# Towards Interest-Based Negotiation

Iyad Rahwan, Liz Sonenberg
Dept. of Information Systems
University of Melbourne
Parkville 3010, Australia
i.rahwan@pgrad.unimelb.edu.au
l.sonenberg@dis.unimelb.edu.au

Frank Dignum
Intelligent Systems Group
Institute of Information and Computing Sciences
Utrecht University
3508 TB Utrecht, The Netherlands
dignum@cs.uu.nl

## ABSTRACT

Negotiation is essential in settings where agents have conflicting interests and a desire to cooperate. In many approaches, agents are assumed to have pre-set, fixed preferences, and complete awareness of the space of possible outcomes. Such strong conditions are often not satisfied. In this paper, we argue that since preferences are adopted to pursue particular goals, one agent may influence another agent's preferences by discussing the underlying motivations and interests behind adopting the associated goals. We identify concepts that seem essential for supporting this type of dialogue. In particular we demonstrate that arguing about beliefs needs to be complemented by arguing about goals, and we begin an analysis of dialogue moves involving goals.

## Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*intelligent agents, multiagent systems, languages and structures*

## General Terms

Design, Economics, Languages, Theory

## Keywords

Negotiation, Persuasion, Argumentation, Preferences, Dialogue Games, Interaction Protocols

## 1. INTRODUCTION

In multi-agent applications, autonomous components often need to interact with one another. Interaction may be needed for many reasons: seeking information, inquiry, persuasion, negotiation, and deliberation [16]. Negotiation is usually seen as a type of interaction in which agents seek agreement on the division of some scarce resources, while each agent tries to maximize its share or utility. Various frameworks have been proposed to facilitate the automation of this activity [12]. Most existing frameworks assume that agents have complete, fixed and pre-set utilities and preferences. However, consumer preferences are typically fluid and only finalised in the course of the transaction itself.

We base our intuition about consumer decisions on ideas in consumer behaviour modelling [13]. Current theories view consumer perception of products in terms of several attributes. Individual consumers vary as to which attributes they consider most relevant. Moreover, consumers' beliefs or perceptions may vary from the "true" attributes because of consumers' particular experiences and the way they gather and process information. This means that consumers may make uninformed decisions based on false or incomplete information. It also means that different consumers might choose the same product for different reasons. Consequently, consumer preferences are shaped and changed as a result of the interaction with potential sellers, and perhaps with other people of potential influence such as family members or other consumers. Game-theoretic and traditional economic mechanisms have no way to represent such interaction.

Consider the following dialogue between a used car seller $S$ and a potential buyer $B$ who wants to purchase a station wagon.[1]

B: *Can't you give me this wagon a bit cheaper?*

S: *Sorry Sir, that's the best I can do. Why don't you go for a sedan instead?.*

B: *I have a big family and I need a big car.*

S: *Modern sedans are becoming very spacious and would easily fit in a big family.*

B: *I didn't know that, let's also look at sedans then.*

During this negotiation dialogue, the seller was able to persuade the buyer to consider sedans after understanding one of the reasons he initially preferred wagons.

Emphasis on discovering the underlying interests has also been of major importance in the literature on interest-based negotiation among humans or organisations [7]. This body of work advocates the advantage of focusing on *interests* rather than on *positions*. It is argued that by understanding the reasons behind positions, we can redefine the problem in terms of the underlying interests. By discussing these interests, we are more likely to reach a mutually acceptable agreement. Suppose two kids are fighting over a banana.

[1]To avoid ambiguity, we shall use "he/his" to refer to buyers or consumers, while using "she/her" to refer to sellers.


*AAMAS'03,* July 14–18, 2003, Melbourne, Australia.

One of them might actually want to eat it, while the other naughty one wants the peel in order to get the teacher to slip over. By exchanging these reasons, the two kids might reach an agreement that was not initially possible.

This paper advances the state of the art of automated negotiation in two important ways. Firstly, it introduces the first formal account of *interest-based negotiation*, in which one agent may influence another agent's preferences by discussing the underlying motivations for adopting the associated goals. To achieve that, it presents the fundamental structures and needed in the decision model. Secondly, the paper provides an account of the types of arguments that can be presented against a particular goal and how they can influence the agent's adopted goals, and consequently its preferences. It then introduces a set of *dialogue moves* that facilitate this type of interaction and demonstrates these moves through an example discourse between a car seller and a potential buyer.

The paper is organised as follows. In the next section we discuss the fundamental differences between arguing about goals and beliefs. In section 3 we show which types of supports exist for goals and how they lead to different goal preferences. Section 4 discusses different ways to attack goal supports and subsequent changes in goal preferences. In section 5 we indicate some locutions and dialogue rules which are exemplified in section 6. Some areas for future research are described in section 7.

