Technical Report

UCAM-CL-TR-61
ISSN 1476-2986

Number 61

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

User models and expert systems

Karen Spärck Jones

December 1984

15 JJ Thomson Avenue
Cambridge CB3 0FD
United Kingdom
phone +44 1223 763500

http://www.cl.cam.ac.uk/

Technical reports published by the University of Cambridge
Computer Laboratory are freely available via the Internet:

http://www.cl.cam.ac.uk/techreports/

ISSN 1476-2986

User models and expert systems

Karen Sparck Jones

Computer Laboratory, University of Cambridge
Corn Exchange Street, Cambridge CB2 3QG, UK

December 1984

Abstract

This paper analyses user models in expert systems in terms of the many factors involved: user roles, user properties, model types, model functions in relation to different aspects of system performance, and sources, e.g. linguistic or non-linguistic, of modelling information. The aim of the detailed discussion, with extensive examples illustrating the complexity of modelling, is to clarify the issues involved in modelling, as a necessary preliminary to model building.

Keywords

user model, expert system, natural language

Acknowledgement

This work was carried out during the tenure of a GEC Research Fellowship.

User models and expert systems

This paper addresses the questions

1) What do we want user models in expert systems for?

2) What sort of things are they?

3) Where do we get them from?

4) Can we get them without 'true' natural language interfaces?

The aim is to clarify the issues connected with user modelling in expert systems. There is a lot of loose talk about user models: the need for them tends to be taken for granted, and what they should be like is equally assumed to be well understood. The issues are in fact very complicated, so proceeding uncritically is likely to cause confusion. For example, if it is assumed that modelling is required to provide a good base for explanations of system behaviour (e.g. Hasling et al 1984), what does this really imply? This paper tries to spell out what is involved in having user models in expert systems, as a necessary preliminary to research on constructing and exploiting these models. In particular, while user modelling is a feature of current natural language research, because it is clear that language-based communication is improved by modelling, it does not follow, whatever the natural language community may assume, that effective modelling in an expert system depends on the use of natural language. Thus the issues from the expert system point of view are what the modelling requirements of an expert system are, and whether they can (generally) only be met by natural language interfaces. This paper is an initial discussion of the issues, with hypothetical examples. Possible research strategies for further, more detailed investigations of the claim that natural language is required for modelling would have to take specific systems and see how modelling requirements supported by natural language could also be supported by non-linguistic interfaces. Reverse direction investigations, of whether modelling needs met by non-linguistic interfaces could be met by linguistic ones, would equally be required. These studies would be designed to find out exactly what natural language gives you, and whether this can be achieved by other means, in order to answer the question: is a natural language interface not merely suffcient (in general) for user modelling, but necessary?

The paper thus starts from the language end of things, but is intended to examine assumptions about the value of natural language for expert system purposes which have not been critically enough examined, either by those in the

language processing community who are already sold on language, or, for example, by those who call for natural language interfaces because they will provide better explanations, which is the sort of confusion of ends with means we need to avoid (Sparck Jones 1985). In the next section, therefore, I shall briefly note some relevant facts about current natural language research, for reference, before standing back to look at what is involved in user modelling in expert systems, from the expert systems point of view.

Introduction

The need for, and possible roles of, user models in natural language dialogue systems are well recognised, and research on how they can be constructed and manipulated is well established (see Wahlster 1984). The general claim that natural language understanding as such requires modelling is given a specific form: user models are seen as necessary to support cooperative systems for concrete applications purposes. This may mean either that cooperative expert systems require natural language interfaces, which in turn require user models (Boguraev 1985), or that cooperative systems require user models, which in turn require natural language interfaces. In communication via language interfaces, user models face two ways: they support the interpretation of linguistic inputs, for instance by allowing the extraction of implicit questions from explicit ones, and they support the generation of linguistic outputs, for instance by allowing user-tailored amplifications of the raw answers to questions. Exemplary systems with natural language interfaces have ranged from those where the initiative lies primarily with the user (e.g. database query) to those where the initiative lies primarily with the system (e.g. Davies et al 's 1985 tutor), with advice systems somewhere in the middle (e.g. McKeown's 1984 student advisor). Some of these systems have been referred to as expert systems, but from the conventional expert system point of view the system's function is trivial (e.g. identifying train departure times (Allen 1983)). More significantly, the expert system is not seen as an independent module with a language interface: the approach from, and emphasis on, natural language has tended to absorb the expert system function within the interface. The tendency has been to treat the system as a single integrated whole with the expert system reasoning symbiotic with the natural language driven user modelling (or user model based natural language processing). The fact that natural language interfaces for expert systems, whether or not they are associated with any serious user modelling, imply a much closer coupling between front and back ends of the system than is typical of current database systems, for example, is not at issue; my point is rather that in research on user models in the context of natural language interfaces, there are sufficient challenges in, for instance, inferring the user's assumptions from his questions, even for very simple expert systems. Thus in general, in these cases, there has been no pressure for powerful expert systems as such, and the sophistication of the linguistic processing and user modelling is not matched by that of the ostensible system function, for example choosing a course for a student.

Proper natural language interfaces for serious expert systems are at an early stage (Carbonell et al 1983); they are seen primarily as a means of providing a more convenient (flexible etc) means of communication for and with the user, for such user purposes as volunteering new information or seeking explanations, and such system purposes as requesting information in a comprehensible manner and providing readable explanations. Handling the system consequences of allowing free natural language communication, even to support quite straightforward interpretation and generation needs, and certainly to deal with such likely extensions of user activity as asking meta-level questions, is difficult enough, and no material work seems to have been done so far on individual user modelling related to natural language expert system interfaces. Nor has much been done, in the artificial intelligence/expert system community, outside such specialised areas as computer-aided instruction or design, to the idea of user modelling not supported by natural language dialogue. Certainly enough problems arise, as Kukich (1984) shows, in providing an interface capable of dealing even in a straightforward way with, for example, requests for explanations, when a new natural language front end is imposed on a substantial existing back end system, to account for the limited work on individual user modelling that has so far been done within the general expert systems community. [1]

For this paper I shall start from the position that a substantial expert system already exists for some purpose or other and consider the contribution user modelling could make to such a system, adopting a fairly realistic view about user interfaces and about the scope of expert systems, i.e. we are not talking about the totally unrealistic ideal. I am interested in laying a groundwork for sensible strategies for investigating the role and utility of user models for typical expert systems, as an essential preliminary to actually providing them. I shall therefore not take it for granted initially that natural language is required, but leave the means of communication between user and system quite open, returning to the question of natural language dialogue later. Note however, that although I am taking a serious expert system for granted, I am not assuming that providing it with substantial user modelling capabilities is simply a matter of putting an extra front end box onto the system, as indeed it cannot be assumed that putting a more powerful user interface of a straightforward kind, without user modelling, onto the system is a mere business of addition. Effective functionality in a front end, with or without specific user modelling, is likely to require at least close coupling of front and back ends, and indeed the extension of the back end to support the operations provoked by the front end. As Young (1984) puts it, in the form of a 'Third Law of Expert Systems', expert systems have to contain knowledge both about the subject area of the system and about how to communicate with the user. But more particularly, the end product of user modelling may be a completely redesigned system, even in those cases where user modelling cannot be claimed as central to the whole system, as it is in teaching. Thus Boguraev (1985) comments on the need to recognise that the user model

---

[1] This is not to imply that user modelling has not been investigated elsewhere: it is an active area of concern for the human-computer interface community. But whether the interpretation of "user model" is the same there as that developed in this paper, or whether the modelling requirements and methods emphasised there are the same as those of the artificial intelligence/language processing community, is a large topic in itself, and so will not be considered here.

may be distributed over the system, and not confined to its own modelling component. The object of assuming an existing expert system is thus simply to emphasise the fact that we have a system designed to do something non-trivial, like monitoring an industrial plant, diagnosing a disease, designing a circuit, or computing tax liability.

I do not believe, however, that this is the place to go deeply into what makes a user model worthy of the name "model", that is, into what distinguishes a user model from a user description. I shall somewhat briskly assume that constructing a representation of the user and exploiting it to determine system operation, and specifically exploiting the representation to reason about the user, further developing the representation and using this to control the system's reactions, constitutes having a model. I am thus adopting a relatively loose or informal characterisation of model, but this is sufficient because the objective of the paper is to see whether anything satisfying this characterisation, in a weaker or stronger form, is necessary or desirable in an expert system. The paper will thus further explicate the notion of model when applied to expert system users.

Human possibilities and system purposes

In identifying the possible role of user modelling in expert systems, it is necessary to consider three matters:

(1) the location of human beings in the system;
(2) the property types of human beings;
(3) the global purpose of the system.

In discussing these I shall make use of hypothetical, but hopefully not unrealistic, systems as examples.

(1) In locating humans in expert systems, I shall simplistically refer to the system's front and back ends, i.e. I shall separate the system proper, where the essential decision-making processes operate, from the communication interface used to acquire information relevant to the system's operations and to present outputs of its operations. (This distinction is a functional one: I am not implying that user models live, for example, in the front end, or that the front end can be a largely independent 'service' module.)

We then have the following cases:

1) There are no humans anywhere in the system, i.e. the back end does not refer to humans, and the front end does not involve them:
e.g. an automatic controller of mechanical equipment.

2) The back end refers to humans, but the front end does not involve one, i.e. there is a human 'patient', adopting this term to refer to the person to whom the system's decision-making processes apply:
e.g. an automatic monitor of a sick person.

Note, however, that the definition of patient does not apply only in medical contexts: it applies to the human subject of any expert system. In fact, if the patient in not even known (in the sense to be amplified later), let alone is consciously participating in the system's operations, but exists only through recorded decision-relevant data, this case perhaps reduces to (1) above. If the patient is consciously participating, on the other hand, this case reduces to (4) below.

3) The back end does not refer to humans, but the front end involves one, i.e. there is a human 'agent':
e.g. a design system for circuit layout.
That is, the agent is the person at the communication interface with the system.

4) The back end refers to humans, i.e. there is a patient, and the front end involves humans, i.e. there is an agent. This is the tricky case, because there are several different subcases:

a) the patient and the agent are the same:
e.g. someone inquiring about social security benefits for himself;

b) the patient and the agent are different:
e.g. someone inquiring about social security benefits on someone else's behalf.

Unfortunately, this case is particularly complex because we can distinguish further between the situations where

i) the patient is absent, and is not significantly known to the agent;
ii) the patient is absent but is well known to the agent;
iii) the patient is present and is participating, i.e. we have a sort of agent-patient two-headed hydra.

These may seem over-fine distinctions, but I believe they are all relevant to the issue of user modelling, as will be shown below. Separating exactly three cases is of course artificial: the reality is a continuum; but it is useful for discussion purposes to consider three fairly different examples.

What the cases listed show is that "user" is an ambiguous term, in that the agent may or may not be the same as the patient. I shall therefore only use the term where the distinction is not material. Further, the cases suggest, as will become apparent below, that if user modelling is most important where there is a human agent, it is not clear that it is only relevant to the cases where there is a human agent, or that it should be confined to the human agent: the different situations listed under (4) point to more complex possibilities.

(2) The points just made lead naturally to the matter of types of user property. For the present purpose we can distinguish the 'objective' and 'subjective' properties of users. Objective properties are ones such as age, sex, Christian name, medical symptom X, etc. These properties can be divided into

those that are decision properties for the back end system, i.e. ones that define the system's decision classes, and ones that are non-decision properties. Thus for a specific medical system, age, sex, and possession of symptom X may be decision properties, determining the patient's class, but Christian name may not. (Of course a decision property for one system may be a non-decision property for another.) It is clear that decision properties refer only to patients: agents independent of patients cannot have decision properties. However non-decision properties of either patients or agents may have a role to play in user modelling, as seen below. Subjective properties are defined here to be the user's general beliefs, intentions, goals, etc, including his beliefs, etc relating to the domain within which the expert system operates. Thus in the medical case the user's subjective properties may refer, for example, to society, medicine or measles. From the system's point of view subjective properties might be claimed to be 'out there', and hence objective; however it is intuitively useful to distinguish such cognitive or psychological properties of users. Whether subjective properties can be decision properties, and whether subjective properties of patients independent of agents are important, are questions to consider. (The user's specific mental attitudes to the given expert system, and in particular his beliefs, goals etc during interaction with the system, are considered later.)

