Automating the Librarian: Belief Revision as a base for System Action and Communication with the User

ALISON CAWSEY, JULIA GALLIERS, STEVEN REECE AND KAREN SPARCK JONES
Computer Laboratory, University of Cambridge, New Museums Site, Pembroke Street, Cambridge CB2 3QG, UK
ac, jrg, sr, ksj @ cl.cam.ac.uk

This paper describes a current research project investigating belief revision in intelligent systems by modelling the librarian in interaction with a literature-seeking user. The work is designed both to test a theory of agent behaviour based on belief revision proposed by Galliers, and to evaluate a model of the librarian developed by Belkin, Brooks and Daniels, through computational implementation. Agent communication is seen as motivated by and motivating belief changes, where belief revision is determined by coherence, combining endorsement, connectivity and conservatism. The librarian is viewed as a distributed expert system with many individual specialised functions operating in particular belief domains. The paper describes our first implementation of the belief revision mechanism and of a very primitive librarian, designed to test the basic viability of our ideas and to allow us to explore different forms of the distributed system architecture.

Received December 1991

1. INTRODUCTION

The work described in this paper is basic research aimed at a challenging and necessarily very long-term goal, automating the librarian. Searching online bibliographic databases to obtain literature references or documents for end users is an important component of a modern librarian or information officer’s work. It requires professional knowledge and skill, so providing convenient direct access to bibliographic services for end users instead calls for sophisticated interfaces able both to determine the user’s need and to express this in a way suited to searching the bibliographic file. In general, that is, it is necessary both to identify the user’s topic and to specify this in the indexing or classification language used to describe documents in the file. But even when the search language is the natural language of the file documents’ titles, abstracts or texts, professional knowledge and skill is required for effective searching.

Some first steps have already been taken in automating the intermediary by, for example, Pollitt,22 Vickeray et al.23 and Brajnik et al.,3 but what has been done so far has been very limited, especially in the system’s subject scope. More effective systems would call on artificial intelligence (AI) techniques for reasoning on knowledge in interacting with and acting for the user. The project described in this paper is therefore concerned on the one hand with appropriate general mechanisms for agents manipulating beliefs and conducting dialogue, and on the other with deploying these mechanisms within the framework supplied by the literature-searching task and by a model of the librarian’s characteristic knowledge and actions.

The project is intended to investigate key ideas about what is in principle involved in automating the intermediary. But given the complexity of the real librarian’s task it can only attempt an initial, very simplistic, laboratory system: the foundational character of the work which is needed for eventual proper systems means we cannot envisage realistic prototyping. The project is also in progress, so this paper describes the work’s starting point, plan, and what has been done so far, but cannot provide final results or evaluate these. It must further be emphasised that while the research is presented here as directed towards document retrieval, because automating the intermediary requires general capabilities as well as specialised knowledge and developing these capabilities is a concern of AI this work is also seeking to contribute to AI as a whole. Thus while from one point of view the aim is to apply AI ideas to information retrieval (IR), from another IR provides a valuable study context for modelling the way any agents adopt or change their beliefs about the world, particularly through engagement in dialogue. At the same time, while the work is concerned with helping the user to obtain literature, it is not primarily concerned with the type of display-oriented support system developed by e.g. McAlpine and Ingwersen,21 or with ‘plain’ public access as in Okapi.24

Section 2 of the paper describes the general theory of belief revision underlying the whole; Section 3 the view of the intermediary adopted as a project starting point; and Section 4 the current state of the research and intentions for the future. Cawsey et al. provides a fuller account than that given here.4

2. THE THEORY OF BELIEF REVISION

As noted, the aim is to provide the power needed for an automated intermediary by exploiting a general theory of belief revision as a mechanism motivating both the system’s external interaction with the user and its internal problem solving. Intelligent agents are continually revising their beliefs, and this applies to interaction between library users and librarians as much as to other dialogues. Interaction on literature seeking is not driven by fixed goals or manifested in a unidirectional flow of data from one party to the other. The system dialogue fragment of Fig. 1, of the kind recorded for actual sessions, clearly shows both parties revising their beliefs about what is wanted. Thus the librarian, having started by assuming that when people ask for books on plants they want books on growing plants, is obliged to revise this belief to accommodate a request for books on other aspects of plants. But equally, the user, having started

THE COMPUTER JOURNAL, VOL. 35, NO. 3, 1992 221
by saying he wants a book on cacti, revises this belief to accept that books on other sorts of plants may be appropriate. For this illustration we may envisage the eventual output of the mutual belief revision process as a submitted search request of the conventional sort for online services, in the form of a Boolean combination of terms in some controlled indexing language of the kind exemplified by MeSH. Both parties, that is, have collaborated to arrive at the actual search specification aimed at retrieving literature from the file to meet the user's real need.

2.1 Galliers' approach

The particular theory of belief revision the project intends to apply is that proposed by Galliers. This starts from the position that an intelligent agent is obliged, in a changing world of which any agent has only partial knowledge, to operate autonomously. An agent, that is, cannot rely on predictable states of the world, or on predictable behaviour by other agents within the world, and therefore has to do the best with the knowledge and powers it does have in setting its goals and in planning and acting to achieve these. An agent also seeks to behave rationally by maximising its own outcomes, so in a context of uncertainty this implies adaptation. In particular, a continuously changing environment stimulates changes of mental state in agents, i.e. since all knowledge is actually belief, changes in the environment stimulate the revision of beliefs. This revision depends on the agent's goals, but as the environment changes the goals can change too. Equally, having or adopting goals, which is a fundamental property of agents, implies action in and reaction to the world motivated by planning, and especially by strategic planning, to effect changes in other agents' mental states.

The nature of communicative behaviour in interactive dialogues between agents follows from the characteristics of and constraints on agent behaviour in general. Inputs from other agents suggest changes to beliefs, and an agent's own outputs are prompted by potential or actual changes relating to the agent's evolving goals. Thus an agent's contribution to a dialogue may be intended to check candidate changes of belief, that is to gather information to choose between competing beliefs, as well as to do what is normally thought of as simply collecting data or seeking to influence others, which are in fact also processes to be viewed as deploying beliefs, i.e. as revising an agent's beliefs in order to attain or determine goals.