# 2. ARGUING ABOUT GOALS VS. ARGUING ABOUT BELIEFS

It has been proposed that frameworks and mechanisms for argumentation can be used to support negotiation [2, 14]. One observation about such frameworks is that they are based on concepts in argumentation about beliefs [3, 6], which are about the establishment of truth, as opposed to arguing about goals, which is about the establishment of choice. The reason behind this, we suggest, is that most of the early work on argumentation was aimed at applications in legal reasoning, where dialogue aims at reaching an undefeated truth [9], or medical reasoning, where dialogue aims at reaching the most reasonable diagnosis [8]. In trading environments, however, the objective is the *fulfillment of needs*. In what follows we demonstrate, by analysis of a number of examples, that arguing about goals requires a different approach to arguing about beliefs.

We use $H \vdash p$ to denote that the *minimal* set of beliefs $H$ logically implies proposition $p$ (i.e. using some consequence relation $\vdash$ we can construct a *tentative* proof for $p$ by using elements of $H$). $H$ is called the *support* of the argument, and $p$ its *conclusion*. An undercutting argument is an argument for the negation of an element of $H$; a rebutting argument is an argument for the negation of the conclusion $p$.

Now we identify some key differences between goals and beliefs that shape the later analysis.

## 2.1 Informational vs. Motivational

First, note a fundamental difference in the nature of beliefs and goals. Beliefs are informational and therefore, in principle, can be checked against the current world for their correctness. Goals, on the other hand, are motivational and therefore by nature intrinsic to an agent. The correctness of a goal does not follow from the state of the world, but from the attitudes of the agent only. So, one might be able to establish that another agent's belief is not correct; but can only attempt to establish that a goal seems unachievable, not useful, unsupported, etc. but not incorrect.

## 2.2 Objective vs. Subjective Resolution

As noted earlier, when people argue about belief they aim at reaching the truth, which is somewhat objective and independent from what the participants initially believe. Moreover, as conflicts among beliefs arise, the resolution is usually performed by comparing the structures of the supporting arguments for the conflicting claims [6] or using a preference relation over the elements of the support (i.e. the content as opposed to the structure) [1]. However, when we deal with goals, an agent might adopt a goal or a subgoal for purely subjective reasons, and might be unwilling (and should not be expected) to drop that goal based on rational arguments in the sense usually seen with beliefs.

If a car is clearly red, it would be difficult (and hopefully impossible) for an agent to maintain a belief that the car is green. However, it would make sense to allow a buyer to maintain the goal of getting a green car, and reject any opposing argument on the basis of subjective judgement alone. After all, colour taste is a personal matter.

## 2.3 Argumentation Objective

The objective behind the argumentation process is a major difference between arguing about beliefs and goals. When arguing about beliefs, the main goal of the discourse is to persuade the other agent that its belief cannot be supported. However, when arguing about goals, an agent attempts to persuade another of adopting alternative goals. To illustrate this difference, consider the following dialogue.

B: *I believe Emirates airlines offers a free stopover because it is the holiday season.*

which the seller might attempt to undercut:

S: *Actually it is not the holiday season because it is July in Australia.*

Clearly, the seller's goal is to persuade the buyer to drop or reverse his belief. She is effectively saying "I think you should not believe $x$ anymore because of such and such". Now, consider the following example, which deals with goals.

B: *I would like to fly with Emirates airlines because this means I will pay little money and arrive quickly.*

The travel agent might attempt to convince the customer to fly with Qantas airlines instead. To do that, she might show one downside of choosing Emirates:

S: *But flying with Emirates means changing your flight arrangements isn't very flexible.*

or she might attempt to make Qantas more appealing:

S: *But if you fly with Qantas instead, you get cheap upgrade to business class.*

or she might correct a misconception:

S: *Actually flying with Emirates can be quite expensive if you factor the insurance cost.*

Obviously, it is in the seller's interest to convince the buyer to fly with Qantas (e.g. she gets a higher commission). This is what motivates the argumentation process.

## 2.4 Derivability vs. Instrumentality

The nature of the relationship between the support and the conclusion of an argument differs when arguing about goals and beliefs is related to . In belief argumentation, one belief is *derivable* from a set of beliefs if we can construct a tentative proof that takes the set of beliefs as premises and derives the new belief as a conclusion using logical implication. On the other hand, one goal cannot logically imply another goal. But a goal can be said to be *instrumental* in achieving another goal (or set of goals). Moreover, there are many reasons for choosing or not-choosing a goal, such as as feasibility, ethical considerations, costs/benefits, etc. These dimensions are relevant to choosing goals, but not beliefs.

## 2.5 Support vs. Purpose

The above distinction also leads to an important issue relating to the direction of reasoning about beliefs and goals. With goals we cannot only argue about what is instrumental to achieve a goal, but also about the superior goals that this goal achieves. For example, to prevent an agent from achieving a goal $g = buyTicket$, we could look at the set of subgoals $SubG = \{goToAgency, payMoney\}$ that need to be fulfilled to achieve $g$, written $achieve(SubG, g)$. If one of these subgoals is prevented (say there is not enough money), then the goal $g$ is prevented. Another way is to look at what supergoals $SuperG = \{goSydney\}$ our goal $g$ contributes to (or is instrumental for), written $instr(g, SuperG)$. If we can persuade the agent that goals in $SuperG$ are not worth achieving (say the conference in Sydney was cancelled), the agent will have no reason to pursue $g$ anymore (i.e. the agent would no longer have a reason to attempt to achieve the goal of buying a ticket since the main goal behind that, namely going to Sydney, is no longer a goal). The reasoning in the two cases goes exactly in the opposite directions.

