Lies can be beneficial

古早论文，伊拉克的神秘力量，引用量很高，但感觉时效性不是很强，看了一半就没兴趣了

Incomplete Information and Deception in Multi-Agent Negotiation

Gilad Zlotkin
Jeffrey S. Rosenschein
Computer Science Department
Hebrew University
Givat Ram, Jerusalem, Israel
gilad@cs.huji.ac.il, jeff@cs.huji.ac.il

Abstract

Much distributed artificial intelligence research on negotiation assumes complete knowledge among the interacting agents and/or truthful agents. These assumptions in many domains will not be realistic, and this paper extends previous work to begin dealing with the case of inter-agent negotiation with incomplete information.

A discussion of our existing negotiation framework sets out the rules by which agents operate during this phase of their interaction. The concept of a "solution" within this framework is presented; the same solution concept serves for interactions between agents with incomplete information as it did for complete information interactions.

The possibility of incomplete information among agents opens up the possibility of deception as part of the negotiation strategy of an agent. Deception during negotiation among autonomous agents is thus analyzed in the constrained Blocks Domain, and it is shown that beneficial lies do exist in some scenarios. The three types of interactions, cooperative, compromise, and conflict, are examined. An analysis is made of how each affects the possibility of beneficial deception by a negotiating agent.

1 Introduction

The subject of negotiation has been of continuing interest in the distributed artificial intelligence (DAI) community [Smith, 1978; Rosenschein and Genesereth, 1985; Durfee, 1988; Malone *et* al., 1988; Sycara, 1988; Kuwabara and Lesser, 1989; Conry *et* a/., 1988; Kreifelts and von Martial, 1990; Laasri *et* al., 1990; Kraus and Wilkenfeld, 1991; Ephrati and Rosenschein, 1991]. Despite the large amount of research on this topic, there does not yet exist a universally accepted definition of what the word even means; as Gasser points out in [Gasser, 1991], "'negotiation' [is] a term that has been used in literally dozens of different ways in the DAI literature." Nevertheless, it is clear to the DAI community as a whole that the operation of intelligent autonomous agents would be greatly enhanced if they were able to communicate their respective desires and compromise to reach mutually beneficial agreements.

The work described in this paper follows the general direction of [Rosenschein and Genesereth, 1985; Zlotkin and Rosenschein, 1989] in treating negotiation in the spirit of game theory, while altering game theory assumptions that are irrelevant to DAI.

Much of the research on negotiation assumes complete knowledge among the interacting agents and/or truthful agents. These assumptions in many domains are not realistic, and this paper extends previous work [Zlotkin and Rosenschein, 1989; Zlotkin and Rosenschein, 1990b] to begin dealing with the case of inter-agent negotiation with incomplete information.

2 The Overall Negotiation Framework

$G = \{ s \text{ satisfy } g_i \mid \forall s \text{ (state)} \}$

Each agent *i* in the interaction is assumed to have a goal g_t, and wants to transform the world from an initial state *s* to a state that satisfies this goal. The set of all states satisfying g_i, is denoted by G,. A goal has an associated *worth* to the agent, which is also the maximum the agent is willing to pay in order to achieve that goal.

Because two agents co-existing within the same environment might interfere with actions of the other, there needs to be coordination of activity. At the same time, depending on the particular domain and goals involved, there may be the possibility that the agents will actually be able to help each other and achieve both goals with a lower overall cost.

A *deal* between agents is generally a joint plan, where agents share the work of transforming the world from the initial state to some final state. The plan is "joint" in the sense that the agents might probabilistically share the load, compromise over which agent does which actions, or even compromise over which agent gets its goal satisfied. In this final case, the deal is actually a probabilistic distribution over joint plans (what was called a "multi-plan deal" in [Zlotkin and Rosenschein, 1990c]).

In the broad sense, the utility for an agent of a deal is the difference between the worth of the agent's goal achieved through that deal, and the cost of that agent's part of the deal.