In the remainder of this paper I shall abbreviate objective properties defining decision classes as DPs, objective properties not defining classes as NPs, and subjective properties as SPs. (Where the distinction between DPs and NPs is immaterial, I shall refer simply to objective properties.)

(3) The third matter bearing on user modelling, namely the overall purpose of the expert system, is as significant as the other apparently more relevant ones, though it is not easy to anatomise. The issues here refer to the implications of the overall system purpose for the role of the user in it, and hence for the value and scope of user models in the system's operations. For example, it is not difficult to see that there could be different user modelling requirements and possibilities when we compare a system with a human patient but not agent, for example one for medical equipment monitoring, with a system with a human agent but not patient, say one for building design. But even within systems with the same human location pattern, the modelling potential may vary, as determined by the prime purpose of the system. For instance for systems with human patients and not agents we can compare a medical monitoring system with a bank customer account optimiser; for systems with human agents but not patients we can compare, for instance, a circuit designer and an investment analyser; and for systems with both human patient and human agent (say the same person), we can compare a programming instruction system with a library advisor. It is clear that systems intended to process people in the obvious sense that a teaching system does demand a serious user model; but it is not so clear that a system for industrial plant control, say, does.

Assertions that all expert systems are essentially planning systems, or diagnostic systems, are not very helpful here; more detailed categorisations like Stefik et al's (1983) into interpretation systems, monitoring systems etc are somewhat more useful, but it is perhaps necessary to look at individual system

purposes, i.e. to system properties at the application level, for real insight into the contribution modelling can make. Thus a laboratory experiment planner may be quite different from a holiday planner.

More global differences of expert system role may also, naturally, have implications for user modelling. Thus Young (1984) distinguishes the 'intelligent front end' role of an expert system from the 'advisor' role. Whether the expert system is intended to supplant the user in relation to some back end activity, or to assist him in accessing it could, for example, lead to models placing different emphasis on the user's understanding of how the back end operations are done.

Model functions

Given these preliminary points about the possible locations of humans in expert systems, the types of property they may have, and the influence of the system's specific purpose, we can now consider in more detail what a user model in an expert system could be for, i.e. what functions it could have. For this we continue to assume a simple expert system scenario, with a back end assigning things or patients to decision classes, and with a front end gathering information, either directly or indirectly, and providing information, not necessarily using natural language as the medium of communication.

What, then, are the aspects of expert system performance to which user properties are related, or on which they bear, so implying needs for user models?

We start by distinguishing the expert system's 'effectiveness' from its 'efficiency': effectiveness refers to the correctness of the decision reached, efficiency to the 'ease' with which it is reached. This distinction will be elaborated during the discussion: I am assuming it is adequate as stated to begin with. Thus the question I shall now address is:

how could user models relative to different human locations and property types contribute respectively to system effectiveness and system efficiency?

The DPs of patients are of course directly relevant to the system's effectiveness; they indeed determine it, in the sense that the system's decisions are characterised in terms of the patient's DPs. Thus where systems have human patients, their decision classes constitute primary user models, or, to put it the other way round, a user model couched in terms of DPs is not optional, but obligatory. (Note that a user model based on DPs may not be confined to the known DPs of the patient: assignment to a class may allow other system properties to be predicted of the patient; prediction is what user modelling is all for.)

The NPs of patients, on the other hand, cannot, by definition, contribute directly to effectiveness. If they are exploited because they are systematically correlated with explicit DPs, they are implicitly DPs too: for example in the medical illustration, inferring male sex (a DP) from Christian name (notionally an NP).

The SPs of patients also seem to be completely irrelevant to effectiveness for expert systems like disease diagnosis programs. However it is not so clear that in other applications, SPs have nothing to do with effectiveness, as opposed to efficiency. For example, in a welfare benefits advisory system, the patients SPs may be related to DPs: the patient's background beliefs (as also the user's more specific system-oriented beliefs considered later) may lead him to ask certain questions and provide certain data that are 'misconceived' or 'wrong' in relation to his true needs and characteristics. However this does not imply that the decision class reached on the basis of the information given is incorrect: thus the SPs could or should be more properly regarded as having to do with (in this case misleadingly or negatively) the efficiency with which the true decision class is reached. (The analogy in the case of systems with non-human patients is where inappropriate information is supplied by the agent because of his beliefs about the application domain, say.)

Whether reaching what is in fact the wrong decision on the basis of inappropriate information is a matter of system effectiveness or system efficiency is perhaps dependent on whether one is judging in absolute or relative terms, or adopting a narrow or broad evaluation of the system. From this point of view one could also see cases where the patients NPs could interfere with the system's decision-making operations: for example being anaesthetised could suppress the manifestation of symptoms relevant to disease diagnosis. Thus from the system's point of view, models of the patient's NPs and SPs might be maintained as indirect contributors to effectiveness.

In all this it is important to distinguish SPs that are in fact hidden objective properties from those that are not. Thus in a student course advice system, the patient's underlying goals could properly become DPs (see further later). However his beliefs about what the system covers, which may motivate his particular questions, may not necessarily be taken over as objective properties, but may remain as SPs to be exploited, for example for efficiency purposes in providing a context or point of view for search.

Note that in all of this I am assuming that the specific properties, whether objective or subjective, have been correctly transmitted to and received by the system, and indeed more generally, that the property descriptions being used are neither erroneous nor dishonest. That is to say, I am assuming that the statement that, for example, the patient has spots is true, and equally that the statement that he believes, say, the tax authorities are out to get him is true, whatever the tax authorities may actually be doing. Identification, description, and transmission errors (or frauds) produce difficulties, but not, I claim, special or unique difficulties for user modelling per se. We may perhaps have more difficulty in identifying errors in the characterisation of SPs, but such checking is not different in kind from that required for objective properties. The only specific point about correctly described but wrong SPs is that the need to cope with them may be a particularly pressing reason for user modelling, as seen further below.

Turning to the relation between the agent and system effectiveness, it is clear that the properties of agents as opposed to patients (whether they are actually separate people or merely separate roles), have nothing in their own right to do with system effectiveness: the system's decisions are necessarily only about patients. The agent's NPs and SPs, however, may have the same indirect relationship to effectiveness, or alternatively efficiency, as the patient's SPs, for the reasons just considered, and so could be modelled for their indirect support for effectiveness.

I glossed "efficiency" as the ease with which the expert system's decision is reached. This can, however, be considered from two slightly different points of view. One emphasises the internal operations of the system, for instance minimising the search effort required to reach a decision, e.g. by getting one piece of information before another, or checking information consistency or coherence (though this might be deemed to fall under the heading of effectiveness). The other point of view emphasises the external communications of the system in acquiring information, e.g. phrasing a request for data in a manner best suited to eliciting it, or in supplying it. Aspects of communication with the user like the provision of explanations, jusifications etc are only indirectly related to system efficiency (they may, for example, promote useful inputs), and are more properly referred to the system's 'acceptability'. The same applies further along the line towards vulgar user-friendliness. The relation between user properties and acceptability will therefore be considered separately. The division between efficiency and acceptability is somewhat blurred, but we can make a rough distinction between what is input to or output from the system from the information point of view, i.e. is directly involved with the system's decision making processes, and how it is packaged, amplified, etc. Efficiency would seem to be more often input loaded, acceptability more output loaded.

If we now consider the possible contributions to, or relations between, users and system efficiency, taking patients first and then agents, we find the following.

The patient's DPs may be exploited for efficiency purposes as well as for effectiveness; for example in a medical system to investigate possible diagnoses involving less nasty or time consuming tests first, or in a tax system to avoid spending effort seeking information about properties poorly correlated with known ones. (These constraints on search processes have of course to be checked in case they incidentally lead to incorrect decisions: the problem of tradeoffs between effectiveness and efficiency is well known.) NPs of patients may be similarly exploited: for example if the symptoms of a disease are not sexually differentiated, but in fact women suffer from a disease more than men, the NP of sex may be used to guide the search process.

The patient's SPs may have more subtle and varied influences, of the kind discussed in connection with effectiveness, so they too may be positively exploited to promote efficiency. For example, a sick person's anxiety for treatment could be used to get key but embarrassing facts from him quickly.

Clearly, where the agent is independent of the patient, both his NPs and his SPs can bear on system efficiency (as noted, he has no DPs). For instance, whether one is dealing with a novice or experienced equipment repairman could lead to different styles of interaction (e.g. different terminologies) in the interests of data acquisition. The SPs of the agent may also matter, including both his goals (for example those determining the way he sets about his task on behalf of the patient) and his beliefs. For instance, the repairman's fervent desire to get the horrible thing mended properly this time, or his belief that Jimbo bottle capping machines are lousy.

Finally, we must consider different user properties for their bearing on system acceptability, i.e. for the way they contribute to the system's acceptability to the user, by being exploited by the system to enhance acceptability. In principle many matters of user friendliness like speedy responses and legible screen output fall under the heading of acceptability, but I shall not be concerned with items of the more 'mechanical' kind relevant to all users, for which no individual user model is needed. Taking the patient first, in relation to acceptability, his DPs are necessarily relevant in the obvious sense that if the set from which these properties are drawn, or the individual properties themselves, constitute a poor basis for the system's decisions, these will not be acceptable to the user. For instance if the descriptive resources on which an equipment failure system is based lead to diagnoses subsequently found to be wrong, the system will be unacceptable in the most basic sense. The same applies to NPs influencing the search processes if these lead to incorrect outcomes, and to SPs, for example where the patient's systematic but undetected misconceptions lead to incomprehensible or disagreeable outcomes.

However acceptability in relation to the patient independent of the agent is not only a matter of the patient's global view of the system: it could affect his reaction to individual sessions, and indeed behaviour during sessions, as considered in more detail later. If the patient does not understand why the system is seeking some information e.g. a benefits system asking about apparently irrelevant 'private' matters, or believe its outputs, he may terminate the session. (Of course the patient in some cases may not be in a position to do this.) Alternatively, the patient may seek to define his problem status, for example as a user of a tax estimation system, in such a way as to inhibit 'unfavourable' outcomes, in a manner involving all three types of property. As these illustrations suggest, the patient's SPs may be particularly important in relation to system acceptability, and as much to what may be called local acceptability as to global acceptability. This is especially clear in relation to the user's view of explanations and justifications, which are most important in making the system's decisions acceptable to the user (notably in providing a basis for actions on or by the patient), and which primarily appeal to the patient's mental judgement. Of course in some cases, for example in that of the very sick person, the scope for reflective response in the patient may be small, but there may still be the possibility of, for instance, the system's explanation influencing agreement to therapy. The important point, however, is that whether the system's behaviour is acceptable to the patient may not be only a matter of the patient's response to the system, but of the system's response to the patient, i.e.

be a function of the patient's previous show of SPs.

The system's acceptability to the agent is also a matter of interest, to which both his NPs and SPs are relevant. In general the same factors apply as for the patient, but there may perhaps be more emphasis on the system's behaviour during a session, and less emphasis on the eventual outcome, if the agent is independent of the patient: though as the case of medical systems shows, an agent doctor will wish to feel confidence in the system's decisions. In relation to local acceptability, disregarding the agent's sex, for example, could produce annoyance, while a car repairman's belief that all Swedish cars are well made could be taken into account in packaging a long fault list for him.

The different user properties and system functions, and the contributions the former can make to the latter, are summarised for convenience below (with indirect rather than direct contributions in parentheses):

model functions

	patient	agent
effectiveness		
DP	x	
NP	(x)	(x)
SP	(x)	(x)
MP	(x)	(x)
efficiency		
DP	x	
NP	x	x
SP	x	x
MP	x	x
acceptability		
DP	x	
NP	x	x
SP	x	x
MP	x	x

I have ground through the relation between user properties and system features in some detail because it is necessary, as soon as one starts considering user models, to bear in mind the implications first, of the fact that agent and patient may not be the same person; second, that different kinds of property of either may be involved; and third, that distinctions between both persons and property types can have different consequences for different aspects of expert system operation. Thus even where only one person is involved he may have different roles, and hence be associated with different models geared to his distinct system functions in these roles (also, in the limit, where two people are involved, apparently behaving symbiotically, it could be that conflicting properties are observed which require the construction of distinct models).