With beliefs, on the other hand, reasoning occurs in one direction, namely from the premises (i.e. initial beliefs) to the conclusion (i.e. the new belief). The reason behind this difference is that beliefs are not purposeful. One cannot form a belief only for the purpose of forming other beliefs, but it is natural to form a goal for the sake of achieving other superior, more fundamental goals. This distinction is crucial to the way arguments are formed for beliefs and goals.

Another important point is that when arguing about beliefs, one cannot attack the relation between premises and conclusions imposed by the consequence relation $\vdash$. This is because the relation is defined in terms of the rules of inference, which are fixed and usually agreed upon. In goals, however, one should be able to attack the assumption that one goal is instrumental to achieving another goal.

## 2.6 Inconsistent Beliefs vs. Alternative Goals

Another main difference between arguing about beliefs and goals is in the way conflicts are defined. This further shows that existing frameworks for arguing over beliefs cannot be directly adapted for dealing with goals.

When arguing about beliefs, if we have $H \vdash a$ and $H \vdash b$, then $a$ and $b$ are necessarily consistent (because they both follow logically from the same set of premises). If $a$ and $b$ happen to be inconsistent, then there is something fundamentally wrong in our reasoning process. On the other hand, we might have two goals supported by identical reasons, be conflicting in some sense, yet pose no problem. Consider the following two statements by a customer.

B: *I would like to fly with* ***Emirates*** *because they are cheap and offer good service.*

B: *I would like to fly with* ***Qantas*** *because they are cheap and offer good service.*

Even though the support of both arguments is identical (one might fly with Emirates or Qantas for the same reasons), and the conclusions are in a sense inconsistent (one cannot adopt the goals of flying with Emirates and Qantas at the same time), we do not encounter a problem as with beliefs. The goals of flying with Emirates and Qantas are seen as "alternatives" for achieving the superior goals of paying less and getting good service. This shows how dependencies between goals and subgoals are different from those between conclusions and premises.

# 3. AGENTS AND GOAL SUPPORT

We now define the components of the agents and discuss the relevant elements of the goal support, i.e. what it is that makes an agent adopt a particular goal, and capture them in a more formal fashion. We discuss the types of supports independently and only link them informally to argumentation. We shall leave the discussion about how exactly they can be attacked or defended to the next section.

*Definition 1.* We define an agent $i$ to be a tuple:
$\langle KB_i, IG_i, AG_i, Cap_i, Role_i \rangle$
where $KB_i$ stands for the beliefs of the agent, $IG_i$ stands for the set of intrinsic goals of the agent (it's desires), $AG_i$ stands for the set of adopted goals of the agent (or it's intentions), $Cap_i$ is the set of the capabilities of the agent and $Role_i$ is the role that agent $i$ enacts.

We will not get into a formal definition of roles but suffice to say that roles can be defined by their goals, norms, interaction rules and relationships to other roles. We only use $RoleG(r)$ here to denote the set of intrinsic goals of the role $r$. See [5] for more details on agents and roles.

We assume that an agent is introspective and therefore the knowledge base contains beliefs about it's own goals, capabilities, role and the relations between them.

*Definition 2.* The *knowledge base* for agent $i$ denoted $KB_i$ consists of the following:

1. A (possibly inconsistent) set $Bels_i$ of belief formulae defined using $Bels$ (the set of all possible beliefs). These are called *basic beliefs* to distinguish them from other types of beliefs involving relations among beliefs and goals.
2. A set $IGoals_i \subseteq Goals$ of intrinsic goals the agent is aware of. (where $Goals$ is the set of all possible goals).
3. A set of statements of the form $justify(B, g)$ where $B \subseteq Bels_i$ and $g \in Goals$.
4. A set of statements of the form $achieve(SubG, g)$ where $SubG \subseteq Goals \cup Cap$ and $g \in Goals$.
5. A set of statements of the form $instr(g, g')$ where $g, g' \in Goals$.
6. A set of statements of the form $conflict(g, g')$ that explicitly denote pairs of conflicting goals, i.e. goals that cannot be achieved simultaneously.[2]

[2]Note that conflicts between goals might involve more subtle dependencies, and may not in fact be represented using an explicit relation. We therefore assume that conflicts can be detected through a separate mechanism (e.g., [15]).

7. A set of statements of the form $altGoal(g', g'')$ such that $g'$ and $g''$ are top level *intrinsic* goals (i.e. $\nexists x$ where $instr(g', x)$ or $instr(g'', x)$). $g'$ and $g''$ are viable alternatives (i.e. either of them suffices).