2.1 The Process of Negotiation

The interaction between agents occurs in two consecutive stages. First the agents negotiate, then they execute the entire joint plan upon which they bad agreed. No divergence from the negotiated deal is allowed. The sharp separation of stages has consequences, in that it rules out certain negotiation tactics that might be used in an interleaved process. A more general negotiation framework that allowed concurrent negotiation and execution might, however, be approximated by concatenating several negotiation /execution processes together, provided that each agent remembers and uses information about the preceding negotiations.

We assume that negotiation is an iterative process: at each step, both agents simultaneously offer a deal. Our protocol specifies that at no point can an agent demand more than it did previously—in other words, each offer either repeats the previous offer or makes a concession to the opponent's position. The negotiation can end in one of two ways:

- Conflict: if neither agent makes a concession at some step, they have by default agreed on the (domain dependent) "conflict deal."
- Agreement: if at some step agent A offers agent B more than B himself asks for, they agree on A's offer, and if both agents overshoot the others' demands, then a coin toss breaks the symmetry.

The result of these rules is that the agents cannot "stand still" in the negotiation, nor can they backtrack. Thus the negotiation process is strongly monotonic and ensures convergence to a deal.

2.2 The Concept of a Solution

When we say that we are looking for a "solution'' to the negotiation problem, we mean two things:

1. A precise definition of deals and utility, which may include probabilistic sharing of actions, probabilities associated with achieving final states, partial achievement of goals, and domain dependent attributes (e.g., the nature of the conflict deal). Previous work discussed *pure deals, mixed deals* [Zlotkin and Rosenschein, 1989], *semi-cooperative deals* [Zlotkin and Rosenschein, 1990b], and *multi-plan deals* [Zlotkin and Rosenschein, 1990c].
2. A specification of how an agent should negotiate, given a well-defined negotiation environment.

How should one evaluate solutions? There are several ways of doing this, related to how we evaluate deals, how we evaluate agents, and how we evaluate interactions among the agents.

- Deals: Deals may have a variety of attributes that are considered desirable. Certain kinds of deals provide solutions to more general situations (e.g., semi-cooperative deals offer solutions to conflict resolution, whereas mixed deals do not), thus increasing the size of the *negotiation set* (the set of possible agreements). Deals may have other positive attributes, such as requiring less initial information.
- Agents: Agents are expected to be designed so that they are individual rational (meaning they agree on deals with positive utility).
- Inter-agent interactions: When two agents negotiate, it is desirable that they converge to a *pareto optimal* deal (meaning the only way the deal could be improved for one agent would be to worsen the deal for the other agent).

It is also highly desirable that an agent's negotiation strategy be in *equilibrium*—a strategy S is said to be in equilibrium if assuming that your opponent is using S, the best you can do is to also use S. Thus, no other agent will be able to take advantage of that agent by using a different negotiation strategy. Moreover, there is no need to exercise secrecy regarding the design of that agent—on the contrary, it is actually beneficial to broadcast its negotiation strategy, so that the other agent doesn't blunder and potentially cause both harm. Different types of deals may change the availability of strategies that are in equilibrium.

There is, in a sense, a meta-game going on between the designers of autonomous agents. Each one wants to design an agent that maximizes the designer's utility. There are strong motivations to design your agent so that it uses a negotiation strategy in equilibrium, in that it results in "best performance" in pair-wise competitions between your agent and any other given agent—conflicts will be avoided whenever possible, and deals that are reached will be pareto optimal.[1]

2.3 Negotiation with Incomplete Information

If agents negotiate without having full information regarding the other agent's goal, they need to take this lack of information into account in their negotiation strategy. There are several frameworks for dealing with this:

- In [Zlotkin and Rosenschein, 1989], we introduced the notion of a "—1 negotiation phase" in which agents simultaneously declare their goals before beginning the negotiation. The negotiation then proceeds as if the revealed information were true. There, we analyzed the strategy that an agent should adopt for playing the extended negotiation game, and in particular, whether the agent can benefit by declaring something other than his true goal.
- An alternative approach is for the agents to start the negotiation with incomplete information, increasing their knowledge as the negotiation process proceeds. Methods for increasing knowledge about another agent's goal we call "goal recognition" techniques.[3]

