Conflict and Cooperation in a Heterogeneous System*

Alison Cawsey, Julia Galliers, Steven Reece and Karen Sparck Jones
Computer Laboratory, University of Cambridge
New Museums Site, Pembroke St, Cambridge CB2 3QJ, England
{ac,jrg,sr,ksj}@uk.ac.cam.cl

Abstract

In this paper we describe a framework for negotiation between autonomous agents. Each agent can reason about how and whether to take on communicated information, and the presence of conflict motivates negotiation. In order to test these ideas we have implemented a simple distributed information retrieval expert. Negotiation takes place both between the system agents, and between system and user, also viewed as an autonomous agent.

Introduction: Background Theory and Concepts

In this paper we present an overview of some research concerned with how autonomous agents negotiate in order to resolve conflicting beliefs. The background of this work is a theory of communication where conflict is seen as a positive force, motivating useful communication (Galliers, 1989a), and the resolution of conflict is based on negotiation, with each agent autonomously reasoning about what they prefer to believe (Galliers, in press). Agents are genuinely autonomous, needing no arbitrator or techniques for compromise. Each tries to convince the other of its own point of view, but given new communicated information each agent may revise its own beliefs, and possibly achieve agreement with the other(s).

Previous systems (and models) of communicating agents have generally either avoided conflict (e.g., Georgeff, 1983), assumed that either the speaker or hearer’s beliefs are authoritative (e.g., Cohen & Levesque, 1990, Perrault, 1990), or assumed that conflicts may be resolved by strategies such as compromising or ‘randomly’ changing beliefs (e.g., Lander, Lesser & Connell, 1991). Exceptions are Sycara (1988), who shows how persuasive strategies may be used to achieve agreed compromise, and Carver et al. (1991), who show how uncertainty may motivate communication involving the reasons for beliefs. However, neither system provides a general framework for negotiation and belief revision between cooperating agents. In our work we show how conflicts in the beliefs of cooperating, non-adversary agents in a multi-agent system may be resolved by negotiation and autonomous belief revision.

Our basic approach can be compared to Mason and Johnson’s DATMS framework (Mason & Johnson, 1989). Like them, we are concerned with sharing tentative hypotheses among a set of sceptical agents, revising beliefs in the face of conflict. However, Mason and Johnson completely ignore the issue of preference given alternative consistent sets of beliefs1, and as a result, the form of the negotiations that may take place between agents are very restricted. These negotiations consist simply of one agent informing another of inconsistent sets of assumptions. In our framework, agents choose what to believe out of alternative consistent sets, and negotiation and belief revision depends on the justifications provided for assertions.

The framework for autonomous belief revision has been described in (Galliers, in press). This framework allows each agent to determine both how and whether to revise beliefs in the face of new potentially conflicting information. Preferences between alternative consistent sets of beliefs are determined by a three tiered ordering which considers both the coherence of beliefs, foundations for beliefs, and minimal change.

Coherence is measured with respect to certain salient or core beliefs – propositions which provide additional proofs for these core beliefs increase the coherence of the ‘network’ of beliefs. The ‘strength’ of foundational beliefs is compared using endorsements (cf. Cohen, 1985), which provide heuristic descriptions of their sources. These endorsements refer, for instance, to whether the source of the information was a first hand communication, second hand or based on a default inference2. For communicated information the source may further be

---

1For example, given inconsistent assumptions A and B it is unclear whether their agents would believe beliefs with just A in their assumption set, beliefs with just B in their assumption set, or both.

2Note that first hand here refers to sensory perception, while all linguistic communication is labelled as ‘second hand’.
Endorsement types:

1cs: first hand communication, strong (ie, reliable sensory perception).

1cw: second hand communication, weak (ie, unreliable sensory perception).

2cs: second hand communication, strong (ie, reliable, linguistically communicated information).

2cw: second hand communication, weak (ie, unreliable, linguistically communicated information).

df: default conclusions.

h: hypothesis.