8. A role $Role_i$ that the agent plays.

As can be seen from the above definition, each goal is supported by different types of beliefs. Each goal has some purpose(s) and some justification(s). The purpose of a goal can either be to achieve a superior, more fundamental goal or it can be an intrinsic goal of the agent. In the latter case the purpose follows from the role the agent plays (or the social relation between the agents). For example, an agent that adopts the role of a buyer in a travel agency has the intrinsic goal of buying a ticket. The reason behind the adoption of this goal can be extracted from the reason of the agent to adopt this role.

The justification of a goal also comes from different sources. We follow Habermas [10] to distinguish three spheres of justification, the subjective, the objective and the normative one. In the subjective sphere the justification of a goal comes from the fact that the agent believes that the goal can be achieved (it has a plan that it believes is achievable). In the objective sphere the justification comes from beliefs that justify the existence of the goal. So, the agent believes that the world is in a state that warrants the existence of this goal. Finally, justifications from the normative sphere come from the social position and relations of the agent. E.g. as a director of a company, an agent is responsible for increasing profits, and also for the well-being of his/her employees.

Formally we define the support of a goal as follows:

*Definition 3.* $H = (SuperG, r, B, SubG)$ supports goal $g$, denoted as $support(H, g)$ for agent $i$ if

- $SuperG$ is a set of goals such that $\forall x \in SuperG$, $instr(g, x) \in KB_i$,
- $g \in AG_i$,
- $B$ is a set of beliefs such that $justify(B, g) \in KB_i$
- $SubG$ is a set of goals such that $achieve(SubG, g) \in KB_i$.

A *goal argument* is a tuple $\langle H : g \rangle$ where $g$ is a goal and $support(H, g)$.

Note that, contrary to an argument for a belief, the goal does not "logically" follow from the support! To make this distinction between a goal argument and a belief argument clear we shall use $\Vdash$ (rather than $\vdash$) to represent the support relation. So we shall write the above relation as:

$$(SuperG, r, B, SubG) \Vdash g$$

In the following subsections we will briefly discuss each of the elements that support the adoption of a goal.

## 3.1 Goals and Beliefs

Consumer goals are often linked to their beliefs. For example, a person with the goal of purchasing a big car might base that goal on the belief that he has a big family. A person with the goal of travelling to Sydney might base that on a belief that an important conference will be held there. The beliefs can be seen as the *context* in which the goal holds, and we can say that the beliefs *justify* the goal. If the context turns out to be unsatisfied, the agent would no longer have a reason to keep its goal. We denote the relation between a goal $g$ and a (possibly empty) set of beliefs $B$ that form its justification or context by $justify(B, g)$.

This relation between beliefs and goals is closely related to the notion of conditional goals as introduced by Cohen and Levesque [4]. However, the notion of justification is stronger than that of conditions for goals. The latter are only effective when a goal already exists. The goal will be dropped when the condition is no longer valid. Simplistically (but illustrative) this could be modelled as: $\neg b \implies \neg g$. Justifying beliefs, however, have a causal link to the goal to the effect that they influence the adoption of the goal. This could be modelled as: $b \implies g$.

## 3.2 Goals and Subgoals

Another important factor that influences the adoption of a goal is the set of resources (or subgoals) that are needed to achieve that goal. This idea was initially explored in an argumentation setting in [14] and later in [2] where the conclusion is an intention and the support is a tentative plan for achieving that intention (the plan may also contain assumptions about beliefs). They propose the use of existing techniques in belief argumentation and give it different meaning. For example, an undercutting argument (i.e. an attack on the support) means that one of the resources in the support is not available. A rebutting argument means the other agent has a conflicting intention.

Here is an example of the goal/plan relationship. In order to achieve the goal of going to Sydney, an agent might have to purchase a ticket and sort out accommodation. If, after checking his account balance, the agent discovers he does not have sufficient funds, he can no longer buy a ticket and must drop the goal of going to Sydney (unless an alternative plan is found, say by borrowing from a friend). We use $achieve(SubG, g)$ to denote the relation between a goal $g$ and the set $SubG$ of resources (or subgoals) that need to be acquired (or achieved) in order for $g$ to be achieved. If $SubG$ is defeated, $g$ is no longer achievable and must be dropped.

## 3.3 Goals and Supergoals

Consumers often adopt particular *non-basic* goals because they believe these help them achieve their more basic, superior goals. For example, a car buyer might adopt a goal of getting a car with airbags because that helps him achieve a more basic goal of being safe. Similarly, an academic might adopt the goal of going to Sydney in order to fulfill the more fundamental goal of presenting a research paper. If the goal of presenting the paper is dropped, its subgoal of travelling to Sydney must also be dropped. Note that the goal of going to Sydney is not assumed to be neither necessary nor sufficient, but rather only *instrumental* to achieving the supergoal (which is a weaker statement). To present a paper, one also needs to prepare the presentation slides, pay the registration fee, and so on.