[1] However, there may be ecological motivations for designing agents that don't use equilibrium strategies, since multiple non-equilibrium agents might all benefit from their deviant strategies. For example, ecologically a group of agents that all play Cooperate in the Prisoners' Dilemma will do better than a group using the equilibrium strategy of Defect, even though a Defecting agent will win in a head-to-head competition with a Cooperating agent [Axelrod, 1984]. Nevertheless, we here concentrate on pair-wise equilibrium, and search for equilibrium negotiation strategies.

*The techniques used for goal recognition will depend on the form of the goal itself. For example, a goal comprised

2.4 Discoverable Lies vs. Undiscoverable Lies

We are unwilling to compel the designers of intelligent, autonomous agents not to design lying into their creations. In fact, the point of this research is to analyze whether or not such a design consideration has any advantages—maybe lies are helpful to societies of autonomous agents, and maybe they are not. While unwilling to outlaw lies, we *are* willing to consider what happens when a penalty mechanism is introduced, such that agents that are discovered lying would be punished. At times [Zlotkin and Rosenschein, 1989], we have considered infinite negative penalties when a lie was discovered. Even an infinite negative penalty, however, does not rule out a space of lies that could never be discovered, and agents might still benefit by using those lies.

We are also willing to assume that agents will keep their commitments: if an agent commits to some action as part of a deal, he will carry the action out.

2.5 Worth of a Goal and its Role in Lies

Throughout our research, we have assumed that agents associate a *worth* to the achievement of a particular goal. Sometimes, this worth is exactly equal to what it would cost the agent to achieve that goal by himself. At other times, we have analyzed what negotiation strategies are suitable when the worth of a goal to an agent exceeds the cost of the goal to that agent [Zlotkin and Rosenschein, 1990c]. The worth of a goal is the baseline for calculating the utility of a deal for an agent.

The worth of a goal is intimately connected with what specific deals agents will agree on. First, an agent will not agree on a deal that costs him more than his worth (he would have negative utility from such a deal). Second, since agents will agree on a deal that gives them both equal utility [Zlotkin and Rosenschein, 1989], if an agent has a lower worth, it will ultimately reduce the amount of work in his part of the deal. Thus, one might expect that if agent *A* wants to do less work, he will try to fool agent *B* into thinking that, for any particular goal, /Ts worth is lower than it really is. This strategy, in fact, often turns out to be beneficial, as seen below.

2.6 Lies in the Postmen Domain

The question of when it might benefit an agent to lie during negotiation was raised in [Zlotkin and Rosenschein, 1989]. The Postmen Domain (agents delivering letters over an undirected graph) was introduced, and several negotiation protocols were examined. When negotiating over *pure deals* (a redistribution of tasks among agents), there are situations where an agent can benefit by either hiding a goal or claiming to have a phantom goal that does not really exist. It was shown in [Zlotkin and Rosenschein, 1990a] that the availability of beneficial lies in the Postmen Domain is strongly related to the way the two agents' goals are coupled. When their goals are tightly coupled (i.e., delivery of one set of letters significantly reduces the cost of delivery of the second set), then beneficial *lies* are *likely to be found. When the* agents' goals are decoupled then a beneficial lie cannot be found.

solely of a conjunction of positive predicates could be deduced through a process of set intersection of positive examples.

By introducing the concept of an all-or-nothing *mixed deal* (a probabilistic redistribution of tasks among agents), beneficial lies were effectively eradicated. Thus, in the Postmen Domain, the existence or non-existence of beneficial lies was sensitive to the negotiation protocol being used.

3 The Constrained Blocks Domain

We here quickly review the constrained Blocks Domain, first presented in [Zlotkin and Rosenschein, 1990b].