Beliefs about other agents are clearly important in the interaction, but not just because they are part of the furniture of the world. As any agent has only limited powers to effect action, it needs cooperation to achieve its goals. This, however, in turn requires that it be cooperative. Thus dialogue is a process of negotiating and mutually accepting beliefs and hence intentions to act. Dialogue is a public manifestation of pervasive, goal-motivated belief revision in each participating agent, operating at every grain level in the characterisation of mental states. For example, in the dialogue of Fig. 1 the agents have beliefs (separate or shared) about the whole area of discourse, namely plants; about particular matters within this area, namely diseases or allergies; about needs to seek information; and about the means of seeking information.

As these examples imply, moreover, belief revision is also pervasive because it covers belief changes of all kinds, not just simple reversals but modifications of all sorts, including both changes in content, like more specialisation of beliefs, and changes in status, like less commitment to beliefs. Moreover, as beliefs are inferentially related, revision affects belief sets, not just single beliefs but whole webs of related beliefs. Thus a new belief may allow inferences affecting several other beliefs, and may mean there is more or less support for other beliefs. In general, a change to a single belief stimulated by interaction with the world or other agents affects the evidence supporting a whole network of related beliefs; equally, individual beliefs gain their value from the way they figure in a whole network of beliefs.

This very general picture is relatively uncontroversial: it has of course to be fleshed out in substantive enough detail for computational implementation. Thus, given some new input from a dialogue, how does an agent decide whether and how to revise its existing beliefs to accommodate the new information? More specifically, how does the agent choose what to reject when there is a conflict of any sort? Conflict resolution is not an occasional requirement having the form of a raw true/false opposition for any individual proposition: it is a normal requirement, having the form of a choice about alternative sets of beliefs representing different ways of responding to any change associated with new information. These responses, moreover, refer not only to beliefs embodying an agent's knowledge, but also to beliefs representing an agent's goals.

A proper theory of belief revision must involve three things: a way of characterising beliefs; a set of criteria for preferring some revisions to others; and a mechanism for applying these criteria to identify the preferred set of beliefs. Specifically, as beliefs form webs of related beliefs, what is needed is a means of handling the way individual beliefs contribute to the structure and solidity of a whole web, and of taking account of the propagation effect of changes at the level of individual beliefs, whether the change modifies an existing belief, adds a new one, or deletes an old one.

2.2 Details of the theory

The theory developed in Galliers is essentially pragmatic, and in the spirit of work by Gärdenfors, Doyle and Harman. It focuses on belief revision, characterising beliefs qualitatively rather than quantitatively, and regards all accepted beliefs as certain but variably corrigible, rather than as variably certain. Both of these features of the theory make it rather different from most approaches adopted in AI, where the choice of what revision to make is not normally addressed, and from approaches using quantitative certainty.

In the theory some beliefs are more persistent than
others: because they have more information value or explanatory power, the agent is less willing to abandon them. Ground assumptions, that is those beliefs which are self justified but which inferentially justify other beliefs, are particularly important here. These cannot themselves be justified: they are taken as a baseline. But they are endorsed with source information (cf. Cohen\(^8\)), and some endorsements are stronger than others. For example, beliefs embodying information received at first hand may be more strongly endorsed than those received at second hand. The theory has a number of types of endorsement which can be naturally, i.e. heuristically, ordered to provide a base for discriminating among ground assumptions. Agents will be more unwilling to give up more strongly endorsed assumptions. Moreover, though endorsement is not propagated directly to derived beliefs, since it is not obvious how derived endorsement values can be calculated from several different input values, it does provide an indirect means of discriminating among derived beliefs.

Fig. 2 shows the types of endorsement, and the rank ordering on them. This ordering refers to strength in a general sense: individual endorsement types may themselves embed more specialised notions of strength appropriate to their particular character, for example, strong or weak linguistic communication. Thus strong first-hand communication provides stronger endorsement than any second-hand communication can, but strong second-hand communication is stronger than weak first-hand communication. The latter in turn provides stronger endorsement than values (for example ethical or social ones), and values in turn rate higher than mere hypotheses. It must be emphasised that ‘communication’ is used in a very abstract way here to refer to what is deemed meaningful about the world by the receiving agent, so first-hand communication is limited to direct perception, and linguistic communication, necessarily depending on another agent and hence having some element of indirectness, is second hand. Communication in this paper will normally refer to the linguistic case, relevant to dialogue, so this is second hand in terms of the general theory, but may nevertheless within its own linguistic framework be deemed strong or weak by the receiving agent.

There is, however, more to the general idea of persistence, i.e. resistance to change, than endorsement. It is also necessary to consider the relations between beliefs, i.e. as beliefs are inferentially connected, the way in which connectivity reinforces beliefs. The more support a new belief offers to others, the more useful it is. Thus in evaluating alternative revisions of a set of beliefs, i.e. proposed alternative revised sets as responses to an input, it is necessary to consider how these improve the derivational, and hence explanatory, justification for beliefs as this is embodied in the connectivity among the beliefs in a set.

Endorsement and connectivity together determine the coherence of a set of beliefs: one set of beliefs is more coherent than another because of the way its constituent beliefs are connected and its assumptions are endorsed. Belief revision is thus a matter of evaluating alternative belief sets, constituting different responses to new data, to identify the most coherent set (or, possibly, sets). In general, more connectivity and stronger endorsement give more coherence, but connectivity is treated as more important than endorsement. Thus in considering alternative ways of revising beliefs, connectivity is examined first so alternatives with more connectivity are preferred, and endorsements are only investigated when connectivity alone does not unequivocally determine a single preferred set of beliefs.