For the purposes of discussion I have treated these three system functions as distinct in relation to user properties. But of course the performance of one function can affect that of others: there is an obvious relation between efficiency and acceptability, for example, so models aimed directly at one function may indirectly support another. In the same way, the different types of user property can combine in their effects in relation to one another.

It is also worth noticing that while the assumption being made is that there is an autonomous expert system whose purposes user models may serve in various ways, the function of a system may simply be to construct a user model for the sake of the knowledge it embodies, without any intention of exploiting that knowledge for ulterior system purposes: Hayes and Rosner's (1976) ULLY design was for a a system whose whole joyous purpose was modelling people observed at parties.

Individuality in models

Though it is implicit in what I have said so far, I should emphasise that my concern is with individual user models, and indeed with individual models in two senses. Thus we must allow in principle that each user (and hence each agent and each patient respectively) can differ in himself from every other, though it is quite possible that the descriptive resources we have available are in fact only rich enough to characterise classes of user, e.g. young, female. (Note, however, that the descriptive resources may be richer than they appear to be where e.g. a limited set of decision properties is extended by the use of degrees of possession, and classes have degrees of membership.) Certainly the assumption is that we are not dealing with a standard type of person, or even with a few broadly defined classes, membership in which might be established in a simple and direct way. This applies, in particular, to agents: it may be that the decision classes of the system are few, so the set of possible patient models is limited, though of course in a serious expert system the effort required to make the appropriate decision class allocation may be non-trivial. However even with respect to patients, where the set of decision classes is larger, it is reasonable, even though this set, and consequently the corresponding set of patient models, is known to the system (though it is implicit in the rule set rather than listed explicitly), to think of the process of reaching a decision as one of constructing an individual user model. Certainly the assumption being made here is that with a non-trivial set of agent properties, there will be enough possible agent classes for it to be more sensible to think in terms of constructing an individual agent model ad hoc, rather than of assigning agents to predefined classes.

However user models may also be individual in a stronger sense, namely that users are defined in terms of the history of their interaction with the system. Thus in a natural language dialogue, a model of a participant will be a time-related object which involves the order in which inputs are presented as well as the inputs themselves. Whether or not the global purposes of specific expert systems would be served by considering individual models in this sense is not the point here: it is considered in more detail later after a fuller discussion of these

'dynamic' models, and of issues like the possibility of obtaining dynamic models without natural language interfaces. For the present purpose what must be emphasised is that the idea of user modelling is deemed to involve treating users as individuals (even if they are not so in fact), both in terms of their property descriptions (of any type) and their interaction histories.

In this connection a distinction between surface and underlying models is important. In systems like Gershman's (1981) Yellow Pages assistant or Wilensky's (1984) Unix Consultant underlying standard users are assumed. Thus while particular activities towards which a 'session' is directed are quite specific, their interpretation is in terms of standard motivations and objectives for which, correspondingly, standard plans, say, are appropriate. Similarly Clancey (1979) and Hasling et al. (1984), and Genesereth (1979), treat user behaviour in an instruction context as an overlay on, or deviation from, a standard way of thinking or doing. In all these cases we have, as it were, particular instantiations of prior norms. This idea is carried further, with richer results, in the use of 'stereotypes' in HAM-ANS (Hoeppner et al. 1983) and in GRUNDY (Rich 1979): here negotiating systems model users as variations on, extensions of, or constructs out of, stereotypical users. In these systems (as Wahlster 1984 notes), user models, in the form of prototypes, are supplied at the beginning of sessions, as opposed to being built from scratch. But they are not given for direct use as decision class models: they are a means for assigning users to decision classes.

Illustrative model uses

As the discussion so far has treated points bearing on user modelling somewhat independently, the examples have provided only partial and indirect illustrations of the possible role of user models in expert systems. In what follows I shall consider some exemplary (hypothetical) expert systems as wholes in some detail, to get a more coherent overall view of what sorts of things we want user models for, i.e. of why they would be useful and so of what they would be like: of whom they would be models, what they would contain, and where they would be applied. Following the analysis of the example systems, I shall consider whether we can hope to get such models, especially if we cannot rely on natural language interaction with the user as a means of acquiring information about him.

As examples I shall take a social security benefits system, used respectively by a benefits office clerk and by the applicant for benefits, as Systems 1A and 1B; and a medical diagnosis system, used respectively by an experienced doctor and a student, as systems 2A and 2B. Since the emphasis is on the uses to which modelling information could be put, in relation to system functions, I shall disregard the fine details of how the relevant user information is obtained, whether it is stated, elicited, or inferred, and also of when it is obtained, i.e. I shall not consider either communication means or interaction history. Thus this set of examples is concerned with models referring to user properties, and not to interaction histories. I shall also not consider the internal mechanisms the system must have for creating and manipulating models.

The examples are deliberately designed to cover a wide range of modelling possibilities and possible model uses: whether, from a more detached point of view, these model uses would actually be worthwhile is a separate question, and one which is considered after the discussion of whether such models could be constructed. It should also be emphasised that the various properties have been selected to serve as minimally adequate illustrations of the points to be made: the fact that they do not necessarily constitute convincing sets and, especially, that the DP sets are far too limited to illustrate significant expert system decision making, is irrelevant.

System 1A

Here we have a distinct patient and agent. We will imagine that the patient is an elderly female pensioner in a wheelchair, who lives in Sunderland, is a Catholic, has poor sight, and is painfully honest but deeply suspicious of officialdom. The agent clerk is an experienced man in his thirties who thinks the national benefits legislation unduly favours women. As far as the expert system decision classes are concerned, the age, sex and disability of the pensioner are DPs, but her place of residence, religion, poor sight, honesty and attitude to officials are not.

What user models could we build from this information, for what purposes?

Considering the patient first, we necessarily require the basic user model, i.e. the patient model determining the decision class. Call this P1. It contains the facts that the applicant is 89, female, and wheelchair-bound, and drives the system's central search process to establish the specific benefit entitlement. This is the model tied to system effectiveness.

However P1 can also be used in relation to system efficiency; for example disability may be used in preference to other DPs to control the search of the rule space (this is a rather trivial and forced illustration in this simple example), and the acquisition of further DP information (e.g. is the use of the wheelchair permanent or temporary?).

P1 can further be exploited for acceptability in informing the patient (and also indeed the agent) of the part played by the particular DPs in the system's conclusions (the standard first level of explanation), for example indicating that a benefit of Xpounds is made up of Ypounds for disability and Zpounds for age. Another example would be taking the patient's age into account in presenting or explaining the system's decision in plain, jargon-free language and not over-elaborate form.

What could be gained from exploiting other properties of the patient (assuming they are known to or discoverable by the system)? The patient's NPs, being a Catholic and having poor sight, could form a second model, call it P2, bearing on the system's efficiency. For example, even if the patient is not in direct contact with the system (this is the agent's function), the fact that she is

poorly sighted might prompt a data-checking process designed to clarify misunderstandings or ignorance attributable to inability or reluctance to read benefits instruction leaflets. (Note that I assume here that the agent does not 'interfere' in such matters, i.e. he is a transmitting rather than interpreting intermediary. This is of course too simple, and if he is enough of an interpreter, the situation becomes extremely complicated in relation to modelling the patient in any other aspect than the DP one: as suggested earlier, if the agent is interfering enough, what sort of model of the other aspects of the patient could the system construct? It is worth noting the large confusion potential here from the point of view of user modelling, though I shall not pursue it further.)

P2 can also contribute to acceptability: for example the system's decision may be conveyed to the local Catholic priest for confirmatory or reassuring presentation to the patient.

Finally, the SPs of the patient, if identified, could be used for a further model, P3, to be exploited for efficiency purposes, e.g. if the patient is known to be honest the system could spend less time on doing consistency checks on its data (these would have to be somewhat richer than those hypothesised here); at the same time, the fact that the patient is suspicious might lead the system to seek that information about her that she is least likely, through objections to prying, to object to giving, for example her disability rather than her age. The most obvious utilisation of P3, however, would be in relation to acceptability: for example the explanation of the system's output might be made especially full in an attempt to overcome her suspicions about being done down.

We now consider the potential for agent models in System 1A, treating the fact that the agent is an official clerk, is experienced, male and in his thirties as NPs, and his opinion about the benefits legislation's bias as an SP (as for the patient, the boundary between NPs and SPs is a bit arbitrary, but is maintained as distinguishing mental states or dispositions as SPs). A model of the agent's NPs, A2, could be used in the interests of efficiency, for example to seek information about the patient expressed in terms of official system notions ("disabled" versus "can't walk much"), to reduce the chances of the system's misinterpreting patient data. (As this illustration shows, the agent's properties can be approached only from the point of view of the system-patient relation: the agent may not be of interest in his own right.) A2 could also be exploited in relation to acceptability in, for example, providing motivation for questions the system asks, or justification for answers given, by reference to established benefits legislation concepts presumed known to an experienced clerk, or through pointers to further detail in backup literature. (Here the system is more directly concerned with the agent, but maintaining acceptability to agents can be seen simply as a long-run concern with system patients, at least in a system like the benefits one.)

Finally, a model of the agent's SPs, indicating his view of the bias of the system, could be used in the interests of system efficiency to couch information requests in a neutral, or perhaps deliberately counterbalancing form, and in the interests of system acceptability, in this case to the agent, by making it plain in

explanation that the benefits for elderly disabled women are no greater than those for men. (This use of the model might also be deemed to be of indirect value to the patient.)

But as these illustrations suggest, the role of models in the case where agent and patient differ may be more complex than has been allowed so far. The assumption we have so far made in considering System 1A is that the models of the patient and agent are exploited in the system's internal operations only in relation to the patient and agent respectively. But of course if the agent is the means of transmission of information about the patient, the possibilities of misrepresentation (for whatever reason) have to be taken into account. Thus the system may have to exploit its models of the agent not merely when addressing the agent in his own right, but when addressing the patient. For example the experience and age of the agent may be relied on as grounds for the accurate transmission of the description of the patient, while his sex bias may have to be positively counteracted by seeking checking information. The fact that agent and patient are distinct, but that there may be all kinds of interactions between them, suggest extremely complicated situations for the application of models (as well as, as noted, for their determination).

System 1B

The modelling situation here is necessarily simpler than that of System 1A, since we are dealing only with models of one person, the elderly woman beneficiary. We thus have three models, P1, P2=A2, and P3=A3. The models as patient would be used in the same general way as they were previously, but the fact that the patient is also an agent with quite different properties from those of the agent in System 1A, would naturally lead to significant differences in the details of the system-agent interaction. For example, in relation both to system efficiency and to system acceptability, the fact that the agent has poor sight could lead to minimising the amount of text produced by the system when seeking or providing information. Again, the fact that the agent is not experienced in benefits legislation concepts could mean that system outputs were not couched in legislative jargon. The fact that the patient and agent are the same person may of course also affect the system's behaviour in more subtle ways; for example it may not be enough to respond to the agent's properties regardless of the fact that they are also the patient's properties: the fact that they are the applicant's properties may influence the style of the interaction. For instance, the fact that it is the agent as much as the patient who is suspicious of offialdom could lead to a different formulation of requests for patient information from that adopted when either one but not the other was suspicious.

System 2A

In this case we imagine we have as patient a black boy of five suffering from a high fever and a yellow rash on the chest, coming from an orphanage and terrified of doctors; as agent we have a female doctor of fifty, who is a specialist in

children's illnesses and is most concerned to obtain a diagnosis for a very sick child with unfamiliar symptoms. We assume that the child's age, sex, colour and fever and rash are DPs; his coming from an orphanage is an NP, and his terror is an SP.