There is a table and a set of blocks. A block can be on the table or on some other block, and there is no limit to the height of a stack of blocks. However, on the table there are only a bounded number of slots into which blocks can be placed. There are two operations in this world: PickUp(i) — Pick up the top block in slot i (can be executed whenever slot i is not empty), and PutDown(t) — Put down the block that is currently being held into slot t. An agent can hold no more than one block at a time. Each operation costs 1.

An Example in the Blocks Domain:

The initial state can be seen at the left in Figure 1. g_A is "The Black block is on a Red block which is on the table at slot 2" and g_B is "The White block is on a Red block which is on *the* table at slot 1".

In order to achieve his goal alone, each agent has to execute four Pick Up and four Put Down operations that cost (in total) 8. The two goals do not contradict each other, because there exists a state in the world that satisfies them both, as can be seen on the right side of Figure 1. There also exists a joint plan that moves the world from the initial state to a state that satisfies *both* goals with total cost of 8—one agent lifts the white block, while the other agent rearranges the other blocks suitably (by picking up and putting down each block once), whereupon the white block is put down. The agents will agree to split this joint plan with probability 0.5, leaving each with an expected utility of 4.

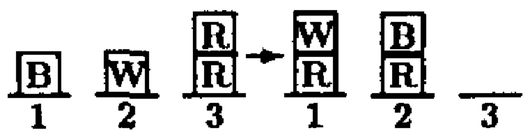

Figure 1: A Simple Cooperative Example

3.1 Types of Interactions — Cooperative, Compromise, Conflict

As presented in [Zlotkin and Rosenschein, 1990b], there are universally three types of possible interactions, from the point of view of an individual agent.

- A *cooperative* situation is one in which there exists a deal in the negotiation set that is preferred by an agent over achieving his goal alone. Here, an agent welcomes the existence of the other agents.

[3]The original result relied on the assumption that a phantom letter might be discovered in a probabilistic agreement. The result has been strengthened, and *it* has been shown that even if the phantom letter cannot be discovered (e.g., the phantom letter can be generated by the lying agent as needed), there are still no beneficial lies when the agents use all-or-nothing mixed deals.

- A *compromise* situation is one where there are individual rational deals for an agent. However, an agent would prefer to be alone in the world, and to accomplish his goal alone. Since he is forced to cope with the presence of other agents, he will agree on a deal. All of the deals in the negotiation set are better for the agent than leaving the world in its initial state *s.*
- A *conflict* situation is one in which the negotiation set is empty—no individual rational deals exist.

It is possible for different agents to have differing views of an interaction—it might be, for example, a cooperative situation for one and a compromise situation for the other. However, a conflict situation is always symmetric.

Given two agents' goals and an initial state, the type of interaction in which they find themselves is a function of the negotiation protocol they are about to use. Part of the reason for introducing new negotiation protocols is specifically to change conflict situations into non-conflict situations. By default, when we refer to the three terms above (cooperative, compromise, and conflict), we mean relative to agents negotiating over *mixed joint plans* that achieve both agents' goals.

Each of these situations can be visualized informally using diagrams. The cooperative situation can be seen in Figure 2, the compromise situation in Figure 3, and the conflict situation in Figure 4. A point on the plane represents a state of the world. Each oval represents a collection of world states that satisfies an agent's goal, *s* is the initial state of the world. The triple lines emanating fioro s represent a joint plan that moves the would to some other state; the agents share in carrying out the plan. The overlap between ovals represents final states that satisfy the goals of both agents *A* and *B.* Informally, the distance between *s* and either oval represents the cost associated with a single-agent plan that transforms the world to a state satisfying that agent's goal.

Note that in Figure 3, the distance from *s* to either agent's oval is less than the distance to the overlap between ovals. This represents the situation where it would be easier for each agent to simply satisfy his own goal, were he alone in the world. In Figure 2, each agent actually benefits from the existence of the other, since they will share the work of the joint plan.

In Figure 4, a semi-cooperative deal is pictured: the agents will carry out a joint plan to an intermediate state *t,* and then will flip a coin with probability *q* to decide whose goal will be individually satisfied [Zlotkin and Rosenschein, 1990b]; each single arrow represents a one-agent plan.