This way of handling beliefs fits a conservative approach to revision which is intuitively plausible. This is to look at the justification for a belief only if it is challenged, and to abandon it only if the result is a more coherent web of beliefs. Thus one belief set embodying some particular belief of concern will be revised, i.e. will be replaced by another, only if there is good reason to do this. There is no need for an agent to evaluate and seek consistency among its beliefs unless this is required, and it is thus quite possible for an agent to have a mass of miscellaneous endorsed and variably consistent beliefs, in fact implicitly representing alternative internally consistent belief sets. But this generally conservative approach also takes a more specific form as a conservation rule, which is applied when there are alternative coherent revisions, to select the one(s) making least change to the previous state.

It must be emphasised that these notions presuppose some delimitation of the universe of beliefs, that is some context-driven means of bringing beliefs from the agent’s overall stock into the current focus of attention. There may thus be beliefs in the current set which are not accompanied by their supporting ground assumptions, i.e. assumptions are ground assumptions with respect to the current context, and as the context changes the assumption set may change. However, for simplicity we are assuming that an agent’s entire stock of beliefs is relevant to the current situation and thus to any ongoing dialogue. It is also the case that further concentration on where revision matters follows naturally from the idea of core beliefs. A pragmatic theory of beliefs accepts that some beliefs (relative to a context) will be held as core, for whatever reason; these may be ground assumptions or derived beliefs. Connectivity, endorsement and conservation are then considered primarily as to how they affect support for these core beliefs.

The way the theory works is illustrated by the highly simplified example of Fig. 3, presented in more detail in Galliers\(^8\) and in Cawsey \textit{et al.}\(^9\).
state 0: J's car bust again, at garage, being fetched:

several belief sets:
A) group for 'J to pay' core belief e.g.
   ...
   garage respectable
   fault mended
   bill to pay (df)
   ...  
   = > J to pay (C)
B) group for 'J not to pay' core belief e.g.
   ... failed mend earlier
   garage feel guilty
   not bill to pay (h)
   ...
   = > J not to pay (C)

2 preferred sets, both A group, equally plausible, differing in
connectivity, endorsement;
content distinction: multiple faults/fault hard to find

J: "How's my car?"
assistant: "It's OK, you can take it away. I don't think there's
anything to pay."

state 1

revising gives for B group e.g.
   ...
   not bill to pay (2cw) 
   = > J not to pay (C)
   ...
   BUT conservation prefers revisions in A group e.g.
   ... bill to pay (df) 
   = > J to pay (C)
   ...
5 alternative sets revising A equally plausible;
content distinction: multiple faults/fault hard to find

J: "He says there's isn't anything to pay."
proprietor: 'No, there's nothing to pay."

state 2

revising gives for B group e.g.
   ...
   not bill to pay (2cs) 
   = > J not to pay (C)
   ...
   revision in B group with stronger assumption now preferable to revision in A
(3 alternative sets for B:
content distinction: garage feel guilty/nothing wrong with car)

Figure 3. Belief revision in car repair example (C = core belief).

Implementing this theory of belief revision computationally requires a specific mechanism for constructing and evaluating all the belief sets which constitute alternative ways of dealing with some new input. The implementation also requires proper definitions of the evaluation criteria as these apply to and allow comparisons between whole sets of beliefs, that is, definitions of connectivity over a set of beliefs, of endorsement for a whole set, and of conservation in a set. It is in addition necessary to develop the theory to handle goals and any other concepts (e.g. intentions, plans) required to drive agent and system action. While the belief revision theory, viewed in a sufficiently abstract way, subsumes these notions, it is necessary to ensure that they are made specific enough to be effective for agents that have to organise and execute actions.

2.3 Increased Coherence Model implementation

The generic class of mechanisms for creating and modifying belief sets, truth maintenance systems (TMSs), has an intrinsically exigent job to do. We have implemented a particular mechanism, ICM (for Increased Coherence Model), which can operate efficiently as the effort of set manipulation is reduced from the general case both by staged processing and by confining this to sets affecting core beliefs.

Processing has four stages, each addressing one of the factors contributing to the preference ordering on sets of beliefs. The first establishes a baseline by identifying all the maximal sets that are internally consistent and self-justifying relative to the context. The remaining three stages deal successively with connectivity, endorsement and conservation in relation to these consistent sets.
Connectivity is investigated to identify those sets offering the most additional derivational support links (proofs) for core beliefs; endorsement is evaluated to identify the sets with the best overall endorsement; and conservation is used to identify the sets making the least change to the previous state. As connectivity is more important than endorsement, and endorsement than conservation, the ordering is significant, with each stage constituting a filter: the processor for the next stage is only invoked where the previous stage has not selected a single preferred set and further discrimination is required. It could thus happen that revision is determined solely by connectivity considerations, or that endorsement has to be taken into account as well, or that conservation has also to be invoked, perhaps even then without final resolution: this reflects the absolute priorities rather than relative status the theory gives to different types of information about beliefs, within its generally conservative framework.

The ICM builds and maintains a ‘cognitive state’, the currently preferred belief set(s), consisting essentially of propositions. Whenever the system receives an input ‘observation’, it reasons about whether to accept this new belief or, necessarily alternatively, its complement. It may be possible after investigation to add one or the other member of the pair to its current set(s) while maintaining consistency. However, if consistency cannot be maintained, new sets have to be built from scratch. This process is nevertheless controlled, because the construction is limited to starting from the ground assumptions in the system’s entire database of beliefs, together with those assumptions which were ‘pervasive’ in the previous cognitive state, i.e. occurred in every one of its belief sets. The ICM thus forms all the new consistent sets for these assumptions, and then applies its connectivity, endorsement and conservation metrics to identify the most preferred set(s) to form the new cognitive state. The actual construction process is driven by an ATMS-type conflict-resolution algorithm, which determines which members of complementary pairs of assumptions should be removed from some network of beliefs, along with their supporting propositions and also any derived ones not themselves assumptions. Overall, the response to observed inconsistencies is to check justificatory chains and delete faulty ones, as these depend on defective foundational assumptions.

