A logic-based approach to conflict resolution

Robert Kowalski
Imperial College London
April 2003
Revised May 2003

Abstract

Many real world conflicts can be understood in terms of logical inconsistencies between goals. In this paper, I present an approach to conflict resolution that unifies logic, goal-reduction and condition-action rules in a cognitive model of intelligent agent. The approach uses goal hierarchies to reconcile goal conflicts by finding alternative, logically consistent ways of solving higher-level goals. It also incorporates the use of decision theory to decide between different solutions, in the attempt to optimise their expected utility.

To illustrate and test the power and generality of the approach, I investigate two applications: the prisoner's dilemma and the Agha-Malley proposed solution of the Israeli-Palestinian conflict.

Introduction

It is usual to think of conflicts in game-theoretic terms. Two or more players need to decide between different actions, with different expected outcomes. Each player needs to decide what to do without knowing what the other players will do. The outcome depends, not only on the player's own decision, but also on the unknown and uncertain decisions of the other players. Each player tries to maximize the expected utility of the outcome for itself.

In zero-sum games, a win for one player is a loss for the other. In other games, if the players co-operate, the outcome can be a win for everyone. Viewed in such game-theoretic terms, conflict resolution can be understood as transforming a competitive win-loose game into a co-operative one in which everyone wins.

For example, it might be possible to transform a conflict between you and me, in which I want to have the cake and you want to eat it, into a co-operation in which I have half the cake and you eat the other half. The original game is a conflict, which only one of us can win. The new game is a co-operation, in which we attempt to maximize the combined utility of the outcome, subject to some constraint on the minimal utility of our individual outcomes.

But conflict can also be viewed in logical terms, as an inconsistency between different goals. Such conflicts can arise both when a single individual is in two, contrary minds, as well as when two individuals are contrary minded. Thus, I can be in conflict with myself

if I want to have the cake and also want to eat it; and I can be in conflict with you if I want to have the cake and you want to eat it.

Viewed this way, conflicts can be avoided in different ways. One of the easiest ways is by timesharing, say by having the cake today and eating it tomorrow. Or by generalizing one or both of the goals and satisfying them in a different way, say by having the cake and eating some other 100% equivalent dessert instead.

In this paper, I will outline an approach to conflict resolution that combines the game-theoretic and the logical views of conflict. From Game Theory, I will borrow the strategy of choosing between alternative actions by attempting to maximize their expected utility. From logic, I will borrow the use of a cognitive model in which problem-solving is viewed as reducing goals to sub-goals and in which conflicts can sometimes be avoided by solving higher-level goals in alternative ways.

Thus, to reconcile a conflict between two goals, we first try to identify what higher-level goals the individual goals are intended to achieve. For example, I might want to have the cake simply to enjoy its appearance; and a person (you or I) might want to eat the cake simply to eat dessert.

Next, we consider other ways of achieving these more basic, higher-level goals. For example, I could enjoy the appearance of the cake by taking a photograph of it, framing the photograph, and hanging the photograph on the wall; and, you (or I) could eat chocolate, ice cream or apple pie instead.

Perhaps some of these alternatives have other advantages. Hanging a photograph of the cake on the wall, for example, would allow me to enjoy the appearance of the cake for longer than if I simply kept the cake itself until it rotted. On the other hand, you (or I) might enjoy eating chocolate even more than eating cake.

Of course, we also need to consider possible disadvantages. Perhaps eating any kind of dessert will harm our diet.

Having identified the various higher-level goals that might be positively or negatively affected by the alternatives, we need to estimate the different degrees to which the alternatives will satisfy these higher-level goals. Finally, we need to choose a combination of alternatives that we expect to satisfy the higher-level goals to the greatest extent.

If the original conflict was a conflict between two different individuals, then we may need to further constrain our solution so that each individual's higher-level goals are adequately satisfied to some minimal extent.

Of course, performing these estimates and calculations is only a normative ideal. In practice, some other, simpler, heuristic technique, which approximates the ideal, will have to be used instead.

In summary:

To reconcile a conflict between two inconsistent goals,
find higher-level goals that do not conflict,
find alternative consistent ways of satisfying those higher-level goals,
infer positive and negative consequences of those alternatives,
estimate the utilities of those consequences,
estimate the probabilities of any unknown and uncertain conditions in the alternatives,
choose alternatives that come as close as possible to maximising the expected utility of those consequences.

This approach to conflict resolution - by identifying alternative ways of satisfying higher level goals – is similar to one proposed by van Lamsweerde et al (1998) for managing conflicts in the requirements engineering stage of software development. Requirements engineering is the first stage in the software development process, in which the requirements of different stakeholders are elicited, conflicts between them are reconciled, and a consistent software specification is constructed.

The combination of logic for thinking with decision theory for deciding is similar to Poole's (1997) Independent Choice Logic, which combines logic programming and decision theory. He applies his Logic to problems in diagnosis, robot control, multi-media presentation and user modelling.

The approach is also compatible with the informal characterisation of thinking presented by Jonathan Baron (1994) in his textbook, "Thinking and Deciding", where on page 4 he writes:

> "Thinking about actions, beliefs and personal goals can all be described in terms of a common framework, which asserts that thinking consists of *search* and *inference*. We search for certain objects and then make inferences from and about the objects we have found."

In our logic-based model, *search* is performed by means of backward reasoning, and *inference* by forward reasoning. The objects that are found are solutions to higher-level goals. Like Baron, we use decision theoretic concepts of utility and probability to decide between different solutions.

In addition to proposing a general, logic-based approach to conflict analysis and resolution, I will consider two examples in greater detail: the prisoner's dilemma and the Agha-Malley (Agha et al 2002) proposed solution of the Israeli-Palestinian conflict. After presenting a brief introduction to these two examples, I will review some of the main cognitive models that have been developed in Cognitive Science and discuss their suitability for conflict analysis and resolution. I will then present a logic-based cognitive model, which attempts to unify the best features of the other models, and which can be combined with the decision-theoretic approach. Finally, I will show in greater detail how the logic-based model can be applied to the prisoner's dilemma and to the Agha-Malley proposal.

The Prisoner's Dilemma

The police have arrested two people suspected of committing a crime. They have only enough evidence to convict both suspects of a lesser offence, unless one of the prisoners turns state witness against the other. The prisoners are taken into separate rooms and offered the following deal:

If one turns state witness against the other, but the other one doesn't turn state witness, then the state witness gets 0 years in jail and the other one gets 4 years in jail.
If both turn state witness against the other, then both get 3 years in jail.
If neither turns state witness against the other, then both get 1 year in jail.

The Prisoner's Dilemma is the problem of deciding which of the two actions (turning state witness or not turning state witness) to choose. The problem is complicated by the fact that neither prisoner knows which action the other prisoner will choose.

According to the norms of decision theory, each prisoner should choose the action that maximises the expected utility of the outcome, minimizing the number of years the prisoner will expect to spend in jail.

Consider the simple case in which the first prisoner judges that the second prisoner is as likely to turn state witness as not. Then the expected utility for the first prisoner of turning state witness is:

the probability of the second prisoner not turning state witness (= .5)
multiplied by the utility of the resulting outcome for the first prisoner (= 0 year) +

the probability of the second prisoner turning state witness (= .5)
multiplied by the utility of the resulting outcome for the first prisoner (= 3 years)

= .5×0 + .5×3 = 1.5 year.

