Revising Beliefs and Intentions: 
A Unified Framework for Agent Interaction*

Alison Cawsey‡, Julia Galliers, Brian Logan, Steven Reece 
and Karen Sparck Jones

Computer Laboratory, University of Cambridge 
New Museums Site, Pembroke St, Cambridge CB2 3QG, UK 
julia.galliers@cl.cam.ac.uk

Abstract. Intelligent agents, interacting with other agents in a changing and uncertain 
world, must be willing to revise their beliefs and intentions in response to new observed 
or communicated information. In this paper we present a unified framework for the 
revision of beliefs and intentions, extending that developed by Galliers [8]. The 
framework has been fully implemented, and uses a notion of belief focus to enable 
practical reasoning given resource bounds. It is being tested by modelling human 
cooperative dialogue fragments.

1. Introduction

An agent, acting and interacting in the world, must be able to revise its beliefs and intentions 
based on new information from the world and from other agents. If that information 
conflicts with its existing beliefs, the agent must be able to choose whether to believe the 
new information, and if so, how this should affect the agent’s other beliefs and intentions. 
The way beliefs are revised should depend on (at least) the authority of the source of any 
observed or communicated information, and on the justificatory connections between beliefs. 
The way intentions are revised must be integrally related to the belief revision process, as 
intentions are based on particular, revisable beliefs about the current state of the world.

In this paper we present a theoretical framework and computational model of belief and 
intention revision, based on a preference ordering over alternative, consistent and jointly 
realisable sets of beliefs and intentions. We extend and develop Galliers’ theory of belief 
revision in communication in which communication was viewed as both motivated by and 
causing belief revision in agents [8]. In particular, belief conflict was seen as a driving force, 
resulting in constructive negotiation and resultant belief change. However, as a result of 
computational implementation and testing, it become apparent that such a purely data- or 
conflict-driven account of communication is inadequate—an agent’s communications (and 
actions in general) should be driven by their intentions, with these intentions revisable in the 
context of changing beliefs.

*This research is supported by grant 90/CS42 of the UK Tri-Council Initiative on Cognitive Science and 
Human-Computer Interaction.

‡Current address: Department of Computer Science, University of Glasgow, Glasgow, UK.
Previous accounts of intention revision in the literature (e.g. [16, 18, 17]), although recognising the dependence of intentions on belief, have assumed different mechanisms for belief and intention revision. Our work, instead, views intention revision as an aspect of ‘belief’ revision in general, governed by considerations of consistency, coherence, and minimal change of the resulting beliefs and intentions. In addition, this previous work has approached the issue of resource-bounded reasoning by emphasising the stability of intentions, only revising when the expected utility of alternative actions is likely to be significantly greater than the current intended action. However, when revision does become necessary, all the agent’s beliefs and intentions are considered, and as a result revision is (potentially) computationally expensive. We instead treat resource boundedness in the context of belief revision as a whole, introducing a notion of focus which serves both to limit the number of possible intentions and reduces the number of beliefs that must be considered when revision is required. Agents maintain the consistency of beliefs and intentions in their current focus, while those outside this focus may be inconsistent.

Although our approach is applicable to any situation where agents are acting and interacting in the world, we are primarily concerned with modelling the communication between two or more interacting agents collaborating on some task. We consider both human-computer dialogues, where a human and a computer agent negotiate to solve some problem, and multi-agent systems, where a set of computer agents (each with their own beliefs and intentions) negotiate with each other. To test both of these cases we are developing a very simple information retrieval expert. The expert consists of a set of interacting agents each of which is responsible for solving a different aspect of the information retrieval task. In addition, the system as a whole must collaborate with the user, who has their own particular knowledge which is needed to solve the task. This test domain (and system) has been described in more detail in [4]: this paper focuses more on recent developments concerning the overall theoretical (and computational) framework.

In the rest of this paper we describe in more detail the basic theory of belief revision in communication; the extension of this to intention revision (illustrated by an example); and initial work on using focus to allow resource-bounded reasoning.