3.2 Beneficial Lies in Non-Conflict Situations

Consider our Blocks World example above, drawn in Figure 1. Recall that the agents' true goals are as follows: g_A is "The Black block is on a Red block which is on the table at slot 2," and g_B is "The White block is on a Red block which is on the table at slot 1".

What if agent *A* lies about his true goal, claiming that he wants a Black block on *any* other block at slot 2? If A were alone in the world, he could apparently satisfy this relaxed goal at cost 2. Assuming that agent *B* reveals

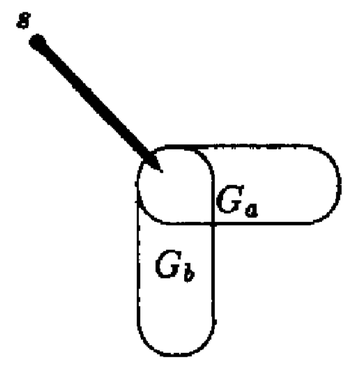

Figure 2: The Cooperative Situation

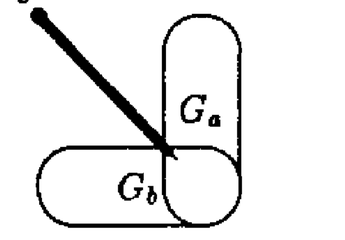

Figure 3: The Compromise Situation

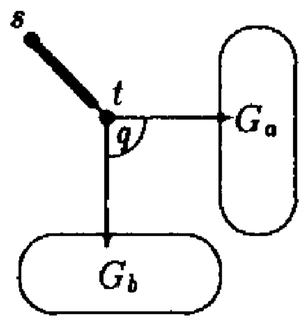

Figure 4: The Conflict Situation

his true goal, the agents can only agree on one plan: *A* will lift a block (either the White or Black one), while *B* does all the rest of the work. The apparent utility for *A* is then 0 (still individual rational), while *B* has a utility of 2. In reality, *A* has an actual utility of 6. *A's* lie has benefited him.

This works because *A* is able to reduce the apparent cost of his carrying out his goal alone (which ultimately causes him to carry less of a burden in the final plan), while not compromising the ultimate achievement of his real goal. The reason his real goal is "accidentally" satisfied is because there is only one state that satisfies B's real goal and ,4's apparent goal, coincidentally the same state that satisfies both of their real goals.

The lie above is not A's only beneficial lie in this example. What if *A* claimed that his goal is "Slot 3 is empty and the Black block is clear"? Interestingly, this goal is quite different from his real goal. If A were alone in the world, he could apparently satisfy this variant goal at cost 4. The agents will then be forced again to agree on the deal above: *A* does two operations, with apparent utility of 2, and *B* does six operations, with utility of 2. Again, *A's* actual utility is 6.

There is a relation here between the two types of lies given above, and the "hiding letter" and "phantom letter" lies in the Postmen Domain. An agent in the Blocks World has the option of relaxing his true goal when he lies; the set of states that will satisfy his relaxed goal is then a superset of the set of states satisfying his true goal.[4] This is analogous to hiding a letter, and is what

[4] An agent will never benefit from pretending that the states satisfying his goals are a *subset* of the true goal states, since that would increase his apparent cost, and worsen his position in the negotiation.

agent *A* did when he said that any block under the Black block would be fine, even though he really wanted a Red block there. Consider Figure 5, where agent A'S expanded apparent goal states are represented by the thicker oval and labeled G'_G.

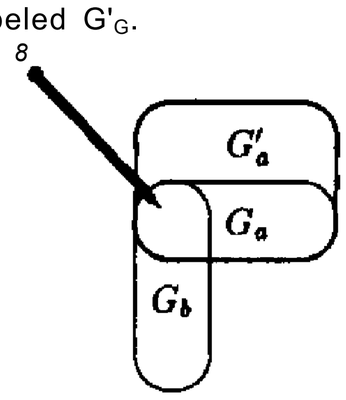

Figure 5: Expanding Apparent Goal States with a Lie

Note that the expansion of the goal states is toward the initial state *s.* This is the meaning of lowering one's apparent cost, and is necessary for a beneficial lie.