Similarly, the expected utility for the first prisoner of not turning state witness = .5×4 + .5×1 = 2.5 years. So, of the two alternative actions, turning state witness maximises the first prisoner's expected utility.

The decision-theoretic calculation can be done for other assumptions about the probabilities of the second prisoner's choice of actions. It can also be done for a different, common goal:

> minimise the expected total number of years in jail, $Y_1 + Y_2$,
> where the first prisoner gets Y_1 years in jail
> and the second prisoner gets Y_2 years in jail.

The calculation is even simpler for the common goal than it is for the selfish goal, because it eliminates the complication of uncertainty. The result of the new calculation is

that neither of the prisoners should turn state witness, and then both of them will get only 1 year in jail, which is better for each prisoner than the selfish solution.

Later, we will see how the Prisoner's Dilemma can be viewed in logical terms.

The Agha-Malley proposed solution of the Israeli-Palestine conflict

In "The Last Negotiation", Hussein Agha and Robert Malley (2002) outline a proposal for a solution of the Israeli-Palestinian conflict. Their proposal is based on earlier proposals, including ones put forward during the Camp David negotiations in 2000-2001. What is most striking about their approach is the way in which it is presented: starting with the "basic interests" of the two sides, translating them into "policy redlines", and then proposing solutions to the resulting "issues". Equally striking is the fact that the authors make no attempt to justify or attack any of the goals and beliefs they attribute to the two sides.

Israel's basic interests, as presented by Agha and Malley, can be summarised as:

1 Preserving the Jewish character and majority in Israel
2 Achieving security
3 Achieving international recognition and normalcy
4 Controlling Jewish holy sites and national symbols and
5 Ending the conflict with Palestinians and Arab States.

The Palestinian basic interests can be summarised as:

1 Living in freedom, dignity, equality and security
2 Ending the occupation and achieving national self-determination
3 Resolving the refugee issue fairly
4 Controlling Muslim and Christian holy sites and
5 Ensuring any solution is accepted by the Arab and Muslim worlds.

Agha and Malley present their proposed solution in the context of five main issues:

1 The territorial issue.
2 Security.
3 Jerusalem.
4 Haram al-Sharif, or Temple Mount.
5 Palestinian refugees.

1Territorial issue. For the territorial issue of the boundaries of the Israeli and Palestinian states, they propose a solution based on land swaps, the purposes of which include the physical continuity of the two States, and transferring the majority of the Jewish settlement areas in the West Bank to Israel.

2 Security. The proposed solution includes the non-militarization of the new Palestine State and the initial presence of an international force to keep the peace.

3 Jerusalem. The proposal is based on demographic and religious self-governance. As they put it, "what is Jewish…should become the capital of Israel, and what is Arab should become the capital of Palestine."

4 Haram al-Sharif, or Temple Mount. The proposal is based on "practical arrangements required to meet both sides' needs."

5 Palestinian refugees. Agha and Malley characterise this as "perhaps the most vexing topic of all". Their proposal is to offer the refugees resettlement in Arab-populated areas of Israel, to include these areas in the land swap with Palestine, and to provide generous financial compensation to the refugees.

My purpose in discussing the Agha-Malley proposal is not to defend or criticise its details, but rather to analyse the argument in terms of the general conflict resolution methodology of this paper. I will return to this analysis, after developing the necessary background.

Overview of agent models

The conflict analysis and resolution framework that I present in this paper is based upon cognitive models that have been developed in Cognitive Science and Artificial Intelligence. In particular, it is based upon a cognitive model that attempts to unify and reconcile several otherwise conflicting approaches. In this "unified model", (Kowalski et al 1999, Kowalski 2001) problem-solving is viewed as a goal-oriented, logic-based activity, in which decisions need to be made between alternative actions.

The alternative actions are solutions of higher-level goals. They can also be thought of as goals in their own right, but at the lowest level of a goal hierarchy. A potential conflict between different actions, or more generally between two different goals, can sometimes be avoided by finding alternative ways of solving goals higher in the hierarchy. In this paper, to decide between different courses of action, I propose combining the unified cognitive model with the decision-theoretic approach used in Decision Theory and Game Theory.

The unified, logic-based cognitive model attempts to combine the best features of production-systems, BDI agent models and goal hierarchies.

Production systems

The production system model (Post 1943, Newell 1973) is undoubtedly the most important and most successful computational model of human intelligence.

Production systems are collections of **condition-action rules** (also called **production rules**), which have the form:

If **conditions** then **do actions**. e.g.

If someone attacks you then attack them back.

Condition-action rules are written in the form of implications. This is consistent with their intended use: to reason forward, when the conditions hold, to derive the actions that are the conclusion of the rule. However, unlike ordinary logical implications, they are procedural rather than declarative sentences, because their conclusions are written in the imperative rather than in the declarative mood.

Production systems embed condition-action rules in an observation-thought-action cycle:

> **To cycle**,
> **observe**,
> **think**,
> **decide** what actions to perform,
> **act,**
> **cycle** again.

Thinking is a form of forward reasoning, which matches the conditions of the rules with statements in a working memory and derives the actions of the rules as candidates for execution. The working memory represents either the agent's own internal short-term memory or its external sensory input. The actions are either internal actions on the short-term memory or external actions on the environment.

If the conditions of several rules are simultaneously satisfied by statements in the working memory, and their corresponding actions are incompatible, then some form of conflict resolution is needed to decide between the alternatives.

The conflict resolution strategies used in production systems are generally very simple. Most often they are determined by assigning different priorities to different rules. For example, the rule

> If someone attacks you,
> then attack them back.

might be assigned a higher priority than the rule

> If it's raining,
> then put up an umbrella.

Such priorities are used to give precedence to actions derived from rules having higher priority over actions derived from rules having lower priority. In theory, however, more sophisticated decision-theoretic strategies that aim to minimise the expected costs and maximize the expected benefits can also be used. In this example, they would no doubt give the same result.

A cycle of violence

Suppose two agents have the same rule:

> If someone attacks you, then attack them back.

and suppose that one of the agents attacks the other. The result is obvious – an unending cycle of violence – or else a fight to the end, until one of the agents can no longer fight back. An agent whose entire behaviour is governed by a fixed set of such rules would be unable to escape the inevitable consequences.

However, because production systems are so simple, it is relatively easy to augment them with a learning module, which modifies the rules in response to any rewards or punishments received from the environment. In our example, such an adaptive production system might learn the new set of rules:

If someone attacks you and they are weaker than you,
then attack them back.

If someone attacks you and they are stronger than you,
then run away.

Or alternatively: If someone attacks you,
then find out what caused the attack and eliminate the cause.

Goals in production systems

Production systems are ideally suited for modelling stimulus-response behaviour, in which goals are implicit rather than explicit. Any appearance that the rules achieve any goals is an emergent property of the rules, which results from their successful adaptation to feedback from the environment.

For example, none of the rules given above for reacting to an attack have any explicit representation of their purpose. Presumably, their implicit purpose is the immediate goal of self-defence, for the ultimate, higher-level goal of survival.

Production systems are not limited to the implementation of stimulus-response rules alone. They can include explicit goals, both as conditions of rules and as components of working memory. For example, they can contain such rules as:

If someone attacks you and the goal is to defend yourself, then attack them back.

The relationship between production rules and explicit goals is a complicated one, which we will address at several places in the remainder of this paper. But first, we turn our attention to alternative cognitive models in which goals are a primary feature.