3. THE MODEL OF THE LIBRARIAN

While the theory of belief revision provides a general base for motivated action including dialogue, it is also necessary, in seeking to automate the intermediary, to consider the task-specific goals and knowledge the intermediary has: what particular characteristics does a librarian have that need to be modelled by the system as the agent interacting with the information-seeking user?

The project is taking work by Belkin, Brooks and Daniels, hereafter referred to as BBD, as a starting point. This work was based on real library dialogues, but it must be emphasised that everything has to be ruthlessly simplified for our project: so whether the real library situation is one where the literature is to hand and the usual means of access is via a conventional catalogue, or one where references to literature are obtained via an online search service is immaterial. The BBD model is a completely general one, intended indeed to apply to all types of information-seeking situation and not just library or literature search service ones. It is also intended to cover the range of enquiries stretching all the way from quite definite requests for known items to very indefinite, barely formulated needs for unknown items. But as the earlier examples suggested, the typical situation is the topic or subject search for unknown items, of the kind associated with online search services. Other research so far by (for example) Pollitt, Vickery et al. and Brajnik et al. has also been concerned, in different ways, with this situation; and Chen and Dhar have proposed a model resembling BBD’s, but simpler, and also based on observed dialogues, for this case. The example of Fig. 1 assumed a subject-based search of an online book catalogue, rather than the more common subject search of journal literature, but the generic situation is the same. Searching in these contexts is of course usually iterative: our initial simplification for experimental purposes is to treat the point at which the first actual search formulation is submitted to the online system as a stopping point; but this does not affect the general form of the agent–user interaction, and iteration can be incorporated later, as it is clearly essential for a realistic and effective system. The situation being modelled will be referred to for convenience as the library situation, regardless of whether there is an actual library with literature to hand, and of whether books or papers are in question.

The essential point about the situation being modelled is that the user has a need for information, and knows what the context motivating this need is, but that he cannot by definition fully characterise the information needed because he has not yet read the documents which supply this information. The user does not have technical knowledge of the access routes to the literature either, i.e. of the indexing vocabulary, classification scheme or whatever, or of the library or information service holdings and coverage. The librarian, on the other hand, does not, indeed cannot, know the user’s individual need, or the user’s personal motivating context. But the librarian does have technical access and holdings knowledge, and typically also has generic subject area knowledge, and user population knowledge. Thus as the earlier dialogue showed, the two parties to the library interaction have mutually complementary starting knowledge, but the process of putting these to work on another is not just a transfer operation: it is a constructive one, since it is necessary to formulate the user’s need sufficiently fully and explicitly for it to serve as a basis for a search specification which is intended to be an effective means, descriptively and selectively, of obtaining relevant literature, given the particular properties of the available document collection or file.

3.1 Belkin, Brooks and Daniels’ approach

After considering all these factors, BBD have suggested that an appropriate way of modelling the librarian is as a set of subtask processors, or functional experts, each with their own specific resources and each satisfying their own data-gathering goals, but in doing this collectively contributing the data required to achieve the overall system goal, namely to enable the user to satisfy his information need. In general this may be done either
A. CAWSEY, J. GALLIERS, S. REECE AND K. SPARCK JONES

directly, or indirectly by providing pointers to documents. But in some cases it may prove impossible to help the user: thus the outcome for the system is more correctly characterised as satisfying the goal of doing the best for the user, as mutually agreed. For the simple experimental case being studied by the project, however, this is taken as agreement on a first-pass search specification.

The justification for the model BBD propose is that very distinctive bodies of knowledge and processes are required for the various tasks contributing to the overall goal of satisfying users' information needs. Thus librarians deploy quite specific knowledge about indexing languages and techniques, for example, and have particular knowledge about individual document collections, even if they also back us this specialised knowledge with a more general 'ordinary' knowledge base. At the same time, forming an effective or adequate search specification calls not only on the topic description itself but on information about the type of user, the type of literature wanted and so forth. Individual processors may also seek data satisfying a variety of subgoals, for example for the user both general educational experience and level of familiarity with the particular area in question.

The complete set of processors BBD propose is quite large. It includes both what may be thought of from the global task point of view as central processors and support processors. The complete set, embodying some compromise between BBD's various publications, and with some renaming for present convenience, is shown in Fig. 4, along with very simple illustrations of the kinds of state they might be in at about (though not necessarily precisely simultaneously) the end of the dialogue fragment of Fig. 1. These illustrations are simply indicative, however, and are not intended to make any claims about the proper way of representing processor results. The central processors are those bearing directly on the user's information need. They include the Problem Description expert, intended to capture the user's topic and its broader conceptual context or subject area, deemed in the example to be conflated as the notion represented by 'cactus cause disease'; the Problem State expert, showing the status of the user's progress with his subject and topic, in this case just starting finding out; the Problem Mode expert, characterising the manner of information gathering taken as appropriate for the user to supply his need, in this case reading (as opposed to, say, talking to someone); the User Model expert, giving the relevant properties of the user, e.g. householder (not horticulturist); and the Retrieval Strategy expert which produces the means of access to the description or document file, in this case taken as a Boolean request in a controlled indexing language.

The supporting subprocesses cover Dialogue Mode for the form of interaction between the user and the librarian, for instance continuing talking about the user's topic, etc. as opposed to looking at actual documents; Explanation Provision, concerned with the kind of information the librarian gives the user about what is going on – in this case we may suppose that a rather broad search specification has been formed because the library holds little material on plants; Input Analysis, designed to interpret the user's natural language input, e.g. 'No, on diseases they cause'; Response Generation, for planning and organising the form and content of system responses to the user, e.g. checking whether material on non-cacti would be appropriate; and Output Synthesis, for producing natural language output, e.g. 'Other house plants ...