Alternatively, the agent can manufacture a totally different goal for the purposes of reducing his apparent cost, analogous to creating a phantom letter. Agent *A* did this when he said he wanted slot 3 empty and the Black block clear. Consider Figure 6, where agent ,4's altered apparent goal states are again represented by the thick outline and labeled G'_a. Note again, that the expansion of the goal states is toward the initial state *s.*

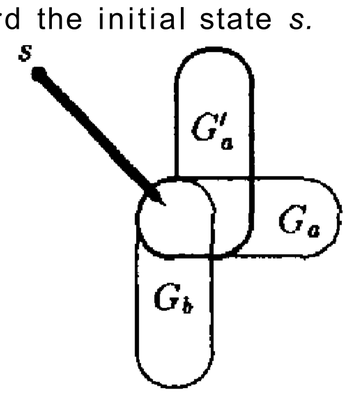

Figure 6: Altering Apparent Goal States with a Lie

The agent then needs to make sure that the intersection of his apparent goal states and his true goal states is not empty. Although this is a necessary precondition for a successful lie, it is of course not a sufficient precondition for a successful lie. Both of the lies in the above example will be useful to agent *A* regardless of the negotiation protocol that is being used: pure deal, mixed deal, semi-cooperative deal, or multi-plan deal.

Lying about the goal set (either expanding it or doing some other arbitrary alteration) is particularly useful when the domain causes tight coupling of the agents' goals. As was seen in the Postmen Domain, the likelihood of finding a beneficial lie grows with the coupling of the two agents' goals. With a phantom goal set, for example, things may have to be (usefully) changed in the real world in order to satisfy the phantom goals. At this point in our research, however, we have no insights regarding how an agent might discover either kind of beneficial lie systematically, For example, lowering the apparent cost of an agent's goal will not necessarily result in a beneficial He—the agent must also ensure that the agreed-upon final state will both satisfy his own goal and not be too expensive (by creating an apparent conflict with the other agent's goal).

Another way of viewing this is that beneficial lies exist in the Blocks World because goals can be "accidentally" achieved, and a lying agent can take advantage of this fact. Accidental achievement may come about both because the agents might share the same goat, and because tight coupling in the domain may cause the goal to be satisfied when some unrelated action is carried out. This is not the case in the Postmen Domain—no one can accidentally deliver your letter. Hiding letters, therefore, rules out accidental achievement *of those* hidden goals, and therefore the lying agent must carry out those goals by himself, to his detriment. In the Blocks World, you can hide part of your goal, and still see it achieved unwittingly by the other agent.

3,3 Lies in Conflict Situations

It might seem that when agents are in a conflict situation, the potential for beneficial lies is reduced. After all, it might appear that the agents' conflict is related to their goals being decoupled, and according to our previous observation, coupled goals aid beneficial lying. In fact, beneficial lying can exist in conflict situations, because conflicting goals do not mean decoupled goals.

"Conflict'' between agents' goals means that there does not exist a mixed joint plan that achieves both goals and is also individual rational. This is either because such a state does not exist, or because the joint plan is too costly to be individual rational. Even when conflict exists between goals, they may be tightly coupled, and therefore a beneficial lie may exist.

Taking Advantage of a Common Subgoal in a Conflict Situation: Let the initial state of the world be as in the left side of Figure 7. One agent wants the block currently in slot 1 to be in slot 2; the other agent wants it to be in slot 3. In addition, both agents share the goal of swapping the two blocks currently in slot 4 (i.e., reverse the stack's order). Formally, the goals are g_A = {At(R, 2), At(W,4), On(B, *W)}* and g_B = {At(R,3),At(IV,4),On(i?,^)}.

