRA: A Natural-Language Access System to Expert Systems. International Journal of Man-Machine Studies 31, 161-195.

Bonarini, A. (1987): User Modeling in Person-Machine Conversation: A Computational Approach. In: Proc. of COGNITIVA-87, Paris, France, 377-382.

Botman, A. M., A. J. Kok and L. Siklossy (1987): Using User Models and Automatic Expertise Acquisition in Man-Machine Interaction. In: Proc. of the 7th International Workshop on Expert Systems and their Applications, Avignon, France, 389-398.

Bruce, B. C. (1975): Belief Systems and Language Understanding. Report No. 2973, Bolt, Beranek & Newman, Cambridge, MA.

Burton, R. R. and J. S. Brown (1979): An Investigation of Computer Coaching for Informal Learning Activities. International Journal of Man-Machine Studies 11, 5-24.

Burton, R. R. (1982): Diagnosing bugs in a simple procedural skill. In: D. Sleeman and J. S. Brown, eds.: Intelligent Tutoring Systems. London: Academic Press.

Carberry, S. (1989): Plan Recognition and Its Use in Understanding Dialog. In: A. Kobsa and W. Wahlster, eds.: User Models in Dialog Systems. Heidelberg: Springer.

Chin, D. N. (1989): KNOME: Modeling What the User Knows in UC. In: A. Kobsa and W. Wahlster, eds.: User Models in Dialog Systems. Heidelberg: Springer.

Cohen, P. R. and C. R. Perrault (1979): Elements of a Plan-Based Theory of Speech Acts. Cognitive Science 3, 177-212.

Dede, C. (1986): A Review and Synthesis of Recent Research in Intelligent Computer-Assisted Instruction. International Journal of Man-Machine Studies 24, 329-353.

Endres-Niggemeyer, B., T. Hermann, A. Kobsa and D. Rösner, eds. (1990): Interaktion und Kommunikation mit dem Computer. Berlin: Springer.

Finin, T. W. (1983): Providing Help and Advice in Task Oriented Systems. Proceedings of the 8th International Joint Conference on Artificial Intelligence, Karlsruhe, W. Germany, 176-178.

Finin, T. W. (1989): GUMS: A General User Modeling Shell. In: A. Kobsa and W. Wahlster, eds.: User Models in Dialog Systems. Heidelberg: Springer.

Georgeff, M. (1983): Communication and Interaction in Multi-Agent Planning Systems. Proceedings of the 3rd National Conference on Artificial Intelligence, Washington, D.C., 125-129.

Hecking, M., C. Kemke, E. Nessen, D. Dengler, M. Gutmann and G. Hector (1988): The SINIX Consultant: A Progress Report. Memo Nr. 28, AI Laboratory, Dept. of Computer Science, Univ. of Saarbrücken, W. Germany.

Hoeppner, W., K. Morik and H. Marburger (1984): Talking it over: the Natural-Language Dialog System HAM-ANS. Report ANS-26, Research Unit for Information Science and Artificial Intelligence, University of Hamburg, W. Germany.

Huang, X. (1990): Student Model Revision: Evolution and Revolution. Proceedings of the Canadian Artificial Intelligence Conference CSCSI/SCEI0-90, Ottawa, Canada (forthcoming).

Jameson, A. and W. Wahlster (1982): User Modelling in Anaphora Generation: Ellipsis and Definite Description. Proceedings of the 1982 European Conference on Artificial Intelligence, Orsay, France, 222-227.

Joshi, A., B. Webber and R. M. Weischedel (1984): Preventing False Inferences. Proc. COLING-84, Stanford, CA, 134-138.

Joyce, C. and L. Wingerson (1983): Can We Adjust to Computer Culture? New Scientist, April 1983, 72-73.

Kaplan, S. J. (1979): Cooperative Responses from a Portable Natural Language Data Base Query System. Ph.D. Thesis, Dept. of Computer and Information Science, Univ. of Philadelphia, Philadelphia, PA.

Kass, R. and T. W. Finin (1988): Modeling the User in Natural Language Systems. Computational Linguistics 14 (3), 5-22.

Kass, R. (1989): Student Modeling in Intelligent Tutoring Systems: Implications for User Modeling. In: A. Kobsa and W. Wahlster, eds.: User Models in Dialog Systems. Heidelberg: Springer.

Kass, R. (1990): Building a User Model Implicitly from a Cooperative Advisory Dialog. Advance Papers of the 2nd International Workshop on User Modeling, HI.

Kay, J. (1991): UM: A User Modeling Toolkit. To appear in *User Modeling and User-Adapted Interaction.*

Kobsa, A. (1983): Präsuppositionsanalyse zum Aufbau von Dialogpartnermodellen. Conceptus 17 (40/41), 165-179.