We use the relation $instr(g, SuperG)$ to capture the fact that the achievement of goal $g$ is instrumental towards the achievement of a non-empty set of superior goals $SuperG$.[3] Instrumentality means that the goal belongs to (or belong to a sub-plan of) a plan that achieves the supergoal. Formally:

$instr(g, SuperG)$ iff for each $s \in SuperG$, there is a set of

[3]Note that the "*instr*" is weaker than the "*achieve*" relation. The latter is a sufficient plan for fulfilling a goal, while the former is only contributing to the goal.

goals $X$ such that either (i) $g \in X$ and $achieve(X, s)$; or (ii) $g \in X'$ and $achieve(X', s')$ and $instr(s', s)$.

The above example can now be represented as follows:

$$instr(goSyd, presentPaper)$$
$$achieve(\{goSyd, prepareSlides, payFee\}, presentPaper)$$

Moreover, finding an alternative means for achieving the supergoal (say presenting at a local conference) might also, in a sense, weaken the initial goal. For example, suppose an alternative plan was presented that achieves the goal of presenting a paper by attending a conference in Perth.

$$achieve(\{goPerth, prepareSlides, payFee\}, presentPaper)$$

The existence of this alternative plan can potentially cause the agent to drop the goal *goSyd* because it is no longer essential for achieving the supergoal. Which plan the agent selects (and hence, which goals it ends up adopting) depends on a comparison between the alternative plans. This might be based on the cost of adopting the plan. In our example, travelling to Sydney costs more money, and would hence be a less preferred plan. We assume that a decision mechanism exists that allows agents to perform such comparisons.

## 3.4 Goals and roles

Finally, justification of a goal can follow from the role an agent plays. Some goals are intrinsic to certain roles and therefore performing the role is in itself a justification for adopting the goal. For example, the travel agent has an intrinsic goal of selling flights to customers.

Sometimes the justification does not follow from the role itself, but rather from the combination/relation of the role with some other roles. For example, a parent might justify behaviour (like telling the child to shut up) towards a child just by stating that he/she is the parent. The same behaviour could not be justified like that towards a partner.

Although roles play an important part in justifying goals in more complicated dialogues, we will not use them in this paper. First, because adequate formalisation would need more accurate formalisation of the basic social concepts involved like "power", "rights", etc. Secondly, the dialogues in which these concepts play a role are very complicated and do not occur in the simple examples explored so far.[4]

## 3.5 Adopting a goal

Having discussed the different supports for goals, we are now in a position to define the mechanism for agents to adopt goals.

In cases of conflict, the agent must be able to choose between conflicting goals. Moreover, the agent needs to be able to choose among alternative plans towards achieving the same goal (or set of goals). The fact that two alternative plans exist for the same goal can be represented by having two statements $achieve(SubG_1, g)$ and $achieve(SubG_2, g)$. So, the agent needs a mechanism that allows it to generate, from its knowledge base, the set of goals $AG$ that it attempts to achieve.

The details of such mechanism are outside the scope of our discussion, so we view it as a black box. However, we assume that the agent attempts to maximise its utility, by considering the costs of adopting different plans as well as the utilities of the goals achieved. One way such a mechanism might be realised is by using influence diagrams [11] to model the different types of support for a goal as probabilistic influences. The goal(s) that are selected are those that have the "strongest" support:

*Definition 4.* A *goal selection mechanism* for agent $i$, denoted $GSM_i$ takes the agent's knowledge base $KB_i$ and returns a set $AG_i$ of *adopted goals* such that $\forall g \in AG_i, \exists H : support(H, g) \wedge \forall g' \notin AG_i, \forall H' : support(H', g') : H \geq H'$

where $H \geq H'$ indicates that the support for $g$ is stronger than the support for $g'$ according to the criteria given above.

Now that we have defined the different supports of the agent's goals and how the networks of support for goals determine the preferred (and thus adopted) goals, we turn to the ways that the preferred goals can be changed by attacking the supports.

# 4. HOW TO ATTACK A GOAL

In this section, we show the different ways in which one can attack a goal, causing it to be dropped, replaced, or even causing additional goals to be adopted. Consider the following goal argument:

$$\langle(\{presentPaper\}, \{confInSyd\},$$
$$\{buyTicket, arrangeAccomm\}) : goSyd\rangle$$

This goal argument intuitively means that the agent has the goal of going to Sydney because he believes there will be a conference there, and he would like to present a paper there. In order to achieve that goal, the agent needs to buy a ticket and arrange for accommodation.

## 4.1 Attacking Beliefs

There are a number of attacks an agent might pose on the beliefs in a goal argument of another agent. An agent might state any of the following:

1. **Attack:** $\neg b$ where $b \in B$
   An agent can attempt to disqualify a goal by attacking its context condition. This is usually an assertion of the negation followed by an argument (i.e. a tentative proof) supporting that assertion. Following our example, another agent might say that the conference in Sydney has actually been cancelled, written $confCancelled \vdash \neg confInSyd$.

   **Effect:** This type of attack triggers an argumentation process similar to the one that is purely belief based, since agents argue about whether proposition $b$ holds. An existing model such as [1] might be used to facilitate that. If the attack succeeds, then $Bels_i = Bels_i - \{b\}$ and link $justify(B, g)$ ceases to exist. Consequently, the goal must be dropped unless it has another justification.