The patient model P1 here will have functions essentially like those of P1 in Systems 1A and 1B, though there is a less obvious role for P1 in relation to system acceptability to the patient: P1's contribution beyond effectiveness would seem to be limited to efficiency, for example the yellow rash might be a highly heuristic property. There is less obvious utility in System 2A for the patient models P2 and P3 than there was in 1A or B. The P2 model covers coming from an orphanage: this could be medically relevant in principle, but as the system does not allow for this, either explicitly or, for example, by treating it as an 'entry' property leading to or correlated with others, e.g. high exposure to risk, it cannot be exploited in the interests of efficiency in any very obvious way. Given that in an expert system of the type in question, and in its specific application to a child, the patient's role is necessarily somewhat passive and mediated by others, it is not clear that there is much for a P2 model to do. (However that a P2 could have a role even in such circumstances is illustrated by considering being black as an NP rather than DP: this could easily be a search guiding property.) The P3 model may be somewhat more important in this case, as possibly affecting the provision of accurate descriptions of the patient's sensations, e.g. having a headache. P3 could therefore be exploited to produce a carefully designed series of simple questions suited in both content and form to elicit descriptions or acknowledgements of his sensations from a terrified child. (As in the earlier example, we simply assume here that the system has some way of establishing that the child is terrified, either directly or via the doctor.)

The role of the agent models in System 2A would appear to be more important than that of the patient models (other, naturally, than P1). Thus the various NPs of the doctor (i.e. all but her concern to obtain a diagnosis) would be used for model A2 which could be exploited to promote efficiency, e.g. by relying on her experience and specialist knowledge to interpret particularly tricky information-seeking questions, for instance "Does the patient exhibit unusual lassitude (for a child)?" A2 could also be used to enhance system acceptability (to the agent) by, for example, stressing uncommon features of the diagnosis, treating this as a useful addition to this agent's existing knowledge. The doctor's concern for the child, the SP constituting model A3, on the other hand could be used in the interests of efficiency because it could be exploited to allow the recommendation of tests requiring unusual dedication of the doctor's part; it also be used to emphasise the need for caution to offset any temptation to act too quickly on a very tentative diagnosis. A3 could be used in the interests of acceptability, say in emphasising that the diagnosed disease is not serious though the symptoms are unusual.

System 2B

We have the same patient here as in System 2A, but as agent a first-year medical student who has never encountered a sick child but who is extremely self-confident.

The patient models here are of course the same as for 2A. The NP agent model A2 could aid efficiency guiding the formulation of system queries both in content and expression terms, for example by not assuming great familiarity with children's illnesses (or indeed any illnesses); A2 could also be used in the interests of acceptability, for instance in relation to explanations by making these rather full and by supplementing them with suggestions for further reading. The SP model A3 could be used in the interests of efficiency to check on the danger of misrepresentation through over-confidence by administering suitably orthogonal checking questions; and it could be exploited in the interests of acceptability to disguise the provision of what is in fact very basic information (this illustration raises nice points about social engineering).

These examples suggest a whole range of detailed possibilities for the application of a range of user models, i.e. they show that there is considerable scope, in improving expert systems, for user modelling to take account both of the complex properties of system users and the complex internal and external functions of the systems themselves. To summarise the position from the system point of view, the examples show that user models have functions beyond the primary one of providing representations of patients who are the subjects of the system's decision-making process, as follows: they enhance the system's efficiency (not necessarily measured in terms of reduced work) in guiding the search process, in validating input data, and in ensuring well-organised output; and they enhance system acceptability by influencing the manner in which input data are sought and output information provided. (Note that though the discussion has focussed on the system's 'intuitions' in seeking information relevant to decision making and providing information about its decisions, the points made also apply to the system's response to information volunteered or questions posed by the user.) However this somewhat simple description masks the rich variety and complexity of the ways in which a system has to manipulate specific models in its detailed operations, and, further, the ways in which models of different people, or different aspects of people (treated in the illustrations as distinct models, as they are logically), may all have a contribution to make to overall system performance. But the examples nevertheless illustrate the considerable range of possbilities to be allowed for in principle, even if in particular system contexts some individual models have no useful role, or some roles may not be required of specific models.

I have not, moreover, considered a whole further area of modelling yet.

Time-dependent modelling

The models we have examined have been essentially static models. That is to say, they have been presented as referring to permanent features of the users which are independent of the behaviour of the expert system, and are consistent over the session. I have also assumed that the user models can be treated as

characterisable in terms of simple property lists: thus possible substantive as opposed to cooccurrence relations between them have to be known in advance to the system.

From one point of view, even the fact that some of these properties are of major importance to the system, as the basis of its decisions, and that they may not all be known at the beginning of the session, does not affect the basic picture: while the system develops its model P1 by gradually acquiring information about the DPs of the patient, the state of the patient himself has remained the same throughout.

But from another point of view, behaving as I did in presenting the examples, as if the system's acquisition of information about user properties is independent of time is not just unrealistic: it is more fundamentally unsound. User modelling itself must be allowed to be time dependent in the most straightforward sense represented by the accumulation of information about the user within an interactive session (or across several sessions where this is natural). We should therefore allow for the possibility that any of the specific kinds of user model presented will be changing models from the system's point of view, treated as changes of user state. These changes will, moreover, be context-dependent, i.e. dependent on the interaction between the user and the system, in the basic sense that a piece of information may be supplied by a user in response to a system question at a particular point in the interaction.

The model changes here are a consequence of the fact that the system does not acquire all its information about germane user properties all at once. We have also to allow for expert systems where property descriptions are time related for external rather than internal reasons, i.e. where the changes are really in the user's states and not in the models, and indeed where the system is geared to take account of the sequence of states. This most clearly applies to patients, as in a system for monitoring a sick person.

From an abstract point of view, whether changes in models are due to changes in the system's knowledge of the user or to changes in the user himself may be immaterial. Moreover systems like monitoring ones may be assimilable to the static view in some cases. Thus when a monitoring system makes decisions reflecting the history of the patient's property states as well as his current state, we have a system which is not essentially different in kind from the type of system considered earlier, since the historical data available for decision making at a given time (which may of course include the past actions of the system as well as states of the patient), is simply a more complicated version of the static property description. Again, if a monitoring system takes successive decisions in (short) real time, based on new patient states, its operation could be treated as concatenation of sessions. It could then be handled within the general framework we have assumed, namely that user interaction with an expert system occupies a 'session', aimed at reaching a single decision: changes in the property descriptions are then either effectively new sessions, or simply attempts to represent the external situation well enough to reach a decision for the single real external problem.

However as these remarks suggest, the static view, while appropriate in some cases, may be artificial or inadequate in others. Thus it may be at least as sensible to treat the continuous monitoring decision case as an essentially time-related adaptive system: this would seem to be particularly appropriate where longer term information covering changes is preserved and exploited to give a trend basis for individual decisions. Equally it might be realistic to take into account the fact that even where properties may in principle be regarded as static, they can change. For example an SP like a global view of benefits legislation might change under the impact of particular information about the legislation disclosed during an interaction, with possible repercussions on the interaction.

But there is a more important sense in which user modelling can be time dependent than that we have just considered, and in some specific contexts, for example teaching systems, and in research on natural language dialogue, models of an essentially different kind than those discussed so far have been investigated. These are models of the user's goals, plans and beliefs in the specific context of the system's operations, i.e. models of the user's system-directed and system-motivated behaviour relative to a session. It is clear that in many applications and operations, expert systems can only perform satisfactorily if they can recognise their user's goals, plans, beliefs etc (Genesereth 1979, Jackson and Lefrere 1984, for example). Models based on these properties are in principle models of a moving rather than static user, or dynamic models. Where the type of model just considered through a medical illustration might be called a changing model, and indeed even a reactive model if the patient's state changes in response to system decisions as well as for autonomous reasons, we are concerned here with changing models that can be claimed to be of a qualitatively different kind because the properties of the user, and specifically his subjective or mental properties, cannot be regarded as permanent, but are a product of the interaction with the system, and alter during the session. At the fine level of detail, such mental properties are local to stages or segments of the interaction. Their existence and qualitative characteristics are as much a result of the history of the session as of the user's and system's longer term attributes.

Of course even in these systems at a particular point in a dialogue the situation may be treated as static, for example if the user has asked a question and the system is seeking to establish the aims and beliefs he had at the time (cf McKeown 1984). But this is an unhelpful or even misleading perspective since it emphasises discontinuity rather than continuity in the system's operations. In the interaction between the user and the system over a session, say of instruction or advice seeking, one can require a continually changing model of the user as his goals, plans and beliefs are changed by the system's behaviour: but while the model changes, the important point is that it is a model of the same one user. (The fact that individual mental properties may not in practice change is immaterial: this has to be treated as the exception rather than the rule.)

As the difficulties of capturing users' goals and so on may be very great because they may be conveyed indirectly rather than in literal utterance meaning (Sidner and Israel 1981), necessitate distinctions between communicative goals

and domain goals (Perrault and Allen 1980, Allen et al. 1982), or be inferred only by deploying extensive knowledge (Pollack et al. 1982), actual systems so far may model this type of user property in a simple and non-evolutionary way (Gershman 1981, Wilensky 1984). But this does not mean that the need for more adequate dynamic models should not be recognised.

A further point about these models is that they may not be appropriately, or perhaps cannot be adequately, characterised by simple lists of property values: they have a structure with e.g. causal relations between elements (this may of course also be true of the 'simple' changing model). Thus there will inevitably be complex relations between the user's goals, beliefs, etc. So the manipulation of the model becomes a much more complicated affair than that of the kind of model considered earlier: though relations between properties have in fact to be allowed in the static and simpler time-dependent cases as well. At the same time, it is more likely that the elements of dynamic models are at least largely, if not wholly, those the system itself uses in its decision making or can be systematically related to this. For example if the system is a student course advisor, the elements of the user model may be courses, prerequisites, dates, etc, just as they are the elements of the system's decision-making apparatus: the user may believe that Course A is given at Time 1 where the system knows it is given at Time 2, but course and time, if not all specific courses and times, are common concepts.

The user's beliefs may thus be wrong, and they may indeed involve general concepts not known to the system itself, for example that there is such a thing as an examples class. But it is not unreasonable to assume that modelling could be effectively supported by the kind of natural structural extension to the core information represented by an ISA hierarchy or by McKeown's 'points of view', i.e. unknown user concepts could be assimilated. This is a quite different matter for the expert system designer from thinking about whether it would be useful in a benefits system, for instance, to know about people's political beliefs.

I shall refer to the properties involved in these dynamic models as 'mental' properties, MPs. Of course mental properties in themselves are the same sorts of things as SPs: they are beliefs, goals, etc; and indeed a user's particular background SP, say his belief about the scope of some objective domain being represented by an expert system, may be identical with an individual MP about the scope of the expert system during the user's initial interaction with the expert system. Indeed a global belief about the scope or function of an expert system could be regarded as the transitional case between 'permanent' SPs and 'temporary' MPs. I shall nevertheless treat SPs and MPs as distinct because they are involved in different types of model, as far as their presuppositions are concerned, with the implication for MPs of some more direct connection with system activity, and (possibly) a more explicit and detailed representation.

The dynamic models just introduced may be superimposed on the static models discussed earlier (simple changing models would more probably just replace them, according to their property types). However it would be dangerous to assume that the static models should be treated as contexts for the dynamic

models, other than in a purely formal sense, since there might be no material connection, let alone a discernible connection, between e.g. the user's NPs and his beliefs about and plans for interaction with the system. That is, even if there is a real connection between properties of different types, and especially, say, between the user's SPs and his plans, we cannot assume that the system will have the knowledge needed to establish this, or the means to exploit such relationships. Thus it is safer to think of the system as operating not with connected or connectible models, but with parallel or conjoint ones. There would however be no reason to associate a distinct dynamic model with each of the patient and agent models, though it is of course proper to allow, where patient and agent are different, for distinct dynamic models for patient and agent respectively (I shall call these DPM and DAM). From the present point of view it is perhaps natural to conflate the sets of static patient and agent models respectively into single patient and agent models (SPM and SAM) each involving several different types of property, with accompanying dynamic patient and agent models. These dynamic models might have a hierarchical character, for example indicating plans and subplans, though it may be necessary to allow for essentially concurrent models for different interests or motives. A hierarchically structured model could incorporate the representation of temporal change: it is in any case essential to recognise that the dynamic models accompanying the static ones change over time.