BBD's claim for the range and nature of the knowledge sources contributing to the librarian's task performance as a whole is based on a detailed and careful analysis of human examples, including protocols taken from dialogues between library users and online search service intermediaries. The analysis also shows that the functional processors may be quite complex, with subprocessors with subgoals to be satisfied in support of a processor's overall goals. BBD thus argue that the natural model for the librarian is as a distributed expert system with multiple agents having their own individual tasks, but cooperating by supplying data any other expert may use by posting messages on a common blackboard. From this point of view, indeed, the user is just another agent, albeit one mediated by the Input Analysis processor.

The motivation for adopting this data-driven model is that the detailed study of human user–librarian interaction shows how very free and flexible dialogue structure is in terms of how far individual goals are pursued at any point, and in what order, when they are revisited, and so forth, and also in terms of the way any individual item of data is obtained. The dialogues show exchange sequences delimited by conversational boundary markers and shifts of discourse topic, with each sequence, or focus, concentrating on one task or another. Overall the dialogue may show a gradual tendency to move from concern with the User Model, through the Problem Description to the Retrieval Strategy, but there is great variation in the detailed pattern reflecting the way in which the needs of different subtask processors are addressed. At the same time the analysis of the dialogues shows that at some times a piece of data required to satisfy some goal comes directly from the user, at other times may be derived indirectly from data primarily relating to another goal. For example, information about the user's expertise relating to the User

Figure 4. Librarian model processors with illustrative information states for the dialogue example of Fig. 1.
Model or about the user’s Problem State may be supplied by the user, or it may be inferred from the type of literature requested, itself a concern of the Problem Description expert; for example, a request for an introductory textbook suggests the user may be a beginning student and/or someone just beginning work in the relevant area. The general presumption is that as the individual processor’s data needs are satisfied, whether via responses from the user to system data requests or contingently via other processors, the system’s collective needs are also satisfied.

3.2 Control problems

As each processing agent in BBD’s model has its own area of knowledge, which it deploys in the context of communications from other agents including the user, it is easy to see that BBD’s approach can be couched in terms of belief revision at the level of the individual processors and hence that of the system as a whole. But there are significant difficulties with the system architecture, in terms both of the structure of relations between agents and of operational control. The control problems refer, moreover, both to overall system management as this affects the internal communications between agents and to the management of external dialogue with the user. BBD’s proposed architecture is a structurally open one with very weak control, since there is no real interference with the blackboard to organise internal communications, and output to the user is apparently triggered simply by ‘overflow’ on the board.

But quite apart from the danger of internal thrashing, effective dialogue with the user cannot be conducted simply by picking off the individual most pressing request for data. The system needs to be able to make an informed evaluation of the state of the blackboard and to have a more controlled organisation of dialogue as a means of data gathering. This in turn seems to imply a controller with the capacity to evaluate message content as well as source, and to prioritise and group messages on this basis.

However, there are three key challenges to be overcome in choosing the appropriate architecture, given the task situation being modelled. These are, first, the weakness of the notion of satisfaction for subprocessors, especially key processors like the Problem Description one. Data gathering cannot be driven, as it can for many other tasks, by a check-list approach. Limited implementations of the automated intermediary like Pollitt’s are able to operate effectively in menu style, with known slots and filler possibilities. But it is not possible in general, for example, to capture topic information by a menu approach because the range of possibilities is too large, and determining whether a given topic characterisation is adequate can only be based on a number of criteria which are individually weak. Simply inviting the user to ‘tell me more’ is not an effective strategy either, but when satisfaction is weak it is difficult to determine what a system’s output should be. The satisfaction problem also applies at the level of the system as a whole: what, in the likely absence of clear indications from individual processors, determines whether the entire ‘information need problem’ has been satisfied?

The second major problem to be resolved for control and dialogue management is the open data sourcing, that is, the fact that useful or desired pieces of information can come from other processors or from the user. For example, the Retrieval Strategy processor may be able to obtain data for a search specification from the Problem Description or User Model or Problem State modules, or from the user via the Input Analysis module. This makes it difficult to determine whether an attempt to obtain information should be forced by embarking on dialogue with the user or should be awaited from any source (including volunteering by the user).

The third problem is the separateness of the user. At the fine-grain information level, there is no predictability in the user, however cooperative they may be both in relation to the task as a whole and in relation to the local dialogue context. This is not so much because individual user responses to system questions or statements may not fit tightly, but because the user is a genuinely independent agent (in the way the other processors are not) who may choose to take his own initiative in the way the dialogue is conducted.

BBD’s distributed data-driven model is attractive in allowing for the heterogeneity of the resources and processes involved in the information task, and for the arbitrariness of the data, in terms both of the nature and the timing of items of information. The issues it raises are whether internal communication can only be in the open, blackboard style; how much control is needed to regulate internal activity and to manage external dialogue; and how these two control processes are related if, as is possible given their rather distinct functions, this involves two distinct system components, a global system controller and a specific dialogue manager.

3.3 Architecture refinement

Belkin, Hennings and Seeger (BHS) began to address some of the questions just raised in simulation experiments designed to study different architectures for the automated intermediary. In these they compared blackboard and actor versions of the distributed model — i.e. architectures where internal communication is via a blackboard with architectures where internal communication is directly between an agent and other specified agents; and they compared uncontrolled and controlled communication regimes — i.e. regimes with no and with some monitoring, prioritising, etc. of message flows. BHS concluded that their experiments showed that a blackboard architecture is appropriate, and specifically that it is superior to an actor one. But they also concluded that it needs a positive control regime: a simplistic free-for-all model is too weak.

Unfortunately, though the experiments were quite carefully conducted, the fact that human agents were involved meant that the simulations were not specified at the level required for machine implementation, and crucial questions about the powers of the controller (called the Blackboard Analyst) and the relationship between the Analyst and Response Generation were therefore finessed. However it is clear that the Analyst needs a strong message interpretation and task evaluation capability, given the observed complexity of, and dependencies among, the experts’ activities.