2. **Attack:** $\neg justify(B, g)$
   An agent might attempt to attack the link between the belief and the goal. The opponent might say that having a conference in Sydney does not really justify going there, written

   $$\neg justify(confInSyd, goSyd)$$

[4]Consequently, in the remainder of this paper, we shall not consider the use of roles in arguments. Arguments will therefore take the form $\langle(SuperG, B, SubG) : g\rangle$.

The justification relation between a belief and a goal is not one that can be objectively tested. Whether such argument is accepted should be therefore based on the social relations in the agent society. For example, the head of department usually has the authority to refuse such justification. It might be that the head of department also requires a paper to be accepted in the conference in order to agree that *goSyd* is justified.

**Effect:** $justify(B, g)$ is removed from $KB_i$, and the goal gets dropped unless it has a different justification.

3. **Attack:** $justify(B', g')$ where $B' \subseteq B$ and $conflict(g, g')$
An agent might present another goal that is also justified by the same set of beliefs, where the two goals are conflicting. In our example, the opponent might argue that the conference in Sydney also justifies the goal of helping students prepare their presentations. Note that this type of attack requires the opponent to know about what other goals the set of beliefs justify. This information might have been acquired from the other agent in an earlier stage in the dialogue, from previous encounters, or be part of the domain knowledge. Or it might be something that the agent can impose by using its social authority.

   **Effect:** The success of this attack adds $justify(B', g')$ (and $conflict(g, g')$ if it is not already in) to $KB_i$ and hence requires the agent to make a decision about which goal is more important, $g$ or $g'$. If $g'$ is more preferred, then $g$ must be dropped. If helping students conflicts with the goal of going to Sydney, then the agent must drop one of the two goals.

## 4.2 Attacking Subgoals

Similarly, an agent might attack the subgoals that are thought to achieve $g$. This might be done in a number of ways:

1. **Attack:** $\neg p$ where $p \in SubG$:
The opponent attacks an element of the subgoals by arguing that it is unachievable. This might be because the associated resources are not available or that there are no successful plans for achieving the subgoal. In our example, the opponent might argue that the agent cannot buy a ticket due to lack of funding.

   **Effect:** Attacking subgoals means that the achievability is undermined. As long as some subgoals remain, which together can achieve the goal, it is still potentially adopted. So to defend against this attack, the agent is required to provide an alternative plan for achieving $p$, or else drop the goal.[5] If all of these plans were defeated eventually, the goal must be dropped. This would cause all $achieve(X, p)$ statements to be removed from $KB_i$.

2. **Attack:** $\neg achieve(SubG, g)$
Here, the opponent attacks the relation between the subgoals and the goal in question. In our example, this would be done by arguing that buying a ticket and arranging accommodation are not sufficient for going to Sydney (say there are other things that need to be done as well).

   **Effect:** $achieve(SubG, g)$ is removed from the knowledge base. If no alternative plan is found, the goal must be dropped.

3. **Attack:** $achieve(P, g')$ where $P \subseteq SubG$ and $g'' \in IG \cup AG \wedge conflict(g', g'')$
In this case, the opponent argues that by executing (part of) the support, another adopted goal becomes unachievable. The opponent might argue that by buying a ticket, the agent would spend too much money and would no longer be able to, say, buy the proceedings. If buying the proceedings is a goal of the opponent, then there is conflict.

   **Effect:** $achieve(P, g')$ is added to $KB_i$. This attack again triggers a comparison between the conflicting goals and the more preferred would be chosen. Another possibility is to find an alternative plan that does not clash with the other goal, formally, $SubG'$ such that $P \nsubseteq SubG'$ and $achieves(SubG', g)$.

## 4.3 Attacking Supergoals

These attacks can be as follows:

1. **Attack:** $\neg instr(g, g')$ where $g' \in SuperG$:
A goal argument might be attacked by arguing against the instrumentality link between it and the supergoals. In our example, the opponent might argue that going to Sydney is not actually instrumental to presenting the paper, written

$$\neg instr(goSyd, presentPaper)$$

   **Effect:** In defense, the agent might either present a plan $P$ where $g \in P$ and $achieve(P, g')$, i.e. to show a plan involving this goal that achieves the supergoal. Otherwise, if authority does not suffice to win, the agent must remove $instr(g, g')$ from $KB_i$, which might weaken the goal and cause it to be dropped.

2. **Attack:** Show set of goals $P$ such that $achieve(P, g')$ where $g' \in SuperG$ and $g \notin P$:
Here, we show an alternative plan which achieves the supergoal but does not include $g$. The opponent might say that going to Perth (instead of Sydney) is also instrumental towards presenting the paper, written

$$achieve(\{goPerth, prepareSlides\}, presesntPaper)$$

   **Effect:** $achieve(P, g')$ is added to $KB_i$. The agent compares the plan $P$ with the existing plan for achieving $g'$, and based on the outcome of this comparison, $g$ might be dropped (with the whole plan to which it belongs).