2. Background Theory: Autonomous Belief Revision in Communication

Gallier’s theory of belief revision in communication assumes that agents are fully autonomous, able to decide both what to believe and how to act [8]. Unlike most previous models of communication, (e.g., [15, 7, 1]), they do not automatically believe what they are told by other agents, or automatically hold on to their existing beliefs. Nor do they automatically adopt other agents’ recognised goals: agents choose what to believe and how to act, given all the information available to them. This autonomous choice will normally result in cooperative and constructive behaviour. However such cooperation is not the result of enforced benevolence or naive acceptance; rather agents cooperate unless they have good reason not to.

We now believe a theory of autonomous belief and intention revision is required if we are to model such autonomous, rational behaviour. The theory should describe how an agent determines what to believe and how to act in the face of conflicting beliefs and goals, possibly revising existing beliefs and intentions. In this paper, we take the first steps towards such a theory. In this section we describe the basic belief revision mechanism, and the extension to intention revision in the next section.

Our belief revision theory defines a preference ordering over alternative maximally consistent sets of beliefs (i.e., extensions), allowing an agent to choose which set(s) to
believe. (The alternative sets are obtained through an ATMS-type mechanism in which the labels represent justifications for beliefs [8]). The preference ordering is based on three principles:

1. **Authority of source:** Beliefs that are communicated, from other agents or from the world, or which are based on uncertain inferences, are variably hard to disbelieve. For example, direct observations are harder to disbelieve than information communicated from other agents, which are in turn harder to disbelieve than simple default inferences. We provide a heuristic description of the source of beliefs, in terms of endorsements assigned to them (see [6]), and use these endorsements when comparing the strengths of competing belief sets.

2. **Connectivity of Beliefs:** Belief sets that form a coherent, mutually supporting set are preferred to those that have less such connections. This notion is operationally defined relative to particular core or salient beliefs, which the agent is actively concerned with.

3. **Minimal change:** All things being equal, an agent will be conservative in its belief revision, i.e. will choose the belief set involving minimal change to its existing beliefs. In our model this also encompasses a notion of positive undermining [11], where an agent will only come to drop existing beliefs if there is some positive reason to disbelieve, and not simply if the initial justification goes away.

These principles are considered in order—first the most highly endorsed sets of beliefs are calculated; then, out of these, the most connected; and finally those which are closest to the agent’s prior beliefs. The result is a set of (equally) preferred belief sets: alternative preferred sets are likely as an agent’s information is normally incomplete, and its preference criteria deliberately course-grained. Propositions that are in each preferred set are viewed as ‘believed’ by the agent\(^1\); those which are in only some of the sets, with the negation of the proposition in others, are uncertain; and those whose negation is in every preferred set are disbelieved. (If a belief set does not contain some (currently focused) proposition, then it must contain its negation.) Agents can be variably committed to their different beliefs, depending on how hard they are to disbelieve. Our revision mechanism has been implemented, and is referred to as the *Increased Coherence Model* or ICM.

The revision mechanism is invoked not only when an agent receives new information, but also when planning communications for other agents. This involves predicting those utterances that will result in intended belief changes in the other agent, given a model of the other agent’s beliefs. For example, simply telling the other agent something may be insufficient to get them to believe it if the other agent holds conflicting prior beliefs. A justification may have to be given, such that the initial proposition, along with the justification, should cause the desired belief change (cf. [14, pp 94]).

As mentioned earlier, this basic framework for belief revision and communication is being tested by developing a very simple information retrieval expert which negotiates with the user concerning how to satisfy the user’s information need. This is a domain where both agents, user and expert, may frequently change their beliefs about each other, about the user’s need, and about appropriate documents or search strategies (see [4] for more details). Our initial prototype used the ICM for belief revision and utterance planning, but there was

\(^1\text{We have previously referred to these as pervasive beliefs in [4]. However, for convenience we will continue to use the term ‘belief’ to refer to any of the propositions in a belief set, whether strictly believed, disbelieved, or uncertain. We will flag the particular use of the term in situations where this could result in ambiguity.}
no explicit overall control of the negotiation—simple local rules were used to decide what
to do/say next depending on new communications and conflict situations [5]. However this
turned out to be inadequate: we now believe a goal driven approach is required, in which
the system can autonomously decide which goals to pursue based on changing beliefs and
requests from the user. The extensions to the framework developed to accommodate this
type of reasoning are discussed below.