Fox’s Coder system, on the other hand, is an actual computational implementation of a distributed blackboard-based system for information retrieval. Its functional experts were envisaged as more restricted than
BBD's, but CODER's original design nevertheless had a powerful strategist/planner controller, with a whole range of directive functions. The implementation is, however, more limited. Thus the user interface is system driven and menu based, and problem mode, state and description have been combined as a single expert, which has become the dominant module since its rule base determines most of the system state changes. The strategist seems to have become more of a middle manager, keeping things running. Unfortunately, there is no information in, for example, Fox et al.\textsuperscript{13} about the detailed character and behaviour of the architecture.

Croft and Thompson's I3R system is a prototype implementation of the intelligent intermediary which has much in common with Fox's system.\textsuperscript{2} It has various experts collaborating within a data-driven blackboard framework, but operation is firmly in the hands of a scheduler carrying out a prior plan, and as the underlying model of indexing and searching is very restricted, control is straightforward.

The architecture of both of these systems is thus less distributed in practice than in principle, and is much like that used in Brajnik et al.'s IR-NLII.\textsuperscript{3} This is essentially a more sophisticated version of Vickery et al.,\textsuperscript{23} using a rule-based approach both for need specification and strategy formulation, supplemented by a user modelling component: this is designed to gather user properties like background, experience and retrieval history, taking stereotypes as a starting point. IR-NLI's operation is essentially system driven through a well-defined, possibly iterative, sequence of steps from information need capture to database search, with communication with the user modulated by the user-modelling component (and not yet in free natural language).

The proposal for a distributed model of the librarian thus has to be evaluated at three levels: whether distributed processing is in itself right; whether processors should interact in blackboard or in actor style; and whether interaction should be essentially uncontrolled or controlled. For all three, there are no independent arguments one way or the other: the particular choice of system organisation depends on the characteristic properties of the librarian's task. The discussion in this section might perhaps suggest that a distributed architecture is not the one for an automated librarian. But there are merits, for a complex, technically based and skilled task, in the notion of distinct special-purpose knowledge sources. It is thus fair to start from BBD's basic position that a distributed architecture is appropriate for the information-seeking case. However, it is not merely desirable in principle, but necessary in practice for computational implementation, to tackle the question of the nature of inter-agent communication and of control, both overall and for dialogue management. There is the further matter, from this point of view, of seeing how Galliers' theory of belief revision works out within the distributed framework. One of the two major objects of our project is to develop, and validate (if simply) by computational implementation, this theory. It fits very well with the notion of distributed processing, as it is a theory of interacting agents. But it is necessary to show that it can be implemented so that both each agent's individual behaviour, and the agents' collective behaviour which defines the librarian, are sensible.

4. PROJECT STRATEGY AND PROGRESS

There are clearly many tricky problems to investigate for our project, associated with the system architecture and the management of dialogue with the user, and with the performance of the belief revision mechanism as the modus vivendi of any system component and hence of the system as a whole. In the longer term, of course, there is the nature of the task and domain knowledge needed for effective functional experts.

4.1 Initial system design and implementation

The approach we have adopted to begin the project is to build a very simple version of the librarian, meeting only the most basic requirements for conducting a simple interaction with a simple user. This is primarily to obtain a working implementation of the essential belief revision and communication apparatus instantiated for the library case, using BBD's functional model with distinct experts. We are currently testing and evaluating alternative architectures and control regimes within the context of this simple initial system.

The initial Mark I system (see Fig. 5) has only a few of the functions of the full BBD model, namely the key Problem Description and Retrieval Strategy ones, the User Model function, and an 'Interactor' function, which is concerned with managing the interaction with the user. This latter function is not strictly one of those enumerated by BBD, but conflates some of the simpler functions of the Input Analysis, Response Generation and Output Synthesis experts. (The appropriate relations between these are not obvious, and BBD's treatment is not convincing.) As the Interactor is the exclusive channel of communication with the user, the user may be deemed from the system's point of view to be the Interactor; in a more general and abstract model the actual user is yet another agent within a larger 'system' embracing the actual computational one.

In our first version system the knowledge contained within each expert consists of a small number of inference rules and some simple data structures, in fact motivated by an actual dialogue about literature on Greek–Turkish relations BBD recorded and analysed (Brooks,\textsuperscript{4} pp. 284 ff.). Thus in Retrieval Strategy, for example, we have data structures representing the attributes of the different literature databases, and inference rules linking desired document attributes with the most appropriate database (much like Chen and Dhar's handles\textsuperscript{7}). The document attributes include subject area, for instance history, to which a whole database might be devoted; document type, for instance journal; and document restriction, for instance date specifiability—meaning, as would be natural for historical materials, that there is a specific field in each document description indicating the time period to which the document refers (e.g. '19th century'). The inference rules are ones like 'subject-area = history $\implies DATABASE \ 6$', 'document-restriction = date specified $\implies DATABASE \ 6$', 'document-type = journal $\implies DATABASE \ 5$ or 6'. Illustrative concepts for the other experts are shown in Fig. 5. Thus Problem Description has 'user-topic = greece', for which an inference rule would derive 'subject-area = geography'. Inference rules can be applied either forwards, to obtain further conclusions based on new data, or backwards, to try and satisfy some goal (such as determining whether a
AUTOMATING THE LIBRARIAN: BELIEF REVISION

Problem Description (PD)
- doc-cont (post-74)
- u-topic (Greece)
- subject-area (geography)

Retrieval Strategy (RS)
- doc-cont (post-74)
- doc-restriction (date)
- DB (history)

User Model (UM)
- u-status (student)
- u-goal (thesis)

Blackboard (BB)
- doc-cont (post-74)
- u-topic (Greece)
- u-status (student)
- DB (history)

User (U)
- Interactor (I)
  - bel (U, doc-cont (post-74))
  - bel (U, u-topic (Greece))
  - bel (U, u-status (student))
  - bel (U, not DB (history))
- User Model (UM)
  - u-status (student)
  - u-goal (thesis)

5a Blackboard architecture

5b Actor architecture

Figure 5. Illustrating blackboard and actor architectures for the Mark I system, working on 'Greek–Turkish relations' dialogue.

particular hypothesised database is appropriate). The form of the knowledge and the inference strategies are the same in each expert module. This is an acknowledged simplification, as BBD have suggested qualitatively different ways of representing the knowledge and reasoning appropriate to the different modules.