Kobsa, A. (1984): Three Steps in Constructing Mutual Belief Models from User Assertions. Proceedings of the 1984 European Conference on Artificial Intelligence, Pisa, Italy, 423-426.

Kobsa, A. (1985a): Benutzermodellierung in Dialogsystemen. Berlin: Springer.

Kobsa, A. (1985b): VIE-DPM: a User Model in a Natural-Language Dialogue System. In: J. Laubsch, ed.: GWAI-84, 8th German Workshop on Artificial Intelligence. Berlin: Springer.

Kobsa, A. (1986): Generating a User Model from Wh-Questions in the VIE-LANG System. In: P. Hellwig and H. Lehmann, eds.: Trends in der Linguistischen Datenverarbeitung. Hildesheim, W. Germany: Olms.

Kobsa, A. (1988): User Models and Discourse Models: United they stand... Computational Linguistics 14 (3), 91-94.

Kobsa, A. and W. Wahlster (1989): User Models in Dialog Systems. Heidelberg: Springer.

Kobsa, A. (1990): Modeling the User's Conceptual Knowledge in BGP-MS, a User Modeling Shell System. To appear in *Computational Intelligence* 6(4).

Konolige, K. and N. J. Nilsson (1980): Multiple Agent Planning Systems. Proceedings of the 1st National Conference on Artificial Intelligence, Stanford, CA, 138-141.

McCoy, K. F. (1989): Highlighting a User Model to Respond to Misconceptions. In: A. Kobsa and W. Wahlster, eds.: User Models in Dialog Systems. Heidelberg: Springer.

Morik, K. (1985): User Modeling, Dialog Structure, and Dialog Strategy in HAM-ANS. Proc. of the 2nd Conference of the European Chapter of the Association for Computational Linguistics, Geneva, Switzerland, 268-273.

Morik, K. (1989): User Models and Conversational Settings: Modeling the User's Wants. In: A. Kobsa and W. Wahlster, eds.: User Models in Dialog Systems. Heidelberg: Springer.

Murray, D. M. (1987): A Survey of User Cognitive Modelling. Report DITC 92/87, National Physics Laboratory, Teddington, England.

Paris, C. (1989): The Use of Explicit User Models in a Generation System for Tailoring Answers to a User's Level of Expertise. In: A. Kobsa and W. Wahlster, eds.: User Models in Dialog Systems. Heidelberg: Springer.

Perrault, C. R. and J. F. Allen (1980): A Plan-Based Analysis of Indirect Speech Acts. American Journal of Computational Linguistics 6, 167-182.

Pratt, J. M. (1987): The Social Impact of User Models. In: H.-J. Bullinger and B. Shackel, eds.: Human-Computer Interaction: INTERACT'87. Amsterdam: North-Holland.

Quilici, A. (1989): Detecting and Responding to Plan-Oriented Misconceptions. In: A. Kobsa and W. Wahlster, eds.: User Models in Dialog Systems. Heidelberg: Springer.

Rich, E. (1979): User Modelling via Stereotypes. Cognitive Science 3, 329-345.

Rich, E. (1989): Stereotypes and User Modeling. In: A. Kobsa and W. Wahlster, eds.: User Models in Dialog Systems. Heidelberg: Springer.

Schefe, P. (1983): Natürlichsprachlicher Zugang zu Datenbanken? Angewandte Informatik 10, 419-423.

Self, J. A. (1974): Student Models in Computer-Aided Instruction. In: International Journal of Man-Machine Studies 6, 261-276.

Sullivan, J. W. and S. W. Tyler, eds. (forthcoming): Architectures for Intelligent Interfaces: Elements and Prototypes. Reading, MA: Addison-Wesley.

Trost, H., E. Buchberger, I. Steinacker and R. Trappl (1983): VIE-LANG: A German Language Dialogue System. Cybernetics and Systems 14, 343-357.

Wahlster, W. (1984): Cooperative Access Systems. In: T. Bernold and G. Albers, eds.: Artificial Intelligence: a Technology at the Dawn of its Economic and Industrial Breakthrough. Amsterdam: North-Holland.

Wahlster, W. (forthcoming): User and Discourse Models for Multimodal Communication. In: J. W. Sullivan and S. W. Tyler, eds.: Architectures for Intelligent Interfaces: Elements and Prototypes. Reading, MA: Addison-Wesley.

Wahlster, W. and A. Kobsa (1989): User Models in Dialog Systems. In: A. Kobsa and W. Wahlster, eds.: User Models in Dialog Systems. Heidelberg: Springer.

Wilensky, R., J. Mayfield, A. Albert, D. N. Chin, C. Cox, M. Luria, J. Martin and D. Wu (1986): UC: A Progress Report. Technical Report UCB/CSD 87/303, Division of Computer Science, Univ. of California, Berkeley, CA.