3. Revising Intentions: Extending the Framework

We are working on the basis that the revision of intentions can be (and should be) managed
simply as part of the overall belief (or now, mental attitude) revision process. Different
propositional attitudes combine to make up an agent’s mental state, with derivational links
between beliefs, goals and predicted future states. An agent’s intentions are determined
via ‘belief’ revision, and defined as its preferred, jointly consistent and realisable goals.
Intentions are those goals ‘believed’ by the agent. Further, we constrain intentions so that
a state cannot be intended if the agent believes it already achieved, or that it will never be
achieved: by definition, such states result in inconsistencies.

The agent’s commitment to its intentions depends in part on its commitment to the beliefs
that ‘support’ that intention. However, commitment also depends on the importance of the
goal state, and the likelihood and difficulty of achieving the goal state. Here we borrow
from decision-theoretic approaches to action choice (e.g., [9]), though we will make no
assumptions about the availability of numerical estimators of utility. Instead, we extend our
notion of endorsements on beliefs to apply to intentions by including heuristic descriptions
of the utility of goal states and the effort required to perform the actions leading to that
state. New goals to achieve particular states are therefore assigned a description of the
expected utility of the intended state, while actions have a crude heuristic description of their
associated expected effort.

Planning, in our framework, involves using the normal inference mechanism to apply
goal generation rules, given existing beliefs, goals and planning operators, to derive new
goals. This is pretty much like standard planning, but the system will be continually
reassessing which mutually realisable sets of goals are currently preferred (these being the
agent’s current intentions). (As with beliefs, goals have ATMS-like labels representing their
justifications.) In general, goals may be conflicting (not mutually realisable) because of
competing temporal, physical or computational resources. Such goals will never be jointly
intended. However, preference also applies between different sets of goals which represent
alternative ways of (i.e., plans for) satisfying a higher level goal. In this case, the justificatory
links between goals are set so that if one alternative is chosen, the other will lose its support.
Subgoals in disregarded sub-plans are not strictly inconsistent with existing intentions, and
so may become preferred (and intended) given an independent reason for that subgoal to be
satisfied.

An agent’s intentions depend on the expected effort required to achieve its goals from
the current state, the utility of the goal, and the strengths of its associated beliefs about
the world. These different types of strength are all captured in the heuristic endorsements
associated with beliefs, goals and actions. An agent is most committed to, and hence will act
on, the action it finds hardest to ‘disbelieve’. Thus planning and acting are interleaved in the
framework, but we minimise the chance of an executed action turning out to be irrelevant
by starting with the most committed intended actions.

It is useful to evaluate our approach to intention revision in the light of Bratman’s desider-
ata [2]. By defining appropriate logical inference rules we ensure that an agent believes that
an intended action can eventually be achieved, and that it has not already been achieved—if either of these conditions fails to hold the intention will be dropped. Because of the basic belief revision mechanism, an agent’s intentions will be consistent with their beliefs, and minimal change ensures that intentions have some, though limited, stability. New intentions must be consistent with old ones—if they are preferred they force intention revision—and the system will attempt to determine ways of achieving existing intentions via its normal inference mechanisms. However, although most of this is consistent with Bratman’s characteristics, it is unclear whether our notion of minimal change provides sufficient stability of intentions (and possibly beliefs too), to ensure that, through the predictability of action, agent actions and interactions may be coordinated. We are therefore considering further ways to provide principled stability to intentions and beliefs.

4. Revising Intentions: Example

We have tested our extended framework by modelling short fragments of dialogues between users and librarians conducting searches of on-line databases. The transcript for one such fragment is reproduced below (taken from [3]):

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

In this example the user is looking for material on ‘Greek-Turkish relations after 1974’. The librarian has tentatively concluded that the Historical Abstracts (HA) database may be suitable because it is the only relevant database that 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. The 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. Below we outline how the system models the belief revision process underlying this dialogue fragment (the reasoning in the example has been simplified for presentation purposes).