BDI agents

Production systems represent an agent's behaviour procedurally, as a collection of condition-action rules. In contrast, BDI logics (Cohen et al 1991, Rao et al 1992) represent an agent's Beliefs, Desires and Intentions declaratively by means of declarative statements. For example:

Belief: You defend yourself,
if whenever someone attacks you, you attack them back.

Desire: You defend yourself.

Similarly:

Belief: You eat dessert
if you eat cake or you eat chocolate
or you eat ice cream or you eat apple pie.

Desire: You eat dessert.

Intention: You eat cake.

The advantage of declarative statements is that they have well defined meanings, namely whether they are true or false. In contrast, procedural statements, such as condition-action rules, do not.

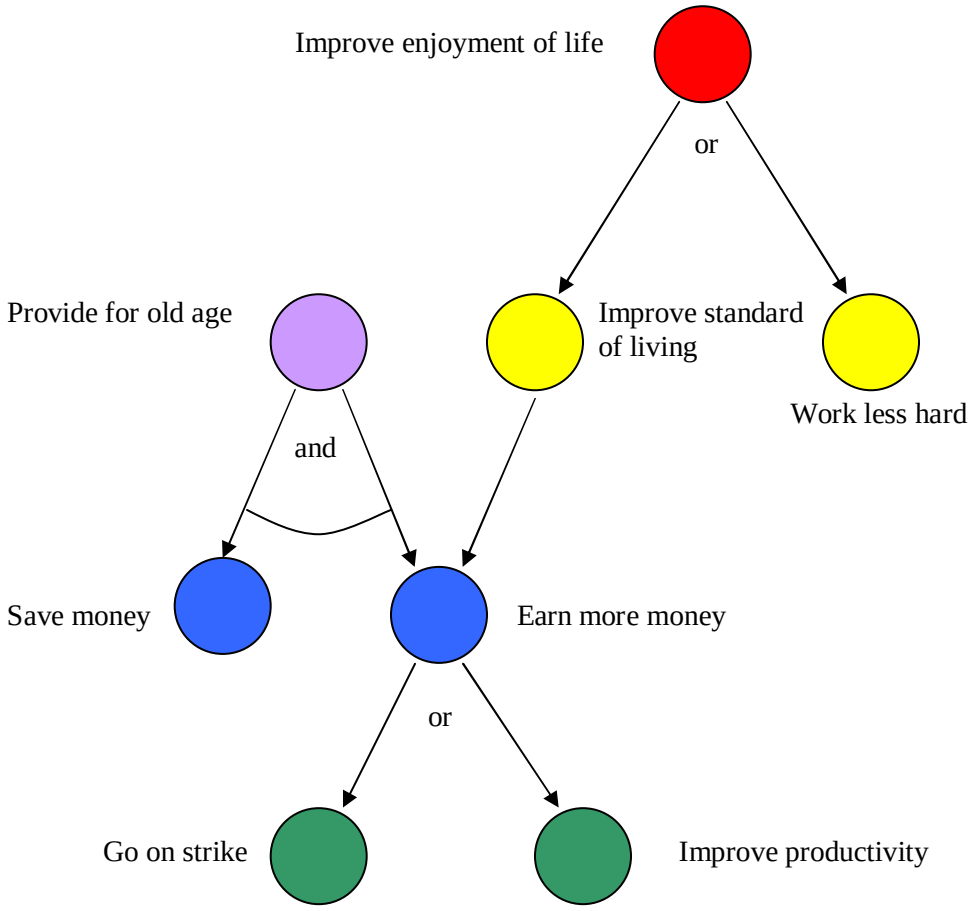
The disadvantage of declarative statements is that it can be difficult to see how to use them to regulate an agent's behaviour. BDI logics, in particular, are usually formalised as modal logics, and these logics are notoriously difficult to implement efficiently. Moreover, many critics of formal logic argue that people do not think in logical terms; and therefore BDI and other logics are unsuitable as formal models of human thinking.

The unified agent model, which I present later in the paper, aims to reconcile the declarative and procedural representations. But first, as a bridge between the two kinds of representation, we will investigate another, alternative representation of an agent's mental state, in terms of its hierarchy of goals.

Goal hierarchies

Instead of viewing an agent's goals as implicit and emergent, or as on the same level as its beliefs and intentions, we can view the agent's goals as the main component of its mental state. Viewed in this way, an agent's goals and sub-goals form a hierarchy, which can be depicted as an and-or tree. One of the earliest uses of and-or trees was by James Slagle (1961) to implement a computer program to solve symbolic integration problems in freshman calculus.

And-or trees are drawn upside-down, so that higher-level goals are higher in the tree, and lower-level goals are lower in the tree. An arc directed from a higher-level node down to a lower-level node represents the (partial) reduction of the higher-level goal to the lower-level sub-goal. Nodes at the bottom of the tree represent irreducible action goals, which can be solved only by performing them successfully. For example:



In fact, as can readily be seen, the tree is actually a more complex graph-like structure, which reflects the fact that a goal, such as earning more money, can contribute to the achievement of several higher-level goals, such as improving your living standard as well as helping to provide for old age.

The and-or tree/graph structure displays the hierarchical relationships between goals and sub-goals. In this case, it shows that going on strike is a sub-goal of earning more money, which is a sub-goal of improving your living standard, which is a sub-goal, in turn, of improving your enjoyment of life.

However, because of the graph structure of goals and sub-goals, the relationship between them is not always rigidly hierarchical. For example, the two goals of improving enjoyment of life and of providing for old age, in the graph above, are at the same level, and are not related to one another in a strictly hierarchical manner.

In addition to displaying the hierarchical structure of goals and sub-goals, and-or trees/graphs show alternative ways of trying to solve a goal. For example, the and-or graph above shows that going on strike and increasing productivity are alternative ways of trying to increase your pay.

Similarly, improving your standard of living and working less hard are alternative ways of trying to improve your enjoyment of life. And-or trees/graphs also show that several conjoint sub-goals might be needed to solve a higher level goal. For example, to provide for old age, not only do you need to earn more money, but you need to save it as well.
And-or trees/graphs clarify the distinctions between goals and sub-goals, alternative sub-goals and conjoint sub-goals. In every day life, it is easy to get them confused. They also clarify the distinction between search spaces and search strategies.