In order to test BBD’s proposal for a distributed architecture we have implemented the different expert modules as parallel distributed processes operating on separate machines. We are currently investigating two contrasting types of communication structure and associated control regime within the distributed architecture, each in its most basic or ‘naked’ form, as well as different hybrids. Thus we are comparing basic blackboard and actor architectures, and some combinations of these.

In the blackboard architecture, as originally proposed by BBD, all the modules communicate via a common data structure, the blackboard, which holds messages reflecting different aspects of the current global problem state. Messages are sender-labelled but not receiver-labelled, and agents have no prior knowledge of the capabilities of other agents, though they make inferences about them from their messages. We assume that any agent can in principle use any message, and also that any agent can in principle supply any message. The results delivered by any agent’s internal operations are in principle communicable, though an agent’s evaluation of its task state may mean they are not—or not immediately—communicated (see below). The messages communicated are of different types, requesting or delivering information, either voluntarily or in response to requests; when hypotheses are communicated which conflict with other blackboard hypotheses, they may be accompanied by their reasons. Fig. 5a illustrates a hypothesis state for the model as it might hold at some point in a dialogue about Greek–Turkish relations.

From the control point of view, the blackboard model is implemented in the simplest possible way. There is no global controller, evaluating the system’s state and manipulating the contents of the blackboard (as there is in Fox’s CODER). The blackboard simply holds all posted messages which are read by all agents, and processing terminates, i.e. the system’s overall goal is deemed satisfied, when all the agents (and hence also the user) have no pending tasks. As messages posted to the blackboard are read immediately by all the agents, when all the agents have no messages or other tasks requiring attention this implies that there are none outstanding on the board either. As a termination condition this is
and causes belief revision, according to the theory between the system and the user, is both motivated by the receiver. Of course strength for the communicator does not necessarily apply to the blackboard model, and in fact in our initial version each expert may communicate specifically with any other expert, as needed. Control for the actor model again has the simplest form, in this case meaning that there is no attempt to restrict the flow of communication between modules: each agent has to decide itself what to do with whatever arrives. The systems processing is again terminated when all the agents have no outstanding tasks. Fig. 5b illustrates the actor model for the same dialogue situation as is shown for the blackboard model in Fig. 5a.

In discussing the behaviour of an individual module in more detail, we will assume agents are communicating with one another directly, as in the actor model. However, the points made generally apply to the blackboard model as well.

Communication between agents is based on a simple common formal language. We are currently not addressing natural language per se, so we communicate with the user 'behind' the Interactor exactly as with any other agent, in the formal language. A message consists of a type of speech action, the sender name and receiver name, a proposition or package of propositions, and optionally a 'strength', which can be weak or strong. For example, the message 'tellref(UM, RS, user-status(student), strong)' informs the Retrieval Strategy module that the User Model module strongly believes the user is a student, while the message 'askwhy(RS, UM, user-status(student))' is a request from the Retrieval Strategy module for further justification for this belief. The speech acts thus embody the simple message types we allow, currently two forms of information request, 'askref' for data and 'askwhy' for explanation, and two forms of information delivery, 'tellref' for hypotheses and 'tellwhy' for explanations ("tellref" is not necessarily an answer to a request). It must be emphasised that strength here refers to the status of a belief for its communicator (i.e. utterer), not for its receiver. It is a simple, ad hoc version of a belief's status as this is associated, for the communicator, with the difficulty of disbelieving it: a strong belief is harder for the communicator to disbelieve, because of its revision consequences for him, than a weak one. Of course strength for the communicator does not necessarily imply strong endorsement, as defined earlier, by the receiver.

Communication between the modules, and thus also between the system and the user, is both motivated by and causes belief revision, according to the theory developed by Galleyars as outlined in Section 1. As we saw there, this theory of belief revision determines how a new communication influences the hearer's beliefs. The communicated proposition may or may not be taken on by the hearer, and if taken on may influence a whole web of related beliefs, whether the new communication is directly sent as in the actor model, or read off the blackboard as soon as posted. Depending on whether the proposition is taken on or not, and on how other beliefs are changed, different messages may be sent or posted in response. These are constrained by a set of Interaction Rules (IR) which, in conjunction with the belief revision process, dictate how an agent may respond to particular communications and communicate belief changes following the operations of belief revision. For example, if an agent is informed of a proposition P but does not take on that belief, then depending on a number of factors such as past communications it has received and its view of the communicating agent's beliefs, the receiver may respond, directly or indirectly via the blackboard, by trying to convince the sender of not-P or by asking the sender why they believe P.

When one agent in the system is attempting to convince another, whether an individual agent in the actor case or presumed other(s) in the blackboard case, of the truth of some proposition, the agent uses simple models of the other to strategically plan a message which it believes will cause the desired belief change. Of course these models of other agents may be inaccurate, so the desired change may not occur. Because of this there may be a complex negotiation between agents as they argue for or against the truth of the proposition in question.

This general picture is complicated by the fact that any agent in the system may have several things to work on at once. Agents may try to determine the truth of some proposition, and may be able to draw new conclusions from new beliefs, to respond directly to incoming messages or to communicate new beliefs to interested agents. An agent in the system has to prioritize these tasks, for which it uses a set of Task Prioritization Rules (TPR). For example, the initial system deals with new messages before old, and puts at low priority drawing new conclusions from weakly held evidence.

As far as the individual agent is concerned, therefore, it has three components: the IR component, the TPR component, and the BR component which embodies the agent's belief revision mechanism (BRM) and sets of beliefs constituting its belief state (BS). In general terms, though the precise relationships between the components have still to be determined, the TPR component manages the task agenda for tasks derived either from incoming messages or from changes in the BS, while the IR component manages the precise form of input-output message linking and output message expression. In the current simple model, information requests just access the BS, while information deliveries stimulate the BRM.