The system’s initial communicative goal is to come to an agreement with the user concerning the appropriateness of the particular database, i.e.:\(^2\)

\[
\text{GOAL(system, AGREE(system, user, suitable-database(HA)))}
\]  

For the purposes of our example, we take this goal as given and intended. Ordinarily, it would be justified by the overall goal of coming to an agreed solution to the user’s information

---

\(^2\)Note that we use the term AGREE(A1,A2,P) to mean that the A1 believes P and believes A2 believes P. We currently make no commitments to issues relating to deeper levels of nested beliefs (cf. [7, 15]). A term such as GOAL(A,State), means that A has the goal to make State true. However, a goal may or may not be actually intended, depending on competing goals.
need; by the belief that deciding on a literature database is crucial to satisfying that goal; and by the belief that the HA database is a possibility.

From this initial intention, the system infers new goals and (potential) beliefs, and in particular, the goal to get the user to believe that the history database is appropriate:

\[ \text{GOAL(system, BEL(user, suitable-database(HA)))} \]  

This new goal is justified by goal (1) above; by the belief that the HA database is an appropriate choice; and by the belief that the user does not yet believe this. If any of these beliefs change the goal will no longer be justified, and may thus no longer be intended.

In the simplified model we have implemented, there are two alternative strategies available for getting the user to believe something: simply informing them; or persuading them (by providing additional justifications for the belief). The former strategy has the least expected effort, and is valid so long as the system does not have a positive reason to reject it. One reason for rejecting a goal to perform an action is if the system predicts that the action will not in fact achieve its target state. In this case, the system predicts the state resulting from the action:

\[ \text{GOAL(system, DO(system, INFORM(system, user, suitable-database(HA))))} \]  

(by comparing its ‘likelihood’ with conflicting states) and chooses, following belief revision, to believe that the action would in fact be successful. That is, it believes:

\[ \diamond \text{BEL(user, suitable-database(HA))} \]  

where \( \diamond \) is an operator denoting predicted belief, and as a result, this is the chosen intended action (with belief (4) having this goal as a justification). The system therefore outputs the utterance:

\[ \text{INFORM(system, user, suitable-database(HA))} \]  

Informing the user of its belief in the suitability of the HA database provides reason to believe (weakly) that the user will adopt this belief. The system therefore drops its intentions (1 and 2) as it weakly believes the goal state has been achieved.

However, in our example the user responds by rejecting the system’s suggestion:

\[ \text{INFORM(user, system, not suitable-database(HA))} \]  

The resulting conflict between what the system predicted the user’s belief state would be and the user’s indicated belief state causes it to reconsider its belief in the suitability of the HA database. However, its existing belief is well supported and the user is not believed to be authoritative concerning literature databases, so the system continues to believe in the suitability of the HA database. (In situations where the user’s opinion is considered authoritative, the system may choose to accept the user’s assertion or at least ask the user to justify its belief in the hope that the system will find the user’s argument convincing. If it does abandon its original belief, the belief revision process will also result in it abandoning any intentions and plans to achieve these intentions based on this belief.) Instead, the system abandons the strategy of simply informing the user that the HA database is suitable, as it no longer believes this will be effective.\(^3\) The alternative strategy of persuasion is now preferred resulting in the goal(s) of getting the user to believe the justification for the proposal, and in particular:

\[ \text{GOAL(system, BEL(user, restriction(date)))} \]  

\(^3\)This of course begs the question of why the user thinks the history database is not suitable.
\[ GOAL(\text{system, } \text{BEL(user, restriction(date) } \Rightarrow \text{ suitable-database(HA)}) \]  

However, as the system already believes that the user knows there is a date restriction on the content of the documents that they want, goal (8) alone becomes an intention, and results in the system outputting the ‘utterance’:

\[ \text{INFORM(system, user, restriction(date) } \Rightarrow \text{ suitable-database(HA)}) \]  

The system now believes (weakly) that the user believes the above rule, and that they therefore agree that the HA database is appropriate. It therefore drops its remaining intentions (1, 2 and 8). If the user replies confirming this view then its commitment to this view will be strengthened.