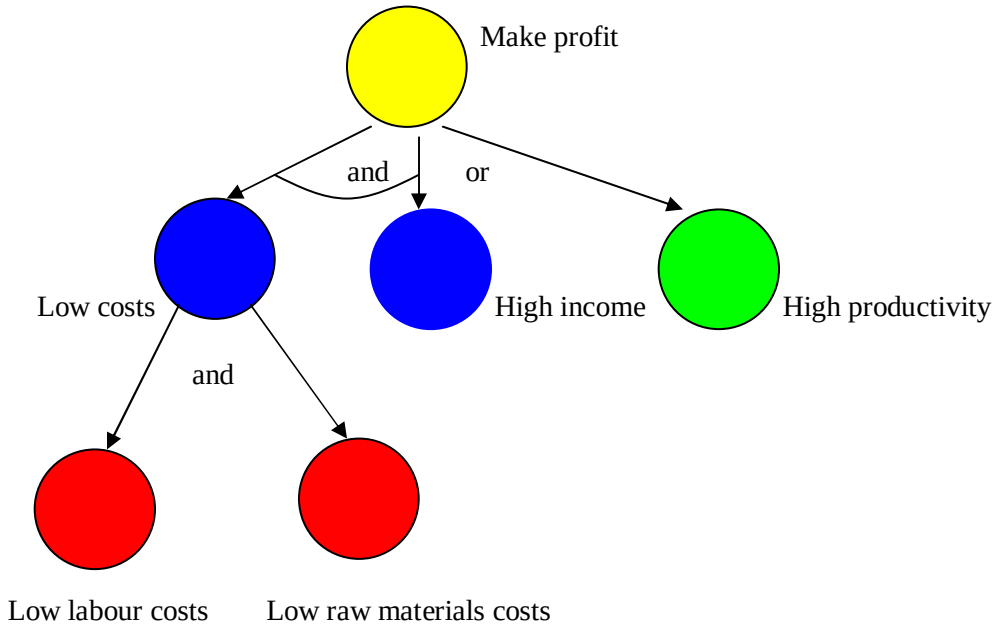
And-or trees/graphs represent only the search space for problem solving. They are compatible, therefore, with many different kinds of search strategy for finding problem solutions. The different kinds of search strategy include depth-first, breadth-first search and best-first search. Branch-and-bound combines depth-first and best-first search and is especially useful for incorporating decision-theoretic methods into the search for optimal solutions.

Like simple depth-first search, **branch-and-bound** explores only one partial solution at a time. Once a complete solution has been found, the degree to which it satisfies the goal is used as a bound, to truncate other partial solutions that cannot be extended to a better complete solution. Under certain easy to satisfy conditions, any new solutions can be guaranteed to improve upon the previous ones. As a consequence, if there is only limited time to search for a solution, then the search can be terminated when the time is up, and the currently best solution will be the best solution that can be found in the time available.

Conflict-resolution in goal hierarchies

Viewing goals in terms of goal hierarchies can help an agent to decide what to do when its actions fail, conflict with other actions, or conflict with other goals. In particular, it can direct the agent's attention to higher-level goals and can help the agent to explore alternative ways of satisfying those higher-level goals. For example, if going on strike does not produce the desired effect of increasing pay, then the goal hierarchy shows that increasing productivity is an alternative way of trying to accomplish the same goal.

Goal hierarchies are especially useful when there are conflicts between the goals of different agents. For example, management might have a different hierarchy of goals, including such goals as:



It is easy to see how labour and management can come into conflict if the workers threaten to go on strike to increase their pay, and management wants to keep labour costs low to make a profit. By analysing the two goal hierarchies, it may be possible to avoid the conflict by choosing alternative ways of satisfying the two agents' higher-level goals, in this case by deciding to improve productivity and to increase pay without going on strike.

Goal-reduction procedures

And-or trees/graphs depict an agent's goals and sub-goals in pictorial terms and highlight their hierarchical structure. However, the relationship between goals and sub-goals can also be represented linguistically, by means of **goal-reduction procedures**.

For example, the goals and sub-goals in the workers' goal hierarchy above can be represented by the procedures:

> To improve enjoyment of life, increase standard of living or work less hard.
> To increase standard of living, increase pay.
> To increase pay, go on strike or increase productivity.
> To provide for old age, increase pay and save money.

Similarly, the relationship between defending yourself and attacking anyone who attacks you can also be represented by a goal-reduction procedure:

> To defend yourself against attack, whenever someone attacks you, attack them back.

In general, a goal-reduction procedure begins with a statement of a goal and is followed by zero, one or more conditions or sub-goals. Viewed in programming language terms, these procedures are a non-deterministic program. They are non-deterministic in two senses: First, they do not determine the order in which alternative procedures are tried to solve the same goal. Second, they do not determine the order in which conjoint sub-goals are explored. These search-related determinations are made by the programming language compiler/interpreter rather than by the program.

A collection of goal-reduction procedures can be understood as a program for constructing an and-or tree, from the top down. By matching the goal of a procedure with a goal in a tree, the same procedure can be applied to different goals in the same or different trees.

Thus, although and-or trees/graphs give a good picture of the hierarchical structure of goals, goal-reduction procedures are more economical, easier to express in linguistic terms, and easier to reuse to solve different goals.

Goal-reduction procedures are a special case of the procedural representations of knowledge in Artificial Intelligence (Winograd 1972), advocated in the 1970s as an alternative to declarative, logic-based representations.

Logic programs

Logic programs (Kowalski 1974) reconcile declarative representations of beliefs with goal-reduction procedures. Logic programs are implications written in the form:

conclusion if conditions.

These are ordinary logical implications written backwards, to indicate that their intended use is to reason backwards, turning the declarative implication into a goal-reduction procedure:

to establish the conclusion, establish the conditions.

For example, backward reasoning turns the implication:

You defend yourself,
if whenever someone attacks you, you attack them back.

Into the goal-reduction procedure:

To defend yourself,
whenever someone attacks you, attack them back.

Logic programs solve the problem with BDI beliefs – that they are hard to implement – by using backward reasoning to turn beliefs that have the form of implications into goal-reduction procedures. They also help with the problem of verifying goal-reduction procedures, by giving them a declarative meaning as implications, which can be used to determine their truth or falsity.

The logic programming connection between goal-reduction procedures and beliefs has consequences for conflict resolution: A conflict between two goals can sometimes be resolved, not simply by solving higher-level goals in alternative ways, but by changing beliefs, so that inconsistent sub-goals no longer arise.

Logic programs and maintenance goals

Logic programs unify goal-reduction procedures and beliefs in BDI agent models. Typically, they are applied only to simple (one-off) "achievement goals" of the kind that are associated with and-or trees. However, they can also be applied to and combined with (on-going) "maintenance goals", which can be viewed as a generalisation of condition-action rules.

Such maintenance goals can occur at the highest-level, e.g.:

If there is an emergency, then you get help.

Maintenance goals at this level are analogous to integrity constraints in databases. Whereas data in a database can be viewed as beliefs that describe some aspect of the world, integrity constraints can be viewed as goals that prescribe the behaviour of the database.

Maintenance goals can also occur at lower levels, as the result of goal-reduction. For example, given the higher-level goal of defending yourself, backward reasoning applied to the logic program:

You defend yourself, if whenever someone attacks you, you attack them back.

gives rise to the lower-level sub-goal:

Whenever someone attacks you, you attack them back.

Except for the declarative form of the conclusion, this sub-goal can be rewritten in the form of a condition-action rule:

If someone attacks you, then you attack them back.

We will see later that this kind of reasoning can often be performed in advance, before a specific problem arises. In such a case, backward reasoning can be used to transform goals and beliefs into the form of condition-action rules that have the same behaviour.

The unified logic-based agent model

The unified agent model combines the agent cycle of production systems with the use of logic in BDI agents to represent goals and beliefs. It also combines forward reasoning using maintenance goals with backward reasoning using logic programs

The cycle in the unified agent model has the same form as in production systems:

To cycle,
observe,
think,
decide what actions to perform,
act,
cycle again.

However, different from production systems, all observations and actions are interactions between the agent and the external environment. Thinking can be interrupted at any time, to assimilate observations and to perform actions. It can be resumed, when there is more time to think.

The combination of backwards and forwards reasoning is illustrated by the following example:

Goal: If there is an emergency, then you get help.