4.2 An illustration

The way our simple models work can be illustrated by what happens when we emulate the fragment of the Greek-Turkish relations dialogue mentioned earlier, which is shown in Fig. 6. In this example the user wants to get hold of documents on Greek-Turkish relations, post-1974. The intermediary suggests the history data-
Intermediary: Um, the only other possibility is Historical


Intermediary: It’s the only database which has really, obviously because it deals with history tried to, cope with this time limitation.

Figure 6. Fragment of a recorded dialogue between a librarian (an intermediary for searching using an online service) and a user seeking literature on Greek–Turkish relations (from Brooks, 1986, p. 293).

base, presumably because it is the only one that allows the user to ask for documents with post-1974 content. However, the user rejects this suggestion, presumably believing that the history database would not have recent material, and not understanding the need for choosing a database which allows documents to be selected through an explicit date restriction. We show below how we can model the crucial features of this dialogue using our belief revision apparatus. The example is simplified for the purposes of illustration, and we assume an actor architecture. It should be emphasised that in this discussion, as we are imagining we are communicating with the real user, we need to separate the Interactor from the user and to refer explicitly to ‘User’ as a communicating agent in messages.

We represent the system’s initial beliefs, and its initial beliefs about the user’s beliefs as follows, where beliefs are either assumptions which are endorsed using the endorsement types of Fig. 2 or are derived from the indicated assumptions:

System’s beliefs:

1. doc-content(post-1974) = = >
   doc-restriction(date): 2cs.

2. (doc-restriction(date) = = >
   database(history): default.


4. (doc-restriction(date): [from 1, 3].

5. database(history): [from 1, 2, 3].


Belief 3 is in the Problem Description module, and the other system beliefs, 1, 2, 4 and 5, are in the Retrieval Strategy module. The system’s beliefs about the user’s beliefs, here belief 6, are held in the Interactor module, as they are needed in order to communicate effectively with the user. Belief 3 is also in Retrieval Strategy, to which it has been communicated.

Initially the system, and specifically Retrieval Strategy, has the task of getting the user to believe that the history database may be appropriate. Using the ICM mechanism Retrieval Strategy predicts (based on its model of the user’s beliefs, transmitted from the Interactor) that simply informing the user (via Interactor) of this should cause the desired belief change. This message is communicated weakly as it is not strongly held by the system, via the Interactor:

RS: tellref(RS, I, database(history), weak)

I: tellref(I, User, database(history), weak)

The user responds by strongly rejecting the suggestion:

User: tellref(User, I, not database(history), strong)

The Interactor now updates its model of the user’s beliefs with the information that the user does not believe that the history database is appropriate. This belief is then passed on to Retrieval Strategy with a strong commitment. However, Retrieval Strategy modifies this to a weak endorsement because it does not believe the Interactor is knowledgeable about database selection. Retrieval Strategy also recognises that the new communicated user belief conflicts with its existing belief (that the history database is appropriate), and considers revising its beliefs. However, the result of the belief revision process is that the Retrieval Strategy module holds on to its existing beliefs. But as it recognises that this conflicts with the beliefs of the Interactor (and hence the user), it responds by trying to convince the Interactor that the history database is appropriate. It does this by providing support in terms of the following information:

RS: tellwhy(RS, I, database(history) because
   [doc-content(post-1974) = = > doc-restriction(date)
   strong &
   doc-restriction(date) = = > database(history) weak &
   doc-content(post-1974) strong]]

(i.e. ‘the reason for choosing the history database is that wanting post-1974 material is a kind of date restriction, and the history database is suitable if there is a date restriction’). The Interactor takes on this justification and passes it to the user, who then decides (for whatever reasons he has) to weakly accept the system’s suggestion:

User: tellref(User, I, database(history), weak)

4.3 Evaluation of the simple models

The example dialogue illustrates the basic functioning of our initial distributed system in one of its modes. But much more work is needed before we can assess the overall approach. The initial system works only for a small range of restricted problems, and is also unacceptably slow. The models we have implemented are very simple indeed, though they are still a helpful base for investigating crucial aspects of our approach to information communication and belief revision in a distributed system. Currently control is entirely in the hands of the agents, who each manage their own operations. But an agent may find it difficult to determine whether the input it receives from other agents, especially in response to requests, is ‘adequate’ and is all it is going to get. With more agents with more knowledge engaged in a more serious literature-seeking interaction, more work will be required and more messages can be expected, suggesting a need for more global control on the operations of the system as a whole. Conducting rational and comprehensible dialogue with the user also suggests a need for more control on the interaction between system and user. Currently the system’s behaviour is unfocused, with messages poorly related to the overall task or to the prior pattern of communication, whether internal or external. The system currently has no serious notion of satisfaction either for any individual agents or
for the system as a whole. These are important issues that have to be addressed, and there are also embedded ones to do with specific aspects of our general approach that need tackling. One of the most pressing of these is developing a proper view of how goals are related to beliefs within the framework of the theory of belief revision: in the present simple models the way goals as represented by tasks are established is somewhat arbitrarily determined by the details of messages and belief revisions. A more coherent and motivated account of the way goals emerge from or are associated with beliefs is required. Finally, it is clearly necessary to give the experts more knowledge to deploy, and to increase the number of agents, for instance by adding a Problem State module.

5. CONCLUSION
As described, the motivation for our project is to combine a general theory of agent behaviour based on belief revision with a specific theory about the librarian as a collective agent. The two fit naturally together, and our aim is to use each to throw light on the validity of the other, and to take the first steps towards automating a more powerful intelligent librarian or information intermediary than the other techniques so far tested allow. The architecture and control investigations and measurements with which we are currently engaged should give us the more solid base we need in order to construct a next version of the model implementation with a little less trivial, if not very extensive, relevant task knowledge.

Acknowledgement
This research is supported by grant SPG8930752 of the UK Tri-Council Initiative on Cognitive Science and Human–Computer Interaction.

REFERENCES