Heuristic ordering on types:

1cs \leq 2cs \leq 1cw \leq 2cw = df \leq h

Figure 1: Endorsement types used in current system and partial ordering between types

strong or weak, depending both on the believed authoritiveness of the speaker and the ‘strength’ with which the information was communicated. A belief will be strongly endorsed if communicated strongly AND the speaker is believed authoritative on the associated topic. A range of endorsement types are given, and a partial order defined among them. The main types used in the current system are given in figure 1. For example, the endorsement type 2cs refers to a strongly endorsed second hand communication (ie, received as a strong communication from an authoritative source).

The system implementing this theory first obtains alternative consistent belief sets (extensions) from the propositions in the system’s database, and then extracts preferred sets according to the three tiered preference algorithm. These sets will each include every proposition or its negation, and define the agents current cognitive state. Propositions which are in only some of the preferred sets are viewed as uncertain, because the proposition and its negation can be viewed as equally preferred.

Unlike previous belief revision theories and reason maintenance systems (e.g., Gardenfors, 1988), our system addresses the issue of preference between alternative consistent extensions in a computationally tractable fashion. We take into account both logical relations between beliefs, and simple heuristic assessments of belief strength based on their source. Preference is considered between whole interconnected sets of beliefs, and not just individual propositions.

In the rest of this paper we will describe the overall architecture for negotiation between autonomous agents, where each agent can autonomously reason about its beliefs as just described. Our ideas are being tested by developing a multi-agent framework for information retrieval.

Conflict and Cooperation in an Information Retrieval Task

In the information retrieval task (described in more detail in Cawsey et al., 1992a) the user and system must negotiate in order to determine how best to satisfy the user’s information need. This need may be satisfied by finding books or papers relevant to the user’s topic, possibly using an online search service as the information retrieval ‘back end’. This is a complex and difficult problem, requiring many types of knowledge. The user’s description of their need is often poorly-specified, and cannot be immediately translated into an appropriate information search request. Bridging the gap between the user’s initial query and an appropriate search request, requires (at least) knowledge of the user (e.g., their status and goals), knowledge of their problem, and knowledge of retrieval strategies (ie, the information retrieval ‘back end’).

Several authors have proposed that the extensive and varied activities involved should be implemented through a set of interacting functional experts (Belkin et al., 1983, Fox et al., 1988, Chen & Dhar, 1987). Belkin et al base their model on very detailed analysis of dialogues between human information retrieval intermediaries and users. They show that the kinds of reasoning and information resources required for the individual functional experts are diverse and complex (Brooks, 1986; Daniels, 1988), and also that the interactions between the modules (functional experts) is also complex and highly flexible. There is no straightforward procedure for determining the information required for the retrieval task – tentative hypotheses are established based on the interaction with the user, and either refined or abandoned as interaction progresses.

Belkin et al suggest a blackboard based architecture for the retrieval task, with interactions primarily data driven (though mediated by a controller). However, in our own work we have found (through empirical evaluation) that an actor architecture is more efficient, as the negotiations and communications are directed more appropriately (Cawsey et al, 1992b). In our early prototypes we used a solely data driven approach, but are now exploring flexible goal based approaches, where these goals are dynamically created and revised based on new information (discussed further below).
In our framework for information retrieval, negotiation must take place both between the user and the system, and between the individual functional experts within the system, as tentative conclusions are considered and revised. To illustrate the type of negotiation which should take place between system and user, consider the dialogue fragment given below, taken from transcripts of dialogues between human users and information retrieval specialists (Brooks, 1988):

1. IR Specialist: Um, the only other possibility is Historical Abstracts but it it
2. User: No.
3. IR Specialist: it is fairly, they can include some recent material...
5. IR Specialist: We'll think about it we'll see we'll put a query by that one. Mm.
7. IR Specialist: It's the only database which has really, obviously because it deals with history tried to, cope with this time limitation.

In this example it has already been established that the user is looking for documents on 'Greek-Turkish relations after 1974'. The information retrieval specialist tentatively concludes that the History Abstracts document database may be suitable because it allows searches for documents with particular date restrictions on the content (e.g., post 1974 material). However their suggestion (utterance 1) is initially rejected (utterance 2). In utterances 3 and 7 the specialist justifies their suggestion, attempting to convince the user that the suggestion should be considered. In deciding how to convince the other, we suggest that each agent has to reason about the other's beliefs, as well as their own. For example, the specialist above appears to base utterance 3 on an assumption that the user might think that a history database only contained old material. The apparent conflict between the beliefs of the two participants motivates the negotiation, while belief revision both determines what beliefs are taken on and how the negotiation proceeds. In general, this kind of negotiation might take place between system and user, or between internal system agents (ie, Belkin et al's different functional experts).