Beliefs: You get help if you alert the driver.
You alert the driver if you sound the alarm.
There is an emergency if there is a fire.

Observation	There is a fire.
Forward reasoning	There is an emergency.
Forward reasoning: goal	You get help.
Backward reasoning: sub-goal	You alert the driver.
Backward reasoning: action sub-goal	You sound the alarm.

Here thinking is initiated by the observation of a fire. The record of this observation triggers forward reasoning using a belief whose condition matches the record of the observation. The conclusion of the belief, in turn, triggers another step of forward reasoning, but this time using the top-level maintenance goal, deriving the conclusion as an achievement goal. This derived achievement goal, "to get help", triggers two steps of backward reasoning, which ends with the derivation of an action sub-goal.

In the simplest case, forward reasoning derives consequences of observations. However, it can also be used to monitor sub-goals for desirable or undesirable consequences. Suppose, for example, that in addition to the goal and beliefs above, we have the beliefs and goal:

Beliefs:	You are warm if you turn on the heating.
	You are warm if there is a fire.

Goal:	You are warm.
Backward reasoning: sub-goal	You turn on the heating.
Alternatively:	
Backward reasoning: sub-goal	There is a fire.
Forward reasoning	There is an emergency.
etc.	

This time, thinking is initiated by a goal, rather than by an observation.[1] Backward reasoning from the goal derives two alternative sub-goals, turning on the heating or starting a fire. Before deciding between these alternatives, forward reasoning can be used to reason hypothetically, to infer other outcomes of the alternatives. In the case of starting a fire, this includes creating an emergency. Calculating the expected costs and benefits of all the outcomes of the different alternatives can help to decide which alternative to choose.

Goals and beliefs can often be compiled into condition-action rules

In computer programming, it is often possible to perform some computation in advance, before specific problems are given as input. This computation in advance is a kind of compiling of the program from a higher-level form, which is easier to develop, validate and maintain, into a lower-level form, which is more efficient to execute.

[1] However, this "achievement goal" might have been derived from a "maintenance goal" triggered by an earlier observation that it is cold.

A similar phenomenon arises when logic is used for problem solving. Consider again the goal and beliefs:

Goal: If there is an emergency, then you get help.

Beliefs: You get help if you alert the driver.
You alert the driver if you sound the alarm.
There is an emergency if there is a fire.

Instead of waiting for a fire, to reason about what to do, it is possible to perform much of the relevant reasoning in advance, both reasoning forward from the assumption that there is a fire and backward from the goal of getting help. The result of these two chains of reasoning is to derive the lower-level goal:

If there is a fire, then you sound the alarm

which has the form of a condition-action rule in declarative mood.

In fact, if the original goal and beliefs were the agent's only goals and beliefs, then forward reasoning using the derived condition-action rule would generate the same behaviour as the combination of forward and backward reasoning using the original goal and beliefs.

The process of compiling a higher-level program into a lower-level program can sometimes be reversed. However, some programs written directly in lower level programming languages do not always have an obvious higher-level counterpart.

A similar phenomenon arises when an agent's behaviour is determined directly by means of condition-action rules, without the direct involvement of any higher-level goals. In some such cases, it may appear that the agent is not rational, because its behaviour does not seem to achieve any sufficiently useful purpose. Nonetheless, it may be possible to rationally reconstruct the agent's behaviour, by ascribing goals and beliefs that, when compiled, would give generate the same behaviour.

Decompiling an agent's condition-action rule behaviour gives a higher-level representation, which, like a higher-level program, is easier to validate and maintain. Maintaining the higher-level representation includes modifying the agent's goals and beliefs, both to respond better to changing patterns of events in the environment and to reconcile conflicts with other agents.

The Israeli-Palestinian conflict is arguably such a case, where the agents' behaviour is not always directed by clearly identifiable goals and beliefs. Nonetheless, the Agha-Malley proposed solution rationally reconstructs the goals and beliefs of the two sides, to identify alternative ways for the agents to achieve their higher-level goals.

The prisoner's dilemma, on the other hand, is a different case, where the two agents have clear-cut and well-understood goals and beliefs. However, both the prisoner's dilemma

and the Israeli-Palestinian conflict share the complicating factor that the two agents have uncertain knowledge about the other agent's intended behaviour.

The selfish prisoner's dilemma

The prisoner's dilemma illustrates the need to augment logic for thinking about goals and beliefs with decision theory for deciding between action sub-goals. Thinking about the problem, in this case, is relatively uncomplicated, and consists of simply reducing top-level goals to action sub-goals. However, deciding what to do is complicated by both prisoner's uncertainty about the other prisoner's intended behaviour.

The prisoner's dilemma also illustrates a conflict between two agents, where the best outcome for one agent is the worst outcome for the other. The conflict can be resolved, as in many other cases, by getting the two agents to agree on alternative solutions of their higher-level goals.

As we have already seen, the two prisoners have the same beliefs, which have the form of implications:

> A prisoner gets 0 years in jail
> if the prisoner turns state witness
> and the other prisoner does not.
>
> A prisoner gets 4 years in jail
> if the prisoner does not turn state witness and
> the other prisoner does.
>
> A prisoner gets 3 years in jail
> if the prisoner turns state witness and
> the other prisoner does too.
>
> A prisoner gets 1 year in jail
> if the prisoner does not turn state witness and
> the other prisoner does not turn state witness too.

Assume the first prisoner has the **"**selfish" goal:

> Minimise the expected number Y,
> where I get Y years in jail.

The goal can be separated into two parts. One part:

> I get Y years in jail.

Together with the four beliefs, determines the search space. The other part:

> Minimize the expected value of Y.

guides the search strategy towards an optimal solution.

The search space goal can be reduced to sub-goals, using the four beliefs backwards, in four different ways. Consider the sub-goals arising from the first of these beliefs:

I turn state witness and the other prisoner does not.

This way of solving the goal has the best outcome for the prisoner, namely 0 years in jail. It has two sub-goals. The first of these is an action sub-goal, namely turning state witness. However, the second is not intuitively a sub-goal at all, but rather a condition about which the first prisoner has no knowledge. But both "sub-goals" must be satisfied for the higher-level goal to be solved. The second sub-goal can be solved only by making an assumption about the second prisoner's behaviour, which may later turn out to be false.

Similarly, the third belief reduces the top-level goal to the sub-goals:

I turn state witness and the other prisoner does too.

This has the outcome of 3 years in jail. Its first sub-goal is the same action, turning state witness, as that derived from the first belief. However, its second condition can be solved only by making the opposite assumption about the behaviour of the other prisoner.

Using classical logic without probability, together with the obvious fact:

The other prisoner turns state witness or
the other prisoner does not turn state witness.

it is possible to conclude:

I get 0 years in jail or I get 3 years in jail.

Unfortunately, this is not very helpful for deciding what to do.

The alternative is to use decision theory to combine the two alternative outcomes into a single expected outcome, which is the average of all the outcomes that would arise, if the situation were to be repeated many times. To perform the decision-theoretic calculation, it is necessary to estimate the probabilities of the two alternative actions available to the other prisoner. If the first prisoner estimates that both alternative actions are equally probable, then the expected outcome is .5×0, for the case where the other prisoner does not turn state witness + .5×3, for the case where the other prisoner does turn state witness, = 1.5 years in jail.