The motivation for having dynamic models is the same as that for static models. Particular emphasis has been placed on them in the teaching system context for discovering the conceptualisation underlying the user's actions, e.g. solution of equations (see, for example, Genesereth 1979, London and Clancey 1983, Sleeman and Brown 1982). Dynamic models are obviously appropriate in the query and advice context too, where they may be used, for instance, to discover the reasons for the user's query, with a view to providing him with the answer the 'real' implicit query rather than the explicit one, with appropriate additional information, or with a tailored explanation (see Shrager and Finin 1982), Carberry 1983, and Pollock 1984, for example). These ways of exploiting dynamic user models are included in those listed earlier for static models, and using dynamic models for further purposes, say in the interests of accurate data collection, are obvious additional possibilities. However, as with the other type of model, the scope for dynamic modelling is naturally variable depending on the particular characteristics of the individual application system, and also, to some extent, on the type of system it is.

Dynamic models have been mainly investigated in types of system where the user necessarily or naturally has a very active role, as in an instruction system, and where patient and agent are identical (and also in cases e.g. hardware design, where there is no human patient). In an important sense the system functions only in relation to the user's wants, as in the advice case, or in relation to his individual behaviour, as in the teaching case, i.e. where the user is not, or cannot be seen as, a mere passive transmitter of independent information, or where the system is necessarily an interactive or even negotiating one. Thus in the HAM-ANS hotel reservation system (Hoeppner et al 1983), the concept of neutral decision classes hardly applies: the user's assignment to a class (room

booking) is the outcome of mutual negotiation and adjustment. In such a case the user's MPs become more like DPs: indeed it is fair to say the user's MPs are the system's DPs. For instance in a teaching system the student may be assigned to a class not simply on the basis of his actual learning behaviour, but on that of his inferred state of knowledge and belief. Certainly treating such systems as essentially the same as one determining what disease a person has, however jusifiable on abstract grounds, is to dilute their distinctive flavour. Modelling in this strongly interactive type of system has been especially associated with the use of natural language but, as the teaching case shows, does not depend on natural language, especially where there are 'visible' user actions constituting an alternative source of information about him. Again, although dynamic models have been associated with agents, i.e. with applications either where agent and patient are identical or where there is no human patient, it is possible in principle to have dynamic models for both agent and patient.

Thus although dynamic models in such applications as teaching are so manifestly important they have dominated and even subsumed models of the other kinds considered so far, they can clearly figure in other kinds of system and need to be analysed in their own right. Again, as advice systems show, dynamic modelling exploiting the user's MPs can accompany other models involving other properties, for example in a travel advice system models using such static DPs or NPs as the age or current physical location of the seeker of advice. As this example suggests, the logical distinction between types of property and their related models are important and need to be recognised. It may be that all the property types, for patient or agent respectively, are put into one big pot for a nominally single model, but this may be a mere matter of technology not implying any real integration of the modelling operations. Real integration, as just suggested, is hard. Moreover attempting integration presupposes a clear understanding of the nature and role of the different kinds of property and of the modelling behaviour appropriate to each. Thus the logical distinctions I am concerned with here, and have been emphasising somewhat artificially by my terminology, and to be recognised.

Illustrative dynamic models

For more specific illustration, if we consider the addition of dynamic patient and agent models to the example systems mentioned earlier, and their possible roles in contributing to system effectiveness, efficiency and acceptability, we can see that these dynamic models have such uses as the following.

In System 1A, the social security system with clerk, there is in fact not much obvious role for a dynamic patient model (DPM), because there is a separate agent: the utility of such a model (indeed also its attainability) depends rather on whether the agent is merely a transmitter, and possibly reformulator, of the patient's individual inputs, or is responsible for the whole organisation, including temporal ordering, of the input in interaction with the system. Even in the first case, the presumption is that inputs would be 'rationalised' for presentation to the system, and hence would be sensibly motivated. However if we assume the

agent is not in complete control over the interaction, we can see that a DPM, for example reflecting the user's successive inputs about age, and queries about age-related benefits, could be used by the system to postulate a patient belief that age is the basis of benefit (rather than disability). The DPM could be exploited, if not obviously for effectiveness, for efficiency in controlling the soliciting of other information, say about possible disability, and for acceptability in explaining that the basis of benefit calculation is disability rather than age.

A separate dynamic agent model, DAM, in this system could be used, for example, in the interests of efficiency to check the agent's beliefs about the relative importance of the patient's DPs, hypothesised from the discourse context; and in the interests of acceptability, exploiting these presumed beliefs, to motivate explanations accompanying system questions. (The agent orientation could be further reflected for the latter by the use of technical language in the interaction.)

In System 1B, the applicant-driven benefits system, the scope for a rich dynamic model, in fact the single one DPM=DAM, would appear to be greater than in System 1A, as without the assistance of even modest mediation through the clerk agent, the system would need to build up a picture of the patient's view of the benefits resources available. It would certainly be highly desirable, if not strictly necessary. Thus the patient's input of age information could be complemented, in the interests of efficiency, by rather wider-ranging further information seeking, to gather all the relevant data, with continual modification of the model of the patient's beliefs in the process. Each single question asked could be profitably supported by an appropriately worded justification geared to the discourse context (as well as to the patient's static properties). This is a classic dynamic modelling situation.

In System 2A, the medical system with experienced doctor, there appears to be virtually no scope for a DPM. There may also be, somewhat surprisingly, little scope for a DAM, even if we assume that the doctor has a pronounced 'patient interpreter' role, analogous to an interpretive rather than merely transmissive benefits clerk. Thus once we have established the doctor's NP of experience, we might simply allow her inputs to drive the system in a straightforward way. However we might also, if we do make the assumption that it is the agent's choice of input datum, or input question, rather than the patient's, allow there to be a role for a DAM representing the system's picture of the doctor's hypotheses about the patient's disease, which the system could exploit, autonomously, for efficiency, in the style described earlier. Note, however, that I have also assumed that the doctor is an experienced user of the specific system as well as a medical expert: if she was not there would be scope for a DAM primarily directed at assisting interaction and providing information about the system, i.e. a DAM aimed at acceptability even more than efficiency. (System 2A also illustrates the possibility, where effectiveness and efficiency are not wholly separable, of using either static or dynamic agent models to support effectiveness; for example the doctor's supposed hypotheses about the patient's disease might be taken as evidence for the system's diagnosis.)

Finally, for System 2B, the medical system with student doctor, there is as little scope for a DPM as in System 2A, and similar scope for a DAM, though the specific DAM would of course be quite different, being as it were the complement of the 2A DAM: for example the model of the agent's hypotheses about the disease could be continuously evaluated, to support explicit responses indicating errors and providing corrective information.

Clearly, as these examples emphasise, the construction and manipulation of the various models, SPM, SAM, DPM and DAM, both independently and in relation to one another, constitutes a major enterprise: all kinds of detailed relations in individual contexts are possible. For example, given the conjunction of some DP with some MP, how should the system respond? These are particularly interesting questions where different models (i.e. model types) might suggest distinct and even clashing strategies, as for example, in the benefits system, 1A, concentrating on the agent's NP, experience, might suggest a brisk approach to information seeking based on a presumption of his helpfulness towards applicants, where his SP, a belief in welfare bias towards women, might suggest a carefully elaborate approach to information seeking designed to counteract possible effects of his bias. Note that such de facto, operational relations between different types of property, and especially between properties of the two major model types, static and dynamic, present problems for the use of models whether or not any intrinsic, e.g. causal, relationship is deemed to hold between them. Any real relationships that can be established between properties, even of the same type, merely add to the complexity (setting aside the special case of DPs). We have in fact little idea of how to build and manage individual models of any sophistication, in relation to underlying expert systems of any complexity, or indeed of what is needed to form and use models of particular types. This remark applies even to user models confined to DPs: that is where work on expert systems starts and where much more research is required (despite Hayes-Roth et al 1983). But taking these models for granted, going beyond very modest additions of one other model type to them, and especially of dynamic models other than ones with severe restrictions as to plausible desires of the user, is already a challenge.

Before the attempt is made to meet this challenge, therefore, it is useful, indeed necessary, to consider how far one could expect to get with user modelling in the absence of a reasonably powerful natural language interface, taking it for granted that natural language interaction is at least especially helpful for user modelling. Indeed we should now, more properly, consider access to modelling information in a neutral way, without any assumptions about the relative value of different types of interface, let alone absolute requirements for any one type.

Sources of modelling information

The question to be addressed here is: what sources of information about users are available to the system as a basis for model construction. A further important question is how far we can expect sources that might be informative in principle to be so in practice. Clearly, if for expert systems of various types, or

application areas of various types, we have no reason to suppose that we could obtain enough information to build a user model, other than the primary one using DPs, it does not matter how helpful in principle a user model could be. Indeed while the system clearly has to build a DP model to work at all, the quality of the user interface can not merely make gathering the necessary information more or less convenient, but mean that the information itself is more or less adequate. Thus while there are familiar strategies for obtaining decision information, e.g. prompting questions or the more subtle 'observation' techniques of teaching systems, the way in which this information is gathered can interact with the performance of other modelling operations, and information even for the primary model can be obtained in different ways, i.e. from different types of source associated with different forms of interface. Equally, given the potential value of other types of model, we have to consider sources of information for these, for which there are also alternatives.

We therefore have a quite general question: what sources of modelling information are there, with a subsidiary more specific question: are there specific kinds of source appropriate to specific kinds of model? For example, it might be that different types of source, implying different interface modes, were required for different modelling purposes within the framework of a single system. At the same time it is necessary to emφphasise again the fact implied by the earlier examples, that even where one interface type, say a natural language one, is optimal, system circumstances may make for poor information sources, where there is no reason to suppose that this is correlated with a lack of need for, or utility in, modelling.

The possible sources of information for expert systems in general are, first, non-linguistic actions, such as drawing a picture, or providing an equation to solve (or, alternatively, solving a system-posed one). Non-linguistic actions are not necessarily confined to systems without human patients e.g. the picture may refer to a patient, but are more commonly associated with such systems, for example design systems. The second type of source is linguistic actions, statements, questions, commands etc in natural language (I am not treating "linguistic action" as synonymous with "speech act"). Both of these types of source can in principle provide inputs not solely in response to system prompts, though natural language, if not mandatory, is typically deemed more convenient for volunteering.

Some apparent intermediate cases can be assigned to non-linguistic or linguistic actions, since both cover more and less restricted means of communication. Thus if the user's contribution to the system has to be expressed within a limited subset of natural language, or alternatively is made by simple, perhaps only YES/NO responses to menus or system questions, we still have an essentially (natural) language-based mode of operation, and can assign these types of source to the class of linguistic actions. The characteristic feature of such systems is that the system provides a linguistic context for the user, and though the user's own linguistic expressions are in practice normally quite limited, they can be interpreted by the system in relation to the richer context it supplies. Whether or not the system itself is generating or interpreting linguistic

expressions in any real sense is not the issue here. From the present point of view, emphasising the user's perceptions, it is fair to describe such menu/prompt interfaces as natural language ones, though they are 'sub-language' rather than full language interfaces, and also system-driven ones, especially as far as specific interactions are concerned, but to a material extent for the session as a whole (e.g. user menu choices apply only within a given menu tree framework).

On the other hand, if the user is confined to an essentially artificial language, like a formal language for circuit design, it may be more appropriate to assign this case to non-linguistic actions, along with e.g. equations submitted to a computational algebra system. This is still appropriate even if an artificial interaction language is dressed up in natural language, as long as it is perceived to be artificial (as a programming language is). Thus 'pseudo' natural language interfaces can be assigned to the non-linguistic class.

These points are more fully discussed in Sparck Jones (1985). As the 'intermediate' examples just discussed suggest, the concept 'language' sensu natural language is so vague that the dichotomy between non-linguistic and linguistic actions cannot be fully sustained, i.e. may depend on a number of intrinsically soft notions like 'expressive resources', 'likeness to our ordinary language' etc. However I shall maintain a broad, informal distinction between linguistic, i.e. using natural language at least non-trivially, and non-linguistic, as this is useful in considering the possibilities for obtaining information about the user.