Our initial very simple implementation of the information retrieval expert currently consists of four autonomous agents, each able to reason about its own and others beliefs. The basic agent architecture is illustrated in figure 2. The agents are heterogeneous in the sense that they all have their own areas on expertise and types of knowledge, but there are (currently) no hierarchical relationships between agents or controlling agents. The user may be considered a fifth agent, with distinct needs and expertise, though interaction with the user is always mediated by the last agent, the interactor. The different agents in the system are described below:

**Figure 2: Agent Architecture**

**The User-Attribute agent** is concerned with reasoning about the user's goals, status and expertise, and how these might influence the proposed information request. For example, rules might state that undergraduates are generally doing coursework, and people doing coursework generally want text books. The agent has general knowledge of user stereotypes.

**The Problem-Description agent** is concerned with filling out and refining the user's description of their problem. For example, rules might state that the user's overall subject is normally related to their academic department, while this overall subject can in turn be used to help complete a description of their problem. The agent has general knowledge of subject relationships.

**The Retrieval-Strategy agent** is concerned with determining an appropriate information search request, given knowledge of the user and their problem. For example, a rule might relate the user's preferred document types, and their overall subject, to a possible database to search. The agent has knowledge of document databases and search terms.

**The Interactor agent** is concerned with communication with the user. It maintains a model of the user's
assumed beliefs, and uses this in negotiation. The
agent has knowledge of how to update this model,
and will have knowledge of how to maintain a coher-
ent dialogue with the user.

The areas of expertise of the agents are overlapping,
and defined in terms of the knowledge types that each
agent can reason about. For example, the user-attribute
agent and the retrieval strategy agent both have knowl-
edge of, and can reason about, preferred document
types (e.g., journal articles). Each agent knows of the
areas of expertise of their fellow agents, and negoti-
ates within this area of overlap. Through negotiation,
the agents aim to agree on hypotheses in their over-
lapping areas of expertise. Negotiation is motivated
by apparent conflict, and at each stage agents reason
first about whether to take on communicated beliefs
of other agents, then (depending on the resulting be-
lieve state) may strategically plan utterances to achieve
belief changes in others.

Communication between agents takes place in a for-
mal language, with four message types currently de-
finite. These allow agents to inform agents of tentative
conclusions (tellref), to ask about some value (askref),
to ask for a justification (askwhy) and to provide a pack-
age of beliefs and rules as justification for some prior
assertion (tellwhy). The desired effects of these mes-
sage types is defined using simple planning operators,
and a simple planner invoked to construct utterances
given particular communicative goals. However, provi-
sional plans must be tested by predicting the effects of
the utterance(s) on the hearer's beliefs, using the belief
revision mechanism and given a provisional and partial
model of their prior beliefs (cf Appelt, 1985). If the sys-
tem predicts that a communication plan will fail, this
plan may be extended or alternative plans constructed.

Example

Our initial simple information retrieval expert allows
dialogues similar (in content and structure) to the one
above. The following example illustrates the basic ap-
proach.

At the beginning of our example the system has been
told that the user wants documents relating to material
after 1974. The system infers that there is therefore a
date restriction, and that the history database is there-
fore a possible choice. These beliefs and inference rules
are represented as follows, including their source infor-
mation (endorsements) and justifications.

System Beliefs:

1. doc-content(post-74) -> doc-restriction(date):
   [endorsement, 1cs]
2. doc-restriction(date) -> database(history):
   [endorsement, default]
3. doc-content(post-74) : [endorsement, 2cs]
4. doc-restriction(date) : [from 1,3]
5. database(history) : [from 1,2,3]

System's beliefs about user's beliefs:

6. believes(user, doc-content(post-74)) : [endorse-
   ment, 2cs]

Belief 3 is in both the problem-description (or PD)
module and the retrieval strategy module, beliefs 1,2,4
and 5 are in the retrieval strategy (or RS) module, and
belief 6 in the interactors (or I) module. However, in this
example we will not be concerned with the inter-agent
negotiation.