The situation is similar if the first prisoner tries to solve the top-level goal, by not turning state witness, using the second and fourth beliefs. With the same assumptions about the expected behaviour of the other prisoner, the expected outcome is .5×4 + .5×1 = 2.5 years in jail.

Thus comparing the two solutions and maximising expected utility, the best decision for the first prisoner is to turn state witness.

The calculation can be redone with other assumptions. In particular, if the first prisoner assumes that the other prisoner will reason just like him, then it is 100 % likely that the other prisoner will also turn state witness. The expected utility of the first prisoner turning state witness is then $0\times0 + 1\times3 = 3$ years in jail, and of the first prisoner not turning state witness is $1\times4 + 0\times1 = 4$ years in jail. So the first decision, to turn state witness, remains the best option[2]. However, the most likely outcome now is that the first prisoner will get 3 years in jail instead of 1.5 years.

The co-operative prisoner's dilemma

As we know, the two prisoners can do better if they both co-operate. This is equivalent to their both agreeing to solve the common goal:

> Minimise the expected total number of years in jail, $Y_1 + Y_2$,
> where the first prisoner gets Y_1 years in jail
> and the second prisoner gets Y_2 years in jail.

As before, the goal can be separated into two parts. One part:

> The first prisoner gets Y_1 years in jail and
> the second prisoner gets Y_2 years in jail.

determines the search space. The other part:

> Minimize the expected value of $Y_1 + Y_2$.

guides the search strategy towards an optimal solution.

The co-operative formulation of the problem treats the two prisoners as one agent, and assumes that the two prisoners will maximise the combined utility of both their actions, removing any uncertainty about their behaviour.

Each of the two top-level goals can be solved in four different ways, giving an initial total of eight combinations. But half of these combinations are infeasible, because they involve incompatible actions. For example, the seemingly best solution, where both prisoners each get 0 years in jail, has the four sub-goals:

> the first prisoner turns state witness and
> the second prisoner does not turn state witness and

[2] In fact, it is possible to show that, no matter what probability is assigned to the expected behaviour of the other prisoner, the first prisoner's best decision is always to turn state witness. However, to show that this is the case requires more sophisticated reasoning on the part of the first prisoner than seems plausible in a cognitive model designed to explain routine, rather than exceptional intelligent behaviour.

the first prisoner does not turn state witness and
the second prisoner turns state witness.

The best feasible solution, where both prisoners each get 1 year in jail, has the four sub-goals:

the first prisoner does not state witness and
the second prisoner does not turn state witness and
the first prisoner does not turn state witness and
the second prisoner does not state witness.

which simplifies to

the first prisoner does not state witness and
the second prisoner does not turn state witness.

The prisoner's dilemma illustrates a weak form of conflict, which arises when two agents pursue their own goals selfishly and independently, achieving a suboptimal solution (a likely 3 years in jail each). It also illustrates a weak form of conflict resolution, by combining the two otherwise competing goals into a single, conjoint, co-operative goal, achieving an optimal solution (a certain 1 year in jail each).

Both versions of the prisoner's dilemma illustrate the use of declaratively expressed beliefs to reduce goals to sub-goals. The sub-goals are of two kinds: action sub-goals, which can be solved only by executing them successfully; and condition sub-goals, whose truth or falsity is determined by the external environment, including the behaviour of other agents.

Some condition sub-goals can be verified before their corresponding actions are selected and performed, as in the case of the conditions of condition-action rules. However, other sub-goals can be verified or falsified[3] only after their corresponding actions are performed, as in the case of the prisoner's dilemma. In such cases, the probability that the conditions will hold needs to be estimated before deciding what actions to perform.

Thus decision theory is a useful adjunct to logic in helping to decide what actions to perform. Conversely, logic is potentially a useful adjunct to game theory, because it shows how the flat analysis of games in terms of actions and their outcomes can be extended to incorporate a more complex structure of beliefs, goals and sub-goals.

The Israeli-Palestinian conflict is a case where game-theoretic concepts alone are not adequate. It has all of the features, such as uncertainty and utility, associated with game

[3] In fact, it may not always be possible to verify the conditions themselves, but only their outcomes. For example, suppose, in addition to the original deal offered to the two prisoners, that the first prisoner gets 3 years in jail if the prisoner turns state witness and the police renege on the deal. Then the first prisoner might turn state witness and end up getting 3 years in jail, without ever finding out whether it was the police or the other prisoner who was responsible for the outcome.

theory, but it also has other features, such as beliefs, goals and sub-goals, that can usefully be analysed in logical terms.

Israeli-Palestine conflict

In their proposal, Agha and Malley first outline the "basic interests" of the two sides, then translate these interests into "policy redlines", and finally present solutions to the resulting "issues". They characterise the relationship between "interests", "redlines" and "solutions" in the following terms:

> "Such a deal must protect both sides' core interests without breaching either party's "redlines", or non-negotiable demands."

To a large extent, their argument can be viewed as a top-down reduction of higher-level goals to lower-level sub-goals and ultimately to solutions. However, most of the detail needed to link goals and sub-goals is missing, and often it is hard to reconstruct.

I have already summarised the two sides' basic interests earlier in the paper. Let me summarize them again here, for ease of reference:

The basic interests

Israel's basic interests:

I_1 Preserving the Jewish character and majority in Israel
I_2 Achieving security
I_3 Achieving international recognition and normalcy
I_4 Controlling Jewish holy sites and national symbols and
I_5 Ending the conflict with Palestinians and Arab States.

Agha and Malley's use of the term "interest" is equivalent to our use of the term "goal". Their use of the qualification "basic" suggests that "basic interests" are higher-level goals.

However, the five basic interests are not logically independent. In particular, the goal I_4 of controlling Jewish holy sites and national symbols can be understood as part of what it means to satisfy the higher-level goal I_1 of preserving the Jewish character of Israel. Thus I_4 is one of several sub-goals needed to satisfy I_1. In fact, it could also be argued that the two parts of I_1 are not independent, and that preserving the Jewish majority is a necessary sub-goal[4] of preserving the Jewish character of Israel.
Indeed, it could be argued that

[4] A necessary sub-goal is a sub-goal that belongs to every alternative way of solving a higher-level goal. It is natural to think, in logical terms, of necessary sub-goals as implied by the higher-level goal. However, in our analysis of the logical relationship between goals and sub-goals, the implication is in the opposite direction: Necessary sub-goals, possibly with other conjoint sub-goals, imply the higher-level goal.

Israel has the character of a Jewish state if (and only if)
Israel has a Jewish majority and
Israel controls Jewish holy sites and national symbols.

Similarly, it can be argued that the goal I_3 of achieving international recognition and normalcy can be reduced to the sub-goal I_5 of ending the conflict with the Palestinians and the Arab States. In fact, it is hard to think of any better way of achieving international acceptance I_3 than by ending the conflict I_5.

It can also be argued that the goal I_2 of achieving security can also be reduced to the sub-goal I_5, because the only significant threat to Israel's security comes from its conflict with the Palestinians and the Arab States. However, it might be possible to achieve some degree of security without ending the conflict, by establishing other, sufficiently powerful means of deterrence. The recent start of construction of a physical barrier highlights the existence of such alternatives. Thus I_5 might be a sufficient sub-goal of I_2, but it is not a necessary sub-goal of I_2.

Because goals, such as I_2 and I_5, can be achieved to varying degrees and because their achievement depends on uncertain expectations about the behaviour of other agents, it makes sense to include both I_2 and I_5 as separate goals.