The form of the source is clearly one major influence on information inputs to modelling. Whether modelling information is typically supplied directly or indirectly is another relevant factor. We can naturally expect to obtain information for the primary (especially static) patient model P1 from what may be called 'driving' inputs, i.e. ones directly supplying decision property values. Whether such driving inputs are sought by the system or volunteered by the user is not material as long as the inputs bear directly on the system's decision-making processes. However, as this suggests, the important feature of inputs relative to decision properties is not whether the user is correct in his beliefs about the relevance of his inputs, but whether the system extracts information for P1 directly from the input. There is, further, no reason in principle why information relevant to other models of the patient or agent should not be indirectly extracted from driving inputs, i.e. such inputs would be indirect sources of information for other models. Equally, non-driving inputs may be used either directly, in relation to 'matching' or 'correlated' models, or indirectly, providing information for other types (it is worth noticing that information about DPs can in principle be derived from non-DP oriented inputs). As with the previous distinction between linguistic and non-linguistic sources, the distinction between direct and indirect sources is not absolute but relative. I am nevertheless emphasising it here because it brings out an important point about obtaining modelling information, namely whether it is reasonable to assume that the user is willing or able, for whatever reasons, to supply all the information needed explicitly.

Comparing non-linguistic and linguistic interfaces, a natural language interface more obviously allows for the possibility of user inputs which are not intended or taken as directly driving the system's primary decision-making process, in the way that the equation to be solved drives an algebra expert system. With a natural language interface, moreover, the expressive resources available also allow for the possibility that information about other user properties than DPs can be derived from the form in which driving inputs are presented. However none of this implies that nothing can be learnt about users (other than relative to P1) from driving inputs submitted through non-linguistic interfaces (see Genesereth 1979), or that in systems without natural language interfaces the user can do nothing but provide driving inputs (they could, for instance, try out the graphic 'alphabet' available in an architectural design system).

As Genesereth, for example, shows, it is perfectly possible to learn something about the user's MPs for dynamic modelling purposes from non-linguistic direct, i.e. driving, inputs, for instance what the user's beliefs about algebra are. It would also be possible in principle, especially over a long information gathering period, to learn something about the user's static properties, for instance in the algebra case whether or not the user remembers what he learns. But it is far from clear how, for example, it could be discovered that the user was a socialist from his input algebra equations (and this could be pedagogically relevant). Similar considerations apply to pictorial inputs (photographs, drawings etc). As noted, the assumption about natural language in general is that as it is more expressive, it correspondingly provides more opportunity for conveying information (intentionally or unintentionally: thus a DP might be discovered from an input not intended to drive the system), both in the 'packaging' of first level, or direct, inputs and in second level, or indirect, ones. For example, the agent's input "The applicant states she is without independent means" could suggest that the agent thinks she in fact has such means, and an agent's request "Tell me about widow's pensions" may suggest that the patient has the property of being a widow.

However as Richard Young points out (personal communication), it is important to recognise the user modelling limitations of some nominally linguistic interfaces, for example using air traffic control-ese, and equally the modelling potentialities of modern interactive terminals offering multi-process, multi-window, i.e. highly flexible, communication resources, albeit non-linguistic or only sub-linguistic ones.

However, whichever type of action is involved, the important issue for model building is property recognition, i.e. knowing that something is a property, as a necessary underpinning of property marking, i.e. knowing that the property holds of the user. Thus while it is not suggested that the marking process is easy, it depends (outside the realm of more ambitious machine learning, aimed at learning wholly new concepts) on the system's familiarity with the property concepts involved. For example, establishing that an agent user is experienced or inexperienced requires that the system 'understands' (at least operationally) the properties of experience and lack of experience.

The major problem, for patient models other than the primary one, or for agent models, is recognising model properties. This applies to static properties, and also to dynamic model properties unless a very narrow view is taken of what constitutes relevant goals, beliefs, etc. Recognition is where the differences between kinds of model are important. The position for DPs is relatively straightforward. The patient's possible DPs are, by definition, known to the system, though recognising them in all the guises and manners in which they come is another matter. Even so, as the system has to know, to function at all, what its sets of decision properties and their possible values are, this gives a lot of leverage when user inputs are not wholly transparent. Equally, the starting assumption about the MPs underlying dynamic models is that even if they are not simply taken as DPs, they are strongly DP-based or linked, e.g. that they are plans relating to DPs (or decision classes), or at least that they are categorisable or organised in terms of higher order concepts or relations known to the system: for example in an algebra system the concepts of precedence and arithmetic operator will be know to the system, allowing identification of a wrong user belief about the specific precedence relations between particular operators. There may of course be more extraction difficulties in the dynamic case, because there may be more complex properties and structural relations between them. In the general case, moreover, we have to allow for user MPs, properties, whether of a relatively elementary or more elaborately structured kind, not known to the system.

The more difficult types of property are NPs and SPs (other than those manifested in or subsumed under a dynamic model). There are potentially a great many properties of the user that could be relevant to the system's efficiency or acceptability, so the system builder is faced with the problem of providing a substantial world description allowing for adequate property recognition (and subsequent deployment). But the mind boggles, for example, at providing an account of religious sects to underpin a medical system (though checking for Christian Scientist parents might be a very useful thing to do in examining a young child), or a comprehensive account of people's views about local and central government sufficient to pick up such general but relevant beliefs as "They always try to do us down" (though beliefs in systematic hostility by <u>them</u> might justify inflated claims by <u>us</u> along the lines "They keep saying we're not really poor, so I'm going to look as poor as I can"). These examples might seem farfetched, but the problem of relevant user NPs and SPs is a major one for any serious attempt to support expert systems with models rich enough to be really effective. Clearly, the only plausible strategy for the expert system builder is to work outwards from the DPs, adding properties obviously related to these and justifiable as improving system performance in relation to observed use. However this is not a simple matter in practice, because the knowledge involved and its structure may not be easily determined, even if a suitable formalism is available for expressing it, which it may not be, and the observational or experimental effort required to establish its value may be far from trivial.

<u>Obtaining modelling information</u>

We now come to the problem of establishing what properties the user has (and especially what properties other than DPs). Clearly, within the expressive resources of whatever means of communication are supplied, he can volunteer a property description, or it can be explicitly elicited from him. However, as noted earlier, some communication media e.g. graphics or artificial languages, are unlikely to be adequate either in principle or in the limited exemplars to be expected in practice, to convey some static property types, e.g. SPs, or the MPs underlying more complex or refined dynamic models, for instance anxiety or growing disillusion with the bases for determining welfare benefits, though a liking for play, or gross paranoia, might be established even in these restricted circumstances. Moreover, as has been frequently observed, there is no reason to suppose that the user is going to volunteer everything that is useful, or will put up with it being extracted by a grinding interrogation. (The same applies, to a more limited extent, to the discovery of DPs.)

Thus it has to be accepted that user properties, other than DPs, cannot generally be obtained directly, but rather have to be obtained indirectly, as a 'byproduct' of the user's 'normal' inputs, i.e. those tied to the system's function. There is no difference in principle between different property types here: the attention paid to dynamic modelling reflects the fact that, given the general presumption that the user is adopting a positive attitude to the system and that he has some idea of what its all for, a great deal of leverage in discovering the user's MPs can be got by starting from the system's own information and functions; and the justification for discovering the user's MPs is particularly strong for those types of system where they bear very heavily on the system's decisions, and hence in turn can have marked effects on acceptability. The system's starting point is far less obvious when it is seeking static model information hidden in the user's inputs. The importance here of the form of user action available is simply that it is far from clear how easily or how much static model information could be derived from non-linguistic inputs: for example hesitant drawing might show that the circuit designer was a novice, but it would not necessarily convey that she was female.

The overall conclusion, therefore, is that there is no reason in principle why the various different sorts of user model could not be derived from either sort of user action (though distinguishing agent from patient in the non-linguistic case might be especially difficult); so there is no reason in either case why the various types of model should not all be sought. However as far as the realities go, it seems clear that there is a much higher chance of extracting information, whether quantitatively or qualitatively, from the richer resources of natural language input. Indeed it is widely held (and not only by the language processing community) that, as modelling has to be based on indirectly obtained information, to avoid oppressing the user with long interrogations, the best method of working is through full natural language interaction, with all the expressive power this allows.

But this is a general conclusion that has to be shown to be correct, and it may in any case not apply in individual cases: the detailed possibilities for obtaining modelling information are of course determined by the particular application

system, and specifically its function (with dependent predictable user community), as well as by its communication media. Both to illustrate the various possibilities in more detail, therefore, and, perhaps, to provide support for the claim for natural language, I shall now consider how the user property information for the example systems could be obtained (though exactly how, and with what precise descriptive system resources I shall not consider), using different means of communication.

Illustrative sourcing

As before, I shall focus on externals not internals, in this case on what can be extracted about the user and not on what the consequent model looks like, or how it is used, either to extract further information or for the other functions listed. In this illustration I shall not consider all of Systems 1A - 2B in full detail, but rather examine alternative forms of system input for one system in some detail, and comment more briefly on the comparable opportunities for the other example systems.

For the main illustration I shall use System 2A, the medical system with the child patient and experienced doctor. I earlier assigned menu interaction to the category of linguistic interfaces; however though in broad terms menu operation is not distinguishable from free natural language, I shall see what could be learnt from the use of these two types of linguistic interface, to illustrate the range of possibilities with linguistic actions. I shall also, to provide the necessary non-linguistic interface, consider an alternative graphic interface for 2A. This is somewhat implausible in practice, but is justifiable for discussion purposes (it also serves, indirectly, to make the point about some styles of interface being more appropriate to some system circumstances (tasks or application domains) than others).

I shall consider the full natural language interface possibilities for System 2A first, assuming that the user is allowed considerable initiative (and also, of course, that the system has the requisite language interpretation capability).

Starting with static models, for the patient's DPs, the assumption is that these would be directly input via the agent as descriptive statements, either volunteered or sought by system query. The fact that the agent is experienced suggests that extracting DP information from the inputs should be relatively straightforward, since it will either be provided in appropriate language or can be sought, by system query, in a manner likely to evoke an apposite, i.e. immediate and knowledgeable, response. Thus the patient's DPs can be expected to arrive in direct style. However with respect to the patient's NPs and SPs, it is not clear how they could be identified, unless they were either volunteered, in recognisable terms, or explicitly sought (with all the underlying assumptions that the system knows it should be interested in such things). It is possible that the NP, coming from an orphanage, could be hypothesised from DP-related inputs, for example that the patient's parents are dead, or that the child is at high risk of contacting disease (i.e. higher than a school?). Again, it is possible that the

fact that the patient is terrified could be inferred from DP-related inputs to the effect that he is conscious but the state of his aches and pains is unknown (but clearly terror is only one potential explanation for this lack of knowledge). But it must be accepted that the chances of the patient's specific NPs and SPs being indirectly conveyed through DP-related inputs, or via any inputs intended for other purposes, is very low.

With respect to the doctor agent: her NP of experience, assuming it is not checked explicitly, might be inferred if she volunteered a comprehensive patient description in proper technical language, or responded rapidly and incisively to system questions about the patient. It is not clear, on the other hand, how the system could establish the agent's SP, concern for the child, without an explicit statement to this effect: repeated pressing of the 'return' key for a quicker system response could just as well be interpreted as due to busyness, or general impatience.

With respect to dynamic models, as noted earlier, there is little scope for a DPM in 2A. A DAM representing the agent's views, including her developing or changing views, of the patient could most obviously be established from direct suggestion ("I believe he may have Snodgrass' disease"), but could also be straightforwardly inferred from a question ("Is yellow rash indeed a symptom of Snodgrass' disease?"); or it could be inferred in less direct fashion from, for example, the input of an extremely detailed rash description accompanied by only minimal information about other symptoms (but again, this is only one possible inference from such a pattern of inputs). A changing DAM could be established (though with difficulty in this relatively simple example) if an initial detailed description of the patient's rash was later followed, without system direction, by a detailed description of the child's fever, or a question about rash was followed by a question about fever.