# 5. DIALOGUES ABOUT GOALS

As discussed above, while beliefs are supported by sets of other beliefs, goals are supported by different elements. An important consequence of this difference is that the dialogue games also get more complicated. When arguing about beliefs the arguments are more or less symmetric. Both agents can attack and defend their beliefs by putting forth other beliefs, which in their turn can be attacked in the same

[5]Note that even in the case where an alternative plan is found, the preference can be weakened, and therefore an alternative goal might become preferred.

way. This leads to a kind of recursive definition of dialogue games, which is always terminated at the level of the knowledge bases of the agents. In the case of dialogues over goals, this does not always hold. However, some similarities arise within the different types of justifications. The existence of a goal can be justified by the existence of another (super)goal. The same attacks can be made to the supergoal as were tried on the original goal. In this way one might traverse the whole tree of goals until an intrinsic goal of the agent is reached. The same can be done downwards for tracing subgoals. On the other hand, when beliefs that justify a goal are attacked one might get into a "classical" argument on those beliefs. What remains different, however, are the possible ways an agent can defend itself against an attack.

## Locutions and Dialogue Rules

We now define a number of locutions that agents can use in the dialogue. Due to space limitations, we shall not provide a complete account of these locutions and the rules that govern them. Instead, we present each locution with an informal description of its meaning. We also do not include locutions that are less relevant, such as those used in opening, closing and withdrawing from the dialogue.

1. ASSERT($b$) The agent asserts a belief $b$ which might be either a propositional belief formula or a relational statement such as $justify(.)$, $achieve(.)$ and so on. The agent must believe the statement uttered. The other agent can respond with ACCEPT($b$), ASSERT($\neg b$) or with CHALLENGE($b$). If $b$ was a relational statement, the other agent's response depends on the different methods of attack as described in section 4. For example, the opponent may assert the opposite and resort to social authority to resolve the conflict.

2. CHALLENGE($b$) One party may not agree with an assertion made by another party. By uttering this locution, an agent asks another for the reasons behind making a belief assertion $b$. This must be followed by an assertion of a (possibly empty) set of beliefs $H$ denoting $b$'s support such that $H \vdash b$. In case $b$ is a relational statement (which we assume for the moment not to have a justification itself) or is a basic belief it can only be answered by asserting social authority.

3. ACCEPT($b$) allows an agent to explicitly acknowledge a belief formula.

4. PROPOSE($g$) is the locution that allows an agent to assert it wants to achieve $g$ and would like the counterparty to adopt its part of achieving $g$. The goal might denote actions to be performed or resources to be exchanged. This locution can be followed by acceptance, rejection, or request of the support elements of $g$. It also can be followed by a counter proposal.

5. REQ-JUST($g$) allows one agent to ask for the justification of a goal proposed by another agent. This must be followed by an assertion involving a $justify(B, g)$ statement.

6. REQ-ACHIEVE($g$) allows one agent to ask another about how it believes it can achieve the goal $g$. This must be followed by an assertion involving an $achieve(G, g)$ statement.

7. REQ-PURPOSE($g$) allows an agent to ask for what supergoal (if any) the goal $g$ is thought to be instrumental towards. This must be followed by an assertion involving an $instr(g, g')$ statement.

8. ACCEPT($g$) allows an agent to accept a proposed goal as part of the final deal.

9. REJECT($g$) allows an agent to reject an offer.

10. PASS allows an agent to pass a turn by saying nothing.

The inclusion of the PASS locution makes it possible to avoid strict turn taking in the dialogue. Especially after a rejection, a party may wish to continue to ask for the reasons or supports of a proposal. Also the party might want to deliver alternatives for the original proposal. Due to space limitations, however, we will not discuss these issues further.

# 6. AN EXAMPLE

In this section, we present a more complex example which involves using some of the above locutions. This demonstrates how it may be used to capture interest-based negotiation dialogues.

*Example 1.* Consider the following dialogue between a buyer and a seller that was presented in section 1:

B: PROPOSE($wagon$)

S: ACCEPT($wagon$)

B: PROPOSE($10K$)

S: REJECT($10K$)

B: PASS

S: REQ-PURPOSE($wagon$)

B: ASSERT($instr(wagon, bigCar)$)

S: ASSERT($instr(\{sedan, modern\}, bigCar)$)

B: PROPOSE($sedan$)

S: ACCEPT($sedan$)

B: PROPOSE($10K$)

S: ACCEPT($10K$)

After rejecting a proposal to sell the wagon for $10,000, the seller asks the buyer for the supergoal that the choice of wagon was based on. The buyer asserts that he wants a wagon because it achieves the goal of having a big car. The seller asserts that modern sedans also achieve that. A deal was facilitated that would not have been possible without arguing about the supergoal.

An alternative strategy the seller could adopt would be to challenge the adoption of the superior goal $bigCar$ itself.