These various relationships among the "basic interests" are relevant to the extent that they affect the argument for the proposed solution. Therefore, we will return to this discussion after we have seen the complete argument.

The Palestinian basic interests:

P_1 Living in freedom, dignity, equality and security
P_2 Ending the occupation and achieving national self-determination
P_3 Resolving the refugee issue fairly
P_4 Controlling Muslim and Christian holy sites and
P_5 Ensuring any solution is accepted by the Arab and Muslim worlds.

Again, there are relationships among these goals. Presumably, the last goal P_5 of ensuring acceptance of any solution by the Arab and Muslim world would follow from a satisfactory solution of the other goals P_1- P_4. The second goal P_2 is synonymous with establishing an viable, independent Palestinian state. It would seem to imply at least the first two of the three sub-goals in P_1. The third sub-goal of P_1 might need independent guarantees that a settlement of the conflict would secure the new Palestinian state against any future security threat by Israel.

The redlines

Agha and Malley state that Israel's basic interests "translate" to the "policy redlines":

RI_1 No mass influx of Palestinian refugees into Israel
RI_2 Jerusalem as capital of Israel
RI_3 Jewish "link" to the Temple Mount

RI_4 No return to the 1967 borders
RI_5 Incorporation into Israel of most Jewish settlements
RI_6 No second army between the Jordan River and the Mediterranean
RI_7 Perpetuation of the Jordan Valley as Israel's Eastern security border.

Agha and Malley do not explain what they mean by the term "translate". Nor do they otherwise explain how the redlines are related to the basic interests. Nonetheless, it is possible to reconstruct some of these relationships.

The redline RI_1, no mass influx of refugees into Israel, is clearly a necessary sub-goal of the goal I_1 of preserving Israel's Jewish majority. The redline RI_2, Jerusalem as capital of Israel, is arguably a necessary sub-goal of preserving Israel's Jewish character (also I_1). The redline RI_3, a Jewish link to the Temple Mount, is a necessary sub-goal of the Israeli goal I_4 of controlling Jewish holy sites and national symbols, which is, as I argued before, another aspect of preserving Israel's Jewish character.

The redlines RI_6 and RI_7 are clearly concerned with security I_2. In our terminology, they are conjoint sub-goals of one alternative way of achieving security. Obviously they are not sufficient. However, together with the goal I_5, of ending the conflict with the Palestinians and Arab States, they would undoubtedly considerably increase the degree to which Israel's security could be achieved.

The relationship of the remaining redlines RI_4 and RI_5 to Israel's basic interests is harder to reconstruct. Clearly, from Israel's point of view, RI_4 and RI_5 are desirable in their own right, even though they might not be as "basic" as its other "interests".

Palestinian basic interests "translate" to the "policy redlines":

RP_1 Palestine state with the equivalent of 100% of the land lost in 1967
RP_2 Refugees given the choice of returning to their homes before 1948
RP_3 Jerusalem as the capital of Palestine
RP_4 Security guarantees for "what would be a non-militarised state".

Clearly, RP_1 is a sufficient sub-goal for the basic interest P_2 of ending the occupation and achieving national self-determination. However, it includes a qualification about the equivalence of the land of the new Palestine state to 100% of the land lost in 1967. This qualification anticipates the need to resolve the conflict with the Israeli redline RI_4 of no return to the 1967 border. It opens the way, therefore, to the proposed land swaps that are an important feature of the solution that Agha and Malley present later in their article.

No doubt, from the Palestinian perspective, RP_2, the unrestricted right of return for the refugees, is the optimal solution of P_2, the refugee problem. However, it conflicts head on with the Israeli redline RI_1, of no mass influx of refugees into Israel, which is a necessary condition of the higher-level goal I_1, of preserving a Jewish majority in Israel. It will have to be reconciled with RI_1 (and therefore "breached") in the final proposal.

RP_3, making Jerusalem the capital of Palestine, contributes to several higher-level "basic interests", including P_2 self-determination and P_4 control over Muslim and Christian holy sites.

RP_4 has two parts. One part, the non-militarisation of Palestine is actually a sub-goal of I_2, the Israeli security interest. The other part, security guarantees for the resulting Palestine state is simply a restatement of the Palestinian basic interest P_1 in its own security, presumably made more difficult to satisfy now by the non-militarisation requirement.

Thus, the Palestinian redlines are somewhat different from the Israeli redlines. They include non-militarisation, which is a sub-goal of the Israeli security goal; and they include goals, like the refugees' right of return to their homes before 1948, that will be "breached" in the final proposal.

The issues and their solution

Agha and Malley conclude their argument by presenting solutions to five issues. They do not explain how these issues relate to their earlier analysis of the basic interests and redlines. However, it is possible to reconstruct most of these relationships.

Again, for ease of reference, I list these issues here:

1 The territorial issue
2 Security
3 Jerusalem
4 Haram al-Sharif, or Temple Mount
5 Palestinian refugees.

1 Territorial issue. This issue directly addresses the Israeli interest I_1 of preserving the Jewish character and majority in Israel and the two redlines RI_4 of no return to the 1967 borders and RI_5 of incorporating most of the Jewish settlements into Israel. Its proposed solution reconciles these with the Palestinian interest P_2 of ending the occupation and achieving national self-determination, i.e. of establishing a viable Palestinian state, and the redline RP_1 of establishing a Palestine state with the equivalent of 100% of the land lost in 1967.

The proposed solution is based on swapping land in the West Bank, transferring a large number of Jewish settlements into Israel, with an equivalent amount of land in Israel proper, transferring the land into the new Palestine State.

2 Security. For the Israelis, this issue is a restatement of the basic interest I_2 and it takes into account the redlines RI_6, no second army between the Jordan River and the Mediterranean, and RI_7, perpetuation of the Jordan Valley as Israel's Eastern security border. It also addresses the security part of the Palestinian basic interest P_1 and incorporates the redline RP_4 of security guarantees for a non-militarised Palestinian state.

In fact, it is the non-military status of Palestine that is proposed as the main solution for Israel's security goal. The introduction of an international peace force led by the United States is the main solution proposed for Palestinian security. To further assuage Israeli fears, the force would, at first, include an Israeli component.

3 Jerusalem. The establishment of Jerusalem as capital of Israel is Israel's redline RI_2, and as capital of Palestine is the Palestinian redline RP_3.

The proposed solution is a division of Jerusalem into two parts, "based on the dual notions of demographic and religious self-governance." As they put it, "what is Jewish…should become the capital of Israel, and what is Arab should become the capital of Palestine." The Jewish areas include those established since 1967 in East Jerusalem. The proposed solution also provides for the Israeli basic interest I_4 and Palestinian basic interest P_4 of control over and "unimpeded access" to their respective holy sites.

4 Haram al-Sharif, or Temple Mount. A "link" to the Temple Mount is singled out as a separate Israeli redline RI_3. The Israeli redline is in direct conflict with the Palestinian basic interest P_4 of controlling Muslim and Christian holy sites.

The proposal is based on "practical arrangements required to meet both sides' needs." It is a special case of the solution to the more general problem of providing control and access to holy sites throughout Jerusalem.

5 Palestinian refugees. This problem of the refugees is reputed to be the problem that led to the breakdown of the 2000-2001 negotiations. It presents a direct conflict between the "non-negotiable" redlines RI_1 of no unrestricted right of return and RP_2 of the right of the refugees to return to their or their ancestors' homes prior to 1948.