Initially the system, and in particular the retrieval
strategy, has the goal of convincing the user (via the
interactor module) that the history database may be
appropriate. The system predicts, based on its model
of the user's beliefs, that simply informing the user will
be sufficient. The belief is communicated weakly, since
it is not strongly committed to it, via the interactors:

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

The user responds by strongly rejecting this sugges-
tion:

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

The interactors now updates its model of the user's
beliefs with the information that the user does not be-
lieve the history database is appropriate. This belief is
passed on strongly to the retrieval strategy. However, as
the retrieval strategy does not believe that the interac-
tor (or user) is authoritative about preferred databases,
its does not strongly endorse the proposition, and as a
result of its belief revision process prefers to hold on to
its existing beliefs. But it recognises that there is now
a conflict between its beliefs and those of the interactors
(and hence user). This conflict motivates the retrieval
strategy to try to resolve that conflict by persuading the
user of its own point of view. It does this by provid-
ing support for its original suggestion in terms of the
following information:

RS: tellwhy(RS, I, database(history) because
doc-content(post-74) ->
doc-restriction(date) strong
& doc-restriction(date) ->
database(history) weak)

Note that inference rules may have two types of
endorsement, based on the source of the actual rule, and
the reliability of the rule's particular conclusions.

4We do not allow natural language input, requiring for
experimental purposes that the user inputs utterances in the
same formal language as that used by system agents.

5Note that inference rules are represented as beliefs just
as any others, and can be disbelieved in the face of new
evidence. As a simplification the example represents a de-
fault inference rule as having a default endorsement. In our
actual implementation the treatment of defaults is slightly
more complex, and inference rules may have two types of
endorsement, based on the source of the actual rule, and
the reliability of the rule's particular conclusions.
In other words, the system says that the reason why the history database is believed suitable is that wanting post-74 material is a kind of date restriction, and the history database is suitable if there is a date restriction. This justification is passed on by the interactor to the user, who decides to accept this suggestion, though with only weak commitment:

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

This example is obviously quite simple, but illustrates the basic belief preference and negotiation mechanisms that we are concerned with. The system has to convince the user that the history database is appropriate, by providing justifications which it believes will cause the change in beliefs. Communication is motivated by the user-system conflict. These basic mechanisms apply, moreover, both to user-system negotiation, as illustrated above, and to negotiation between system agents. The user is viewed as just another agent in the multi-agent system, though communications to the user are mediated by a special agent, the interactor.

**Ongoing Research: Focussing Agent Activity**

There are two interrelated problems with the current system. First, its behaviour (and the resulting dialogue) is unfocussed. There is no notion of the relevance of an inference or communication to the current goals of the system. Second, the belief revision process considers the whole set of an agent’s beliefs. This is both computationally expensive, unnecessary, and may result in unhelpful ‘distractions’ in an agents behaviour. We suggest that an agent reasons about preferences in a restricted subset of their beliefs. Ongoing research is concerned with addressing these two important and related issues.

In order to address the first issue (focus), we are developing a notion of goal preference, analogous to the belief preference heuristics. Given some knowledge goals (cf. Ram & Hunter, 1992), such as to agree on an appropriate literature database, the system constructs possible plans of action. Each action has a heuristic assessment of its likely outcome, based on the expected utility of the outcome, including the uncertainty of the outcome and effort involved. For example, agreeing on the database in the above example would have a high expected utility (as agreeing on something is viewed as more useful than coming to unilateral conclusions), but have an uncertain outcome and would involve moderate effort. Certain preferred goals and associated actions are focussed at any time. Given some new goals resulting from new knowledge states or communications, the system must decide whether the expected collective outcome of the currently focussed actions is preferred to the expected outcome of actions resulting from the new goal. If the two are equally preferred according to the heuristic assessments of outcome, then the notion of minimal change is used, and the system’s goal focus remains the same.