While this is a very simple example, it serves to illustrate the way in which belief and intention revision are combined in one process based on preference between alternative consistent sets. More complex dialogues can be modelled within the same framework.

For example, if the system has misunderstood the user’s intent regarding the date restriction, or if the user changes their mind about the date restriction and responds to utterance 3 in the original dialogue by stating that they want material written after 1974 rather than material dealing with events since 1974 (for instance, the user may be interested in the reinterpretation of historical events prompted by the invasion of Cyprus), the system will revise its beliefs about the user’s intentions. Thus given the input:

\[ \text{INFORM(user, system, not restriction(date))} \]  

the system comes to believe that the user believes that there is no date restriction on the documents the user requires. This leads it to abandon its own belief that there is a date restriction, that the HA database is suitable, and hence to abandon intention (1).

Similarly, suppose the user were to respond to utterance 3 in the dialogue by stating that they had searched the HA database on a previous occasion:

\[ \text{INFORM(user, system, already-searched(HA))} \]  

Again the system comes to believe that the user considers the HA database unsuitable (as a result of the rule already-searched(HA) \( \Rightarrow \) not suitable-database(HA)). Whether it adopts this belief itself depends on whether it can abandon its belief in the suitability of the database without giving up its belief in the date restriction. Belief revision results in (at least) two belief sets, one containing suitable-database(HA) and restriction(date) \( \Rightarrow \) suitable-database(HA), the other containing not suitable-database(HA) and already-searched(HA) \( \Rightarrow \) not suitable-database(HA). Which belief set is preferred depends on the relative endorsement of the two rules. If the former rule is less strongly endorsed the system will abandon its belief in the suitability of HA database, and will consequently abandon intention (1). These cases suggest that while a straightforward planning approach might seem adequate in simple examples like our first one, an approach based on belief revision in which beliefs and intentions are handled in an integrated way is more natural and thus more appropriate for complex situations.

However some relatively simple dialogues are beyond the system’s capabilities. Suppose, for example, that the user were to respond to the original utterance 3 by suggesting another database, say Current Affairs (CA):

\[ \text{INFORM(user, system, suitable-database(CA))} \]  

\(^4\)Note that the system will in fact abandon intention (1) in any case for the quite separate reason that, having exhausted the available negotiation strategies, it believes it to be no longer achievable.
The system will adopt the user’s suggestion if it has no reason not to (for example, it may believe that the CA database has been withdrawn or is not relevant to the user’s query). However it will also believe (weakly) that the user believes the historical abstracts database is suitable as a consequence of uttering (9) above. That is, the system fails to understand that the user’s suggestion of an alternative database is an elaboration of utterance 2 rather than the beginning of a new segment, and it falsely believes that it has been successful in convincing the user of the suitability of the HA database. Whether this false belief is recognised depends on how the dialogue develops.

However these problems arise because the system has little grasp of some aspects of dialogue pragmatics, like rhetorical structure, rather than as a result of the underlying belief revision framework itself. The unified framework for belief an intention revision described above underpins dialogue pragmatics as a whole. Intentions are chosen or dropped depending on predicted and actual belief changes, based on past and planned communications. Intentions change not just because an existing plan choice becomes invalid, but because it comes to be believed that they are not eventually realisable; because the generating intentions or beliefs are dropped; or simply because other conflicting intentions become preferred. The revision of beliefs and intentions underlies both planning and user modelling, in that it manipulates the information used by these processes. Achieving this within a conventional planning framework would effectively involve extending the planner to allow it to revise its beliefs about the world. Our framework accommodates all of these requirements, and provides a more comprehensive and thorough treatment of intention revision than the conventional planning approach.

5. Current Work: Reasoning in a Limited Focus

The main problem with the approach outlined so far is its computational complexity. Considering an agent’s every belief during revision is impractical. The only realistic solution to this problem is to work out ways of reasoning with a restricted set of focused beliefs, as those relevant to the agent’s current activities.