Of course I have not indicated the detailed mechanisms by which this information could be extracted: I am making a simplistic assumption that the system could learn what the human auditor could learn from ordinary language inputs. I shall make a similar assumption for the alternative communication media.

Thus if we now consider communication via menus for System 2A, the obvious point is that there is no means of volunteering information, so unless the system explicitly seeks the different types of property information, the chances of obtaining non-DP properties are much lower than they are with a full language interface. However, assuming a not completely trivial menu interface, there is still the possibility of inferring some data: for example menu responses could be the basis for inference about the patient's NPs and SPs, as they were previously. Similarly, it is possible to obtain some information about the agent. For example, her experience may be established not only by such obvious devices as her lack of use of a 'Help' facility (though this alone is ambiguous between medical and system experience), or ready supply of menu slot fillers, but from, e.g. the overall content of her responses, say that they have provided consistent data. Clearly, this would all be very tentative and would need verification. As

before, it would be very hard to establish the agent's SPs. It would also be much more difficult, because of the fact that the communication initiative lies with the system, to construct a DAM. It is just possible that one might be constructed, for example from noting that only one or very few menu slots were filled (out of a larger set of related slots) and inferring that the agent thought these were especially important.

Finally, we can imagine extracting information from a hypothetical pictorial input system (for which we make an unrealistic assumption about the system's capabilities for picture interpretation, as working with an iconic 'alphabet' would make things too linguistic). We assume that this is a system allowing user initiatives, rather than a system-driven one. Unfortunately, imagining how a pictorial communication version of 2A would work leads one into the wilder realms of the imagination. Even hypothesising that the system has the image interpretation capabilities that we have does not remove the difficulties of conveying, even with the intention of conveying directly, the agent's particular SP, for instance: single pictures would appear to be insufficient, so a whole sequence could be required to convey such a concept as a doctor's anxiety for a patient. One does not see either, how direct images could also (without extension) convey non-DPs indirectly.

The mere idea of pictorial communication for such a system may seem ludicrous: but it is worth looking at the possibility simply to be clear about the implication, which is that obtaining user modelling information from non-linguistic actions is likely to be feasible only where the use of a non-linguistic interface is natural for the underlying expert system, as it is in design cases, and that a non-linguistic interface for its own sake is in general likely to be of limited utility.

In general terms, the situation for the other example systems is as for System 2A. Thus with a full natural language interface in System 1A, one can again see that the patient's DPs would be relatively easily obtained, given an experienced agent. There is the same difficulty of identifying putative NPs and SPs if they were not volunteered in recognisable form or explicitly sought. The patient's poor sight NP might be hypothesised from the fact that an agent is being used to work the system (assuming this itself has been established), while her SP of honesty might be inferred from the supply of income date with the fine details of pence as well as pounds. On the other hand, the difficulty of establishing NPs and SPs is well illustrated by the fact that although the patient is suspicious of officialdom she is approaching the system through a clerk. The agent's NP of experience might be inferred as in System 2A, though it is not obvious how the clerk's SP of a belief that the benefits legislation is biassed towards women could be established, except possibly through the form of his request for an explanation of the system's decision. As noted earlier, there is little scope for a dynamic patient model in 1A; a dynamic agent model, reflecting the clerk's view of the relevance of certain benefits to the applicant, could be constructed from the sequence of his inputs. the status of a more restricted menu interface, or of a non-linguistic graphic interface, would be much as for System 2A.

In spite of these similarities, however, it is worth noticing that very considerable challenges of a rather different type also arise in the cases of Systems 1A and 1B, because of the particular form the patient-agent relationships take. In systems like 2A and 2B the distinction between patient and agent would appear not to present material problems because the system's view of the patient is very 'object-oriented', so to speak, and is based primarily on properties not regarded as relevant to any potential agent, for example having a rash. Thus the possibilities of confusing patient and agent appear small, making the task of extracting information about them that much easier. However it is not difficult to see that with a different patient, for example an adult of the same sex as the agent, the ease with which agent and patient could be distinguished in relation to their respective NPs and SPs could be significantly reduced.

Even so, with Systems 1A and 1B, the system's difficulty in separating information about the two people (for 1A) or roles (for 1B) may be much greater than with 2A and 2B, unless simplifying assumptions are made, which effectively means restricting the system's modelling potential. This is simply because disjoint property descriptions, particularly for the case where there are genuinely two different people, are less less likely to be a natural consequence of the system's area of decision. Thus, for example, unless the person reference is made explicit, either by a direct description in volunteered input, or by implicit connection with the system's indicated reference when the system is prompting, it is not clear how the system can distinguish the patient from the agent for inputs stemming from user initiatives. It would seem that the system can only operate on the general assumption that input content refers to the patient, because the system is primarily geared to taking decisions about patients, even if this limits the system's scope for agent modelling. Alternatively, it could be assumed that input style is attributable only to the agent, but this too would be restrictive. (Modelling on this basis would have to be very conservative, because of the high chance of error.) It would appear that the distinctions between patient and agent required to motivate separate models of a more ambitious kind could only be made in the long term, via the system's perception of coherence in property descriptions, and hence separability of descriptions. Thus while obtaining information about a property may be as easy or as difficult in these systems as in 1A, establishing that it is a patient as opposed to agent property, when the people are different, is a major problem. Nor is it reasonable to assume that where, as in System 1B, there is only one person involved, there is no need to separate the patient and agent roles.

An integrated example

To pull the previous examples bearing on individual points together, I shall conclude with a more extended one, illustrating the way in which, over a period of interaction, an expert system could acquire modelling information and utilise it. For this purpose I shall take System 1A version of the benefits expert system, as having distinct patient and agent, with both 'active'. To show the most favourable modelling situation, and at the same time, even its difficulties, I shall

imagine free natural language as the means of communication, but, in the interests of some degree of realism about the state of the computational art, make the slightly artificial assumption that the system can interpret its linguistic inputs sufficiently without extensive knowledge of the world far outside its own decision scope. In the imaginary interaction below, the supposed user inputs and system outputs are labelled I and O respectively (the former being input by the clerk agent), with comments on the notional modelling behaviour of the system indicated by c. P1, P2 and P3 are the models for the patient's DPs, NPs and SPs, and A2 and A3 the models for the agent's NPs and SPs; DPM and DAM are the dynamic patient and agent models. The symbol "+" indicates the addition of information to a model, or its reinforcement if already present; "?" indicates the models are hypothetical, not certain. For the example, some properties additional to those originally given are introduced. Note that I am not concerned here about the plausibility of the assumed benefits legislation and its system implementation. (We assume we have already gone through the initial logon sequence.)

O: Please give me details of the applicant.

I: The applicant is female, aged 89, widowed, grade 3 disabled, and I imagine should get a pretty lavish benefit.

c: P1 + female, 89, widow, disabled 3
P2 + nil
P3 + nil
DPM + nil
A2 + expert?
(use of "grade 3" jargon)
A3 + hostile to patient?
(use of word "lavish")
DAM + nil

O: What are the applicant's sources of income?

c: response motivated simply by need for further DP information, primarily to promote effectiveness

I: She says she's getting widow's pension of 17pounds and 3pence and otherwise the priest sometimes gives her 2pounds out of his poor box. Her clothes seem rather worn.

c: P1 + pension 17.03, charity occasional 2.00
P2 + catholic?
(reference to "priest")
P3 + honest?
(pence details; mention of odd gift)
P3 + ignorant?
(does not realise housing status relevant)
P3 alternative + dishonest?
(concealing housing status)
P3 + suspicious (of clerk)?
(only replying to direct questions; note not necessarily an alternative)

DPM + no obvious beliefs, expectations about system?
A2 + nil
A3 + hostile to patient?
(reinforced by use of "she says", "seems")
A3 alternative + cautious observer?
(not experienced in evaluating state of old women's clothes)
DAM + exploring influence of modifications on basic income?
(apparent interest in clothing allowance)

O: Find out whether she buys clothes frequently. But has she been receiving anything for her disability? And ask her whether she lives with anybody.

c: response motivated by need to elaborate on P1, for effectiveness and possible efficiency; also to check on P3s of honesty and ignorance, attempting to select from alternative hypotheses, for the same reasons; by use of homely language attempting to bypass hypothesised suspicion, with additional motive of enhancing acceptability to patient; further motivation to gather information sufficient to start a useful DPM. With respect to agent response motivated by attempt to distinguish alternative A3s and sidestep hostility, primarily for effectiveness but also for acceptability; also using DAM suggesting focus on ancillary income to redirect attention to main sources of income

I: She says the hospital gave her a wheelchair and that she lives by herself but her daughter sometimes visits and the District Nurse attends at basic frequency. She finds it very difficult to afford clothes.

c: P1 + no disability allowance, home owner?
P2 + nil
P3 + honest?
P3 + ignorant?
(apparently does not know of disability allowance, clothing allowance)
DPM + no notion of system and its capabilities?
A2 + experience?
("basic frequency" jargon)
A3 + nil
DAM + nil

O: Ask her whether she owns her home, and if she is still paying anyone for it (check for mortgage freedom). Tell her we may be able to help even if she has a house.

c: response pushing for more P1 information, for effectiveness and, we will suppose, efficiency and acceptability; also reacting to and testing P3s and DPM, promoting acceptability and perhaps changing the patient attitudes embodied in the DPM. The system is exploiting the agent's experience in A2 (reference to "mortgage freedom") and trying to reduce his hostility and render the system more acceptable by exhibiting care and exhaustivity in its operations

.....

The illustrative inferences are not the only ones that could be drawn; the responses too are not the only possible ones; and I have not commented in detail on the way possibly conflicting inferences and models (e.g. re P3) are balanced. But the example is sufficient to show the way in which individual models could be manipulated both singly and jointly to improve performance in its different aspects. Clearly, within the scope of this short dialogue, it is not possible to demonstrate that a crass, non-modelling approach could not work as well, or at least well enough; however I maintain the kind of model-supported behaviour shown would be very hard to emulate.

Conclusion

I have not said anything in this paper about the specific mechanisms with which the system forms, develops, and applies user models. I have been primarily concerned with the roles of user models in expert systems, and with whether user - system interaction could in principle supply the information required for modelling. I have simply blandly assumed that if a human being could create and utilise a model in a given expert system context, the computer could do it too.

My main aim has been to illustrate the complexity of modelling that stems not from the manipulation of a single model for one system purpose (though this can be complicated enough), but from the multiplicity of models, with different bases and different functions, of which an expert system could take advantage. Indeed this complexity extends beyond that of models based on one kind of property rather than another, and beyond that of static versus dynamic types of model. Even for a single model type, based on a single kind of property, for example a DAM, different individual DAMs are possible in a given system, according to the specific purposes for which the model is used, and hence the choice of individual properties on which it is based. For example, in the discussion of System 2A, one dynamic model function could be to speed the gathering of patient information, and another to improve the agent's professional knowledge: models for these different purposes might not merely make different use of the same user information, for example, input order, but could be essentially based on different subsets of the user information, for example clarity of descriptions and fullness of descriptions respectively, and naturally also on different ways of manipulating this and extending it by inference. Of course, as with the static models, whether one calls it one model with different aspects or different models is to some extent a matter of taste. But for system design it is useful to distinguish the different functional models that could be derived even from the same specific property set, let alone from distinct subsets, especially but not necessarily of different kinds, drawn from a larger set of properties. Having recognised the range of functional possibilities, we may then choose to put everything back together as one big multi-facetted, multi-purpose model. But it is still useful, in my view, to maintain separate patient and agent models, if not for all expert systems, for many individual systems, and not only those where patient and agent are distinct people.