The proposed solution reconciles the conflict by solving the higher-level Palestinian goal P_3, of which RP_2 is a sub-goal, in an alternative way. The alternative is to offer the refugees resettlement in Arab-populated areas of Israel and to include these areas in the land swap with Palestine. To make this solution more attractive, the refugees would be provided with generous financial compensation.

Agha and Malley argue that this solution has a better outcome for the Palestinians themselves than their redline RP_2. This is because RP_2 has the undesirable consequence that the refugees would be resettled in "what has become an alien land". In the alternative, proposed solution, "the refugees would get to live in a more familiar and hospitable environment – and one that would ultimately be ruled not by the Israelis, but by their own people."

Discussion of the Agha-Malley argument

The Agha-Malley argument is on the whole an impressive example of both the goal-reduction approach to problem solving and the goal/sub-goal approach to conflict resolution. To a large extent, my analysis confirms their characterisation of their proposal as a deal that "protects both sides' core interests without breaching either party's "redlines".

Many of the redlines are necessary sub-goals, which belong to all ways of trying to achieve the higher-level "basic interests". The redlines RI_1 and RI_2, for example, are arguably necessary sub-goals of I_1, and the redline RI_3 is a necessary sub-goal of I_4.

The solutions of the five issues are a selection from among the alternative ways of solving the basic interests, while satisfying the redlines, and reconciling conflicts between other ways of solving the higher-level goals. For example, the proposed solution of the territorial issue avoids a potential conflict between RI_4 and PI_1. The solution of the security issue reconciles a potential conflict between the higher-level security interests of both sides. Similarly, the solutions of the problems of Jerusalem and the Temple Mount, or Haram al-Sharif, avoid potential conflicts between their respective higher-level goals.

The proposed solution of the refugee problem is more interesting. It reconciles a direct conflict between the non-negotiable redlines RI_1 and RP_2. It does so by solving the Palestinian higher-level goal P_3 in an alternative way that does not conflict with RI_1. To make the alternative solution more acceptable to the Palestinians, it is linked to the provision of generous financial compensation to the refugees, which contributes to the achievement of other higher-level goals, which were not initially identified as "basic interests".

Although much of the Agha-Malley argument lives up to their characterisation of it, there are other places where it fails. The most notable example is the characterisation of RP_2 as a non-negotiable redline. Another example is the redline RP_4, which combines a redundant repetition of the higher-level Palestinian security goal P_1 with the Israeli redline RI_6, no second army between the Jordan River and the Mediterranean. Other examples are RI_4 no return to the 1967 borders and RI_5 incorporation most Jewish settlements into Israel, which are not obvious translations of the higher-level Israeli basic interests.

In addition, as mentioned earlier, the basic interest for Israel I_3 of achieving international recognition and normalcy and for the Palestinians P_5 of ensuring any solution is accepted by the Arab and Muslim worlds would both be consequences of solving the conflict. In any case, they are not addressed by any of the redlines or solutions of the five issues.

The Agha-Malley analysis of the Israeli-Palestinian conflict implicitly combines both logic for reducing goals to sub-goals and decision theory for assessing the desirability of alternative solutions. The authors implicit use of decision-theoretic concepts is most obvious in their discussion of the refugee issue: They mention that during the 2000-2001 negotiations the Israelis failed to appreciate the importance of the refugee issue to the Palestinians and the Palestinians failed to appreciate the degree to which the Israelis believe that an unrestricted right of return would spell the end of Israel as a Jewish state.

Conclusions

There are two sides to the work presented in this paper. On the one hand, it applies the logic-based agent model to problems of conflict resolution. On the other hand, it challenges the model with applications outside its ordinary range.

The conflict resolution application highlights the importance of several features of our agent model, the most important of which is the use of backward reasoning using beliefs to reduce goals to sub-goals. This feature of the model points the way to resolving conflicts between lower-level sub-goals, by finding alternative ways of solving higher-level goals.

The application also highlights the use of forward reasoning using beliefs to infer consequences of candidate actions/solutions. This feature of the model enables it to identify otherwise unforeseen consequences of the candidate solutions, including any other goals that might opportunistically solved, as well as any constraints that might be broken.

The agent model distinguishes thinking about actions from deciding what actions to perform. It is open to different decision-making strategies, including strategies that seek to optimise the expected utility of the consequences of actions.

On the other hand, the conflict resolution application draws attention to the need to extend the agent model to deal with uncertainty about conditions whose truth or falsity is determined by the external environment, including other agents. I have illustrated how such an extension, incorporating probability, might work in the example of the prisoner's dilemma, but I have not developed a general approach. Poole's (1997) Independent Choice Logic is a possible approach for developing such an extension.

It might be argued that our use of logic and goal hierarchies to help reconcile conflicts is not realistic, because people are not rational. This argument draws strength from the success of production system models, in which goals are implicit rather than explicit. However, as I have argued, production systems can sometimes be "decompiled" into higher-level logic-based representations. Such higher-level representations, like higher-level computer programs, are easier to maintain in the context of a changing environment, including an environment in which conflicts arise with other agents.

Acknowledgements

I am grateful to Phan Minh Dung, Bashar Nuseibeh, David Poole and Ken Satoh for helpful discussions and useful advice.

References

Agha, Hussein. and Robert Malley. 2002. The Last Negotiation. *Foreign Affairs.* 81: 10-18.

Baron, Jonathan. 1994. Thinking and *Deciding,* 2nd ed. Cambridge: Cambridge University Press.

Cohen, Philip and Hector Levesque. 1991. Intention is Choice with Commitment. *Artificial Intelligence* 42: 213-261.

Kowalski, Robert. 1974. Predicate logic as programming language. In *Proceedings of IFIP Congress.* North Holland Publishing Co. 569-574.

Kowalski, Robert and Fariba Sadri. 1999. From Logic Programming towards Multi-agent Systems. *Annals of Mathematics and Artificial Intelligence* 25: 391-419.

Kowalski, Robert. 2001. Artificial intelligence and the natural world. *Cognitive Processing* 4: 547-573.

van Lamsweerde Axel, Robert Darimont, Emmanuel Letier. 1998. Managing conflicts in goal-driven requirements engineering. *IEEE Transactions on Software Engineering* 24: 908-926.

Newell, Alan. 1973. Production Systems: Models of Control Structure. In *Visual Information Processing.* Edited by W. Chase. New York: Academic Press. 463-526.

Poole, David. 1997. The independent choice logic for modelling multiple agents under uncertainty. *Artificial Intelligence* 94: 7-56.

Post, Emil, 1943. Formal reductions of the general combinatorial problem. *American Journal of Mathematics* 65: 197-268.

Rao, Anand and Michael Georgeff. 1992. An abstract architecture for rational agents. In *Proceedings of Knowledge Representation and Reasoning.* Edited by C. Rich, W. Swartout, and B.Nebel. 439-49.

Shoham, Yoav. 1993. Agent-oriented programming. *AI Journal* 60: 51-92.

Slagle, James. 1961. A heuristic program that solves symbolic integration problems in freshman calculus: Symbolic Automatic Integrator (SAINT). Report no. 5G-0001, Lincoln Laboratory, PhD dissertation. Massachusets Institute of Technology.

Winograd, Terry. 1972. *Understanding natural language.* New York: Academic Press.