S: REQ-JUST($bigCar$)

B: ASSERT($justify(bigFamily, bigCar)$)

S: ASSERT($justify(young, \neg bigCar)$)

Here, the seller asks for the belief that justifies the supergoal $bigCar$. She finds out that the buyer bases this on the belief that he has a big family. It is very difficult to argue with someone about the size of their family. So the seller chooses a different route, stating that the buyer seems young, and that being young justifies not getting a big car (say because smaller cars are more cool). Now if the buyer agrees on the statements $young$ and the relation $justify(young, \neg bigCar)$ and if he believes it is stronger than $justify(bigFamily, bigCar)$ then he would accept the argument. Now there is one less reason to get a wagon, and it could cause him to request a sedan instead.

# 7. DISCUSSION AND CONCLUSIONS

In this paper, we sought to work with negotiation scenarios in which agent preferences are not predetermined or fixed. We argued that since preferences are adopted to pursue particular goals, one agent might influence another agent's preferences by discussing the underlying motivations for adopting the associated goals. We demonstrated how this process differs from arguing about beliefs, hence distinguishing our work from other argumentation based approaches. We described the main concepts that support the selection of a particular goal and used it to show the various ways in which goal selection might be attacked. We then presented a set of dialogue moves to support this type of interaction and demonstrated an example discourse that makes use of them.

Throughout the paper, we presented scenarios between buyers and sellers. However, we believe our framework would also be applicable to a wide range of distributive bargaining problems, such as resource allocation and labour union negotiations. We also believe our framework can be extended to deal with more than two agents.

It is important to note that in situations where agents are familiar with the domain and have complete and fixed utilities, there is less incentive to share the underlying reasons behind their choices. In fact, it may be that hiding information (or the true valuation of outcomes) from the opponent can give the agent an advantage. This is often the case in auctions, for example. Our work, on the other hand, concentrates on situations where the agents' limited knowledge of the domain and each other makes it essential for them to be, in a sense, cooperative. In other words, there appears to be some kind of tension between the willingness to provide information about the underlying motives (which can potentially help improve the outcome), and the tendency to hide such information for strategic reasons. We take the position that in some settings where agents have incomplete information, sharing the available information may be more beneficial than hiding it.

In the future, we intend to provide a full account of the dialogue game protocol and analyse some of its properties, such as termination, success, complexity, and so on. We also intend to explore the different strategies that an agent can adopt in an interest-based negotiation dialogue. For example, one agent might prefer to explore the instrumentality of a goal, while another might attempt to argue about its achievability. Another important question is how strategy selection can be guided by the agent's beliefs about the opponent or the domain. For example, an agent that has a strong belief that a certain goal cannot be achieved would rather attack the subgoals instead of the supergoals.

# 8. ACKNOWLEDGMENTS

The authors acknowledge the support of a University of Melbourne International Collaborative Research Grant in the development of this work. The first author is very grateful for motivating discussions with Peter McBurney and for useful feedback from Nick Jennings.

# 9. REFERENCES

[1] L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1-3):197–215, 2002.

[2] L. Amgoud, S. Parsons, and N. Maudet. Arguments, dialogue, and negotiation. In W. Horn, editor, *Proceedings of ECAI2000*, pages 338–342. IOS Press, 2000.

[3] C. I. Chesñevar, A. Maguitman, and R. Loui. Logical Models of Argument. *ACM Computing Surveys*, 32(4):337–383, Dec 2000.

[4] P. Cohen and H. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261, 1990.

[5] M. Dastani, V. Dignum, and F. Dignum. Organizations and normative agents. In *Proceedings of first EurAsia conference on advances in ICT*. Springer-Verlag, 2002.

[6] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.

[7] R. Fisher and W. Ury. *Getting to Yes: Negotiating Agreement Without Giving In*. Penguin Books, New York, NY 10014, USA, 1983.

[8] J. Fox and S. Parsons. Arguing about beliefs and actions. In A. Hunter and S. Parsons, editors, *Applications of Uncertainty Formalisms*, number 1455 in LNCS, pages 266–302. Springer Verlag, 1998.

[9] T. F. Gordon. The pleadings game: formalizing procedural justice. In *Proceedings of the fourth international conference on Artificial intelligence and law*, pages 10–19. ACM Press, 1993.

[10] J. Habermas. *Theorie des kommunikativen Handelns*. Suhrkamp, Germany, 1981.

[11] F. V. Jensen. *Bayesian Networks and Decision Graphs*. Springer-Verlag, New York, 2001.

[12] S. Kraus. *Strategic Negotiation in Multi-Agent Environments*. MIT Press, Cambridge, USA, 2001.

[13] G. L. Lilien, P. Kotler, and S. K. Moorthy. *Marketing Models*. Prentice-Hall, USA, 1992.

[14] S. Parsons, C. Sierra, and N. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3):261–292, 1998.

[15] J. Thangarajah, M. Winikoff, L. Padgham, and K. Fischer. Avoiding resource conflicts in intelligent agents. In F. van Harmelen, editor, *Proceedings of the 15th European Conference on Artificial Intelligence (ECAI 2002)*, Amsterdam, 2002. IOS Press.

[16] D. N. Walton and E. C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. SUNY Press, Albany, NY, USA, 1995.