Thus summarising the modelling possibilities discussed in the paper, we have something like the following:

```
    (--DPs--SPM1 )     exploiting any combination      )
    (--NPs--SPM2 )     at any time for system actions  )
P-  (--SPs--SPM3 )-    for affectiveness, efficiency,  ) -
    (--MPs--DPM )      and acceptability               )      )
                                                              )- ditto
    (--NPs--SAM2 )                                     )      )
A-  (--SPs--SAM3 )-    ditto                           ) -
    (--MPs--DAM )                                      )
```

(where each model may in fact be a set of models,
because of uncertain data or distinct purposes)

This diagram clearly indicates some of the complexity of user modelling, above the implemented process level, but even so it completely conceals major problems. It says nothing about the way in which individual properties and individual models are related, and in particular what causal relationships (whether tightly or loosely defined) hold between them. It says nothing about the problem of how the modelling information available is manipulated, especially in inference procedures, both in exploiting models for the the system's intrinsic purposes and for its extrinsic 'public relations' purposes: there is a difference for the system between ensuring that its decision information is correct and telling the user that he is wrong; in particular, it says nothing about model bootstrapping: using the model to extract further information with which to develop the model. And the diagram says nothing about balancing prediction with uncertainty in its manipulation of its models, and about the whole business of testing models.

The potential practical complexity of user models, suggested particularly by the example Systems 1A and 1B, is well brought out by asking who the system is using the models it has for. The simple answer is that the system uses whatever models it has for the benefit of the patient. But this is too simple. One can use a model, in fact the same model, and any model, for the benefit of the agent as well as the patient; for example to improve the agent's understanding of the system. It is specifically necessary, moreover, to distinguish, in model application, the beneficiary from the addressee. Thus in a system like System 1A, the system's interactive outputs will be addressed directly to the agent, but may or may not be addressed indirectly to the patient; and any agent model involved may be exploited for the (supposed) benefit of the patient, i.e. may be viewed as constructed in the interests of assisting the patient (for example to maximise the amount of money he gets), or for the benefit of the agent, i.e. may be constructed in the interests of the agent (for example to improve his understanding of the social security legislation). Indeed an agent model constructed in the interests of the patient may be applied to the interests of the agent. When the agent is deemed informed about and helpful towards the patient there may be no problem here, but it is easy to see how complicated things could get if the system formed the hypothesis that the agent is hostile to

the patient, for example the benefits clerk thinks the applicant is a dishonest layabout. Another manifestation of the complexity of modelling even in relation to individuals is suggested by what may be called the user's persona: are the user's attitudes his own, for example, or those of the organisation for which he works (Wilks and Bien 1983)? It is not difficult to apply this distinction to the benefits system clerk or to the doctor, if we interpret professions as organisations.

All of this discussion, too, has assumed a single patient and a single agent: we have to allow for the dauntingly more complex model manipulation and system apparatus required for an expert system involving multiple agents and patients, for example multiple agents in an industrial plant control context, or multiple agents and patients in a hospital (say in dealing with an epidemic) or business management system. Even in the type of example considered in the illustrations, things would be much more complex in real life because we would in general expect to have to manipulate a lot more user properties.

That these complexities are indeed the reality emphasised by Belkin's (1984) analysis of information retrieval. A librarian, helping someone to find the documents he needs, has to model his client from several different points of view and to exploit these models to serve a whole range of functions in interacting with the client to establish his need. As readers are ultimately interested in the contents of documents, not their descriptions, but both are critically language-dependent, librarians in real life are far more sophisticated expert systems than any we have now. The set of models, listed in Belkin, required for an information system as a whole, i.e. those of its readers, and even writers, as well as its librarians, show how far modelling can extend beyond the exemplary cases considered in this paper.

The second general point brought out by the examples is the variable scope for modelling in different expert systems. Even if we take it as given that any expert system involving a human being anywhere can benefit from user modelling, the benefits will clearly be larger in some cases than others (This point concerns models other than the primary patient model P1.) It is easy to see that in a system like a teaching system there is large scope for user modelling, in this case by progressing from a static to a dynamic model; in, say, a circuit design system, one could gain a lot from user modelling, though one could also get a long way without it, which is much more doubtful in a teaching system. In other cases, while the idea of a cooperative interface, in the deep sense dependent on modelling rather than in the superficial sense, is very attractive, it does not follow that users will be put off by its absence. If the value of, and need for, a system is sufficiently great, the user will be very persistent. It might not even matter very much that he made mistaken inferences through the system's lack of a model which could be used to prevent misleading system responses, in the style of Joshi et al (1984). Thus it seems that user models are of most importance, and, especially, of great importance, where the user's SPs and MPs make a real difference to the system's validity, i.e. where the user's SPs and MPs make a real difference to the system's own operations, because it is important that the system should be able to adapt its decision behaviour to respond to bias of any kind in the user's inputs. From this point of view, it appears that dynamic models may

in general be more desirable, because more valuable, than static ones. Certainly the human advice-giving contexts studied by Pollack et al. (1982) and Belkin (1984), illustrating strongly interactive and negotiating behaviour on the part of the user, imply the need for such models in constructing computer-based expert systems.

The final general point to be emphasised is the manifest difficulty of obtaining much modelling information that could in principle be of utility in system operations. This applies not only to static models but also, as the language-processing literature makes clear, to dynamic models in the most favourable case with full natural language interaction. How much modelling information we can realistically expect to be able to get? Another way of approaching this is to ask how useful very simple models would be, given the difficulty of obtaining much information (especially about properties other than DPs). In particular, how useful would simple static models involving NPs or SPs be? As the examples discussed show, they could be of some use, but not apparently of great value. Equally, it is possible that in some cases, any dynamic models would have to be simple, as a straightforward consequence of the underlying expert system's purpose, and so might well not be worth the trouble of construction and maintenance. On the other hand, the fact that dynamic models are system geared and time spanning improves the chance of getting relevant information, compared with that for static models. But in general, the ambiguity of simple models may be so great, i.e. what the set of properties perceived could imply is so uncertain, that exploiting such models may be more dangerous, because the system may respond inappropriately, than helpful. That is to say, given that the primary purpose of modelling is prediction, we have to face the fact that the predictive value of very simple models may be extremely low, because they are based on undiscriminating property sets (or values), so it is possible that trying to exploit them may do more harm than good.

(The fact that the predictive potential of simple models may be very limited serves also as a reminder of the need for proper use of the expression "user model". It should not be trivialised by being used to mean little more than "user description", for something with minimal application in the system's operation: for example to effect a choice between standard 'experienced' and 'inexperienced' user interaction modes. It is certainly not very helpful, however formally true it is, to say that an expert system has a user model when all it has is the same built in interface model for all users. Equally, applying the phrase "user model" only to the DP-based patient model, in the static case, may be misleading.)

All this suggests that other than in specialised direct action cases like teaching or design, though modelling may be possible in principle and even in practice without a linguistic interface, it may be difficult to do any very useful modelling in such circumstances. Further, while modelling may be possible with a restricted linguistic interface, it is in general unlikely that any powerful modelling can be done unless free natural language interaction is allowed. This has considerable implications for expert system design as a whole. I have already noted (and see also Sparck Jones 1985) that providing a natural language interface in itself, without any reference to user modelling, is not merely a matter

of providing the obviously necessary (but far from trivial) language-oriented interpretive resources, but is likely to require increased capabilities in the underlying expert system: the convenience of the interface leads the user to provide or seek information in new ways, or to provide or seek new information, including new types of information. The system's capabilities have to be extended to match the greater (if unconscious) expectations of the user, implying a closer coupling of front and back ends than might initially be expected. In fact, as the comparison with the database case shows, even with very modest linguistic interfaces, more is required than just the addition of specifically language-processing capabilities like syntactic parsing: an enhanced, or differently expressed, characterisation of the backend system is needed to support the lexical, semantic and pragmatic aspects of input text interpretation.

The attempt to provide for user modelling in an expert system leads in itself to major pressures on the system, not only in such obvious cases as teaching systems, but in general, because the modelling is as intimately bound to the system's core decision-making activity as it is to its public relations side. Adding both user modelling and natural language interaction capabilities to expert systems makes their design a whole new ball game. There is nevertheless a good strategy with which to approach this game: start not with the user, but with the system. That is, don't start with all the properties an expert system user could have, with all the ways in which he may be modelled, with all the uses to which these models may be put. Start with the system's function, and ask what modelling needs to be done to make that specific system work better, in terms of its effectiveness, its efficiency, and its acceptability: what does the system really need to know? Then one can think about how to get it.

Acknowledgement

I am very grateful to Steve Goudge for discussions stimulating this paper, and to Richard Young of the Medical Research Council's Applied Psychology Unit for his most helpful criticisms and comments.

References

Allen, J. et al. (1982) ARGOT: the Rochester dialogue system. AAAI-82, 66-70.

Allen, J. (1983) Recognising intentions from natural language utterances. In Computational models of discourse (ed Brady and Berwick), Cambridge, MA: MIT Press.

Belkin, N.J. (1984) Cognitive models and information transfer. Social Science Information Studies, 4, 111-129.

Boguraev, B.K. (1985) User modelling in cooperative natural language front ends. In Social action and artificial intelligence (ed Gilbert and Heath), London: Gower Press.

Carberry, S. (1983) Tracking user goals in an information-seeking environment.

AAAI-83, 59-63.

Carbonell, J.G. et al. (1983) The XCALIBUR project: a natural language interface to expert systems. IJCAI-83, 653-656.

Clancey, W.J. (1979) Dialogue management for rule-based tutorials. IJCAI-79, 155-161.

Davies, N.G. et al. (1985) TUTOR - a prototype ICAI system. In Research and development in expert systems (ed Bramer), Cambridge: Cambridge University Press.

Genesereth, M. (1979) The role of plans in automated consultation. IJCAI-79, 311-319.

Gershman, A. (1981) Finding out what the user wants - steps toward an automatic Yellow Pages assistant. IJCAI-81, 423-425.

Hasling, D.W. et al. (1984) Strategic explanations for a diagnostic consultation system. International Journal of Man-Machine Studies, 20, 3-19.

Hayes, P.J. and Rosner, M.A. (1976) ULLY: a program for handling conversations. AISB Conference Proceedings.

Hayes-Roth, F. et al. (eds) (1983) Building expert systems, Reading, MA: Addison-Wesley.

Hoeppner, W. et al. (1983) Beyond domain independence: experience with the development of a German language access system to highly diverse background systems. IJCAI-83, 588-594.

Jackson, P. and Lefrere, P. (1984) On the application of rule-based techniques to the design of advice-giving systems. International Journal of Man-Machine Studies, 20, 63-86.

Joshi, A. et al. (1984) Preventing false inferences. COLING-84, 134-138.

Kukich, K. (1984) Presentation at the Generation Workshop, Stanford, July 1984.

London, B. and Clancey, W.J. (1982) Plan recognition strategies in student modelling: prediction and description. AAAI-82, 335-338.

McKeown, K.R. (1984) Natural language for expert systems: comparison with database systems. COLING-84, 190-193.

Pollack, M.E. et al. (1982) User participation in the reasoning processes of expert systems. AAAI-82, 358-361.

Pollack, M.E. (1984) Good answers to bad questions: goal inference in expert advice-giving. Department of Computer and Information Science, University of Pennsylvania.

Rich, E. (1979) User modelling via stereotypes. Cognitive Science 3, 329-354.

Shrager, J. and Finin, T. (1982) An expert system that volunteers advice. AAAI-82, 339-340.

Sidner, C.L. and Israel, D.J. (1981) Recognising intended meaning and speakers'

plans. IJCAI-81, 203-208.

Sleeman, D. and Brown, J.S. (eds) (1982) Intelligent tutoring systems, New York: Academic Press.

Sparck Jones, K. (1985) Natural language interfaces for expert systems: an introductory note. In Research and development in expert systems (ed Bramer), Cambridge: Cambridge University Press.

Stefik, M. et al. (1983) Basic concepts for building expert systems. In Building expert systems (ed Hayes-Roth et al.), Reading, MA: Addison-Wesley.

Wahlster, W. (1984) User models in dialogue systems. Invited talk at COLING-84; to appear in Computational Linguistics.

Wilensky, R. (1984) Talking to UNIX in English: an overview of an on-line UNIX consultant. The AI Magazine, 5, 29-39.

Wilks, Y. and Bien, J. (1983) Beliefs, points of view, and multiple environments. Cognitive Science 7, 95-119.

Young, R.M. (1984) Human interface aspects of expert systems. MRC Applied Psychology Unit, Cambridge.