To support this style of reasoning our basic agent action cycle has the following stages (similar to Kiss et al, 1988). First, incoming messages are processed. The agent revises its beliefs based on new communicated information, and creates new goals based on requests from other agents (ie, askwhy and askref messages) and the new belief state (e.g., uncertainty or inter-agent conflict). Next, a particular knowledge goal is selected based on the heuristics mentioned above. A plan is constructed to (attempt to) achieve that goal, and executed until either completed or requiring a reply of some kind from another agent. This plan is then suspended, awaiting such a reply, and the action cycle started again. Unlike many systems, reasoning and planning continue while an agent awaits a reply.

In order to address the second, closely related issue we are developing a notion of belief relevance, with respect to current focussed beliefs. These focussed beliefs are based on the current knowledge goal, discussed above. For example, if a current knowledge goal is to agree on the history database as an appropriate literature source, then the belief in that database will be focussed. However, certain related beliefs are also clearly relevant, and must be considered in any process of belief revision. Previous work on focus in belief revision (e.g., Huang et al., 1991) has failed to address the issue of exactly what makes a belief relevant given some current agent activity, taking it for granted that the application would supply that relevant set. We are looking at what makes a belief relevant in the context of a negotiation.

We suggest that the relevant beliefs are those which, if they change their status (e.g., from believed to disbelieved), may influence the status of the focussed belief(s). Moreover, relevant beliefs are those which will most ‘easily’ produce a shift in status of the focussed belief. Thus, in the history database example, the uncertain inference that a ‘date restriction suggests the history database’ would be a relevant one. The system could easily disbelieve this inference (it is not strongly committed to it), and if disbelieved it would cause a change in the status of the belief in the history database.

We have implemented this notion of relevance as follows: The system searches the alternative consistent sets of beliefs to find the most preferred set(s) in which the focussed belief has changed status. Beliefs which have different statuses in the original and new preferred sets are those which could most easily cause a change in status in the focussed belief. These beliefs then, are the most relevant beliefs, given the negotiation context. This approach and algorithm has been implemented and tested for a small number of cases, and is effective at deriving appropriate relevant subsets of beliefs. Unfortunately, the overhead in constructing relevant sets for focussed beliefs is quite high. How-
ever, we expect that once the overall behaviour of the system is more focussed, the belief relevance mechanism will rarely need to be invoked ‘from scratch’, and techniques for changing the relevant set given only minor changes in the focus are being developed.

Summary

In this paper we have presented a framework for negotiation between autonomous agents, in the presence of conflict. This framework is based on a general theory of communication (Galliers, in press) and is appropriate for both negotiation between user and system and between system agents. Conflict is viewed as a positive force motivating useful communication, and conflict is resolved through negotiation, with each agent autonomously choosing what to believe. System-user negotiation, however, has to be mediated through a special agent, who has knowledge about other agents areas of expertise and how to update assumptions about the user’s knowledge. This particular agent will also require knowledge about how to manage a coherent dialogue, as the constraints on dialogue between system agents and between user and system are not the same.

We have focussed on the negotiation that occurs given inconsistent beliefs, to try and establish agreement. However, most of the ideas also apply to negotiations concerned with an agent’s goals and activities, where communications may result in changes in an agent’s intentions (cf. Cohen & Levesque, 1990; Grosz & Sidner, 1990). Agents should autonomously choose what to do next, based on prior wants and intentions and on new communications. In the face of (goal) conflict agents must be able to decide what they prefer to do (Galliers, 1989b). Intentions are therefore determined by preference between competing goals and activities.

The main ideas have been illustrated using a distributed information retrieval system which has to cooperate with the user in order to decide on appropriate ways of satisfying the user’s information need. In this domain, negotiation between system and user is vital, while individual system agents must be able to share their tentative conclusions and resolve any apparent conflicts. In our first prototype we showed how fragments of negotiations between a human user and information retrieval expert could be simulated. In our ongoing work we are extending the framework to allow more flexible goal driven and focussed behaviour, enabling more extended dialogues.

Our methodology in this work is to develop increasingly more complex prototype multi-agent systems, based on the theory of negotiation and belief revision. As more complex systems are developed, new issues emerge and have to be addressed. At each stage of the development we are empirically evaluating alternative architectures, with the first ‘round’ of evaluations discussed in (Cawsey, 1992b). We hope eventually to develop an information retrieval expert of sufficient complexity to demonstrate the effectiveness of both the belief revision theory and the overall theory of belief revision and communication.

References