There has been some past work on using a notion of focus in belief revision and reason maintenance (e.g., [13, 12]). However, this work has been concerned with how to reason within focus sets, once these have been supplied by the problem solver. We are concerned more with how these sets should be determined, in our specific context of agent interaction and negotiation. We suggest that the beliefs which are relevant to an agent’s activities should be those which influence the agent’s intentions, and in particular its most committed intention(s)⁵.

In our current work we are exploring how to maintain such an appropriate and relevant set of focused beliefs. This involves recalling previous (possible) beliefs into the focused set, and ‘throwing out’ those which are irrelevant to the current situation.⁶ Both the recall and focussing processes involve a notion of relevance to the most committed intention(s), which we will refer to as the central intention(s). We keep a record of past focused belief sets, and use this record when recalling beliefs given a new central intention. Beliefs are initially recalled if they are relevant to beliefs (and goals) that are associated with the central intention in certain predefined ways. New inferences are drawn from the belief set

⁵This approach to belief focus relates to Groz and Sidner’s notion of intentional and attentional structure in discourse [10]. We would expect the discourse focus space in negotiation to be closely related to our belief focus set.

⁶In this section we use the term ‘belief’ to mean ‘possible beliefs’, i.e. propositions.
incorporating recalled beliefs, and the resulting set is 'pruned' by discarding beliefs that are not now considered directly relevant to the central intention. This results in a new focused set.

The crucial concept in all this is the notion of relevance to some central intention. In our context of action and interaction in an uncertain world, the relevant beliefs (and goals) are those which have most impact on a particular decision to act. An agent’s relevant beliefs and goals are those which can most easily change their ‘status’ (e.g., from believed/intended to disbelieved/not intended), and which if changed, would cause the agent to change the status of its central intention(s). The agent must therefore remain consciously aware of the status of these beliefs and goals, if it is to rationally decide when to change its central intention(s).

We have developed a straightforward algorithm for efficiently calculating these relevant beliefs, given the set of variably preferred belief/goal sets provided by the belief revision mechanism. First, we look for the most preferred of the given sets which contains the negation of the central intention. We then look for the beliefs and goals that have changed status between these two sets. It is these beliefs and goals which are most relevant to the status of the central intention, as it is most easy to disbelieve them, and hence revise the central intention.

We have implemented this algorithm and tested it using dialogues involving two or three utterances and a small number of beliefs through a single focus-recall cycle. However there are a number of problems which only become apparent with longer dialogues involving substantial refocusing and recall. In particular, there are issues concerning the details of the recall mechanism, the control of inferring within the focus, and the relation between these and the relevance mechanism. This is ongoing work.

6. Conclusion

It is crucial that an agent, acting and interacting in an uncertain and changing world, has the ability to revise its beliefs about the world (and other agents) and its intentions to act. Revisions of beliefs and intentions are integrally related, so are best considered together, as aspects of a single process of change of mental states. This revision process must take into account both the source of ground-level beliefs and intentions (i.e., the authority of the belief source, or the utility of the intended state) and the connectivity of the whole network of beliefs and intentions. We have developed a framework for such belief/intention revision, which has been computationally implemented and tested in the domain of cooperative dialogue modelling. An agent engaged in a cooperative dialogue will be frequently revising beliefs based on statements from the other agent, and revising communicative intentions based on requests from that agent and its own changing beliefs. Belief/intention revision is therefore crucial in a principled model of such dialogues.

It is of course impractical for an agent to consider all their beliefs when revising. The agent must reason about a restricted subset of currently critical beliefs. We have developed a measure of relevance to the intentions currently at issue, as influencing immediately future actions. We maintain a set of beliefs currently in focus, by recalling associated beliefs from past foci, and (following limited inference) throwing out irrelevant beliefs. This aspect is still ongoing work.

The main contributions of this work are: a) in establishing a unified general framework for autonomous belief/intention revision which involves both the consistency of alternative sets, and a principled measure of preference between competing sets; b) in demonstrating this in the domain of cooperative dialogue modelling, where a consideration of belief/intention revision is vital for modelling behaviour which is cooperative, but not naively compliant.
References