The cost for an agent of achieving his goal alone is 10. Negotiating over the true goals would lead the agents to agree to do the swap cooperatively (at cost of 2 each), achieving the state shown on the right of Figure 7, and *then* flip a coin, with a weighting of i, to decide whose goal will be individually satisfied. This deal brings them an overall expected utility of 2 (i.e., 1/2(10 - 2) - 2).

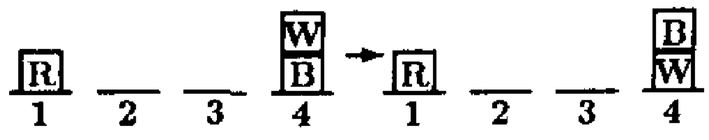

Figure 7: Taking Advantage of a Common Subgoal

What if agent *A* lies and tells *B* that his goal is g'_A = {At(R,2),On(B,W)}? Agent *A* thus "hides" the fact that his real goal has the stack of blocks at slot 4, and claims that he does not really care where the stack is. The cost for agent *A* of achieving his apparent goal is 6, because now he can supposedly build the reversed stack at slot 3 with a cost of 4. Assuming that agent *B* reveals his true goal, the agents will still agree to cooperatively bring the world to the same state shown on the right of Figure 7, but now the weighting of the coin will be 4/7. This deal would give agent *A* an apparent utility

of $1\frac{3}{7}$ (i.e., $\frac{4}{7}(8-2)-2$) which is also B's real utility (i.e., $\frac{3}{7}(10-2)-2$). A's real utility, however, is $2\frac{4}{7} = \frac{4}{7}(10-2)-2$. This lie is beneficial for A.

The situation is illustrated in Figure 8, where A's lie modifies his presumed goal states so that they are closer to the initial state, but the plan still ends up bringing the world to one of his *real* goal states.

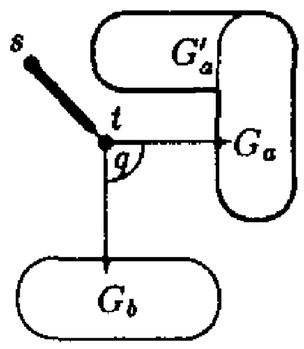

Figure 8: Lying in a Conflict Situation

In the example above, the existence of a common subgoal between the agents allowed one agent to exploit the coupled goals. The lying agent relaxes his true goal by claiming that the common subgoal is mainly its opponent's demand—as far as he is concerned (he claims), he would be satisfied with a much cheaper subgoal. If it is really necessary to achieve the expensive subgoal (he claims), more of the burden must fall on his opponent.

One might think that in the absence of such a common subgoal, there would be no opportunity for one agent to beneficially lie to the other. This is only partially true. When the goals are decoupled and the agents are negotiating over semi-cooperative deals, then there does *not exist* an *intermediate state t to which the agents* would agree to cooperatively bring the world. All the deals that the agents can agree on will have t — s, and will look like Figure 9. In this case, the utility for both agents of any deal (i.e., any value of q) will be 0.[5]

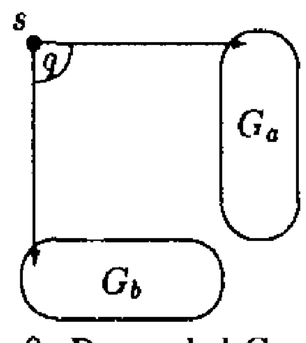

Figure 9: Decoupled Goals

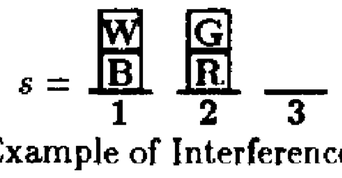

Figure 10: Example of Interference Decoy Lie

Another Example of Beneficial Lying in a Conflict Situation: The initial state s can be seen in Figure 10. A's goal is to reverse the blocks in slot 1, and to leave the blocks in slot 2 in their initial position, B's goal is to reverse the blocks in slot 2, and to leave the blocks in slot 1 in their initial position. To achieve his goal alone, each agent needs to do at least 8 PickUp/PutDown operations. This is a conflict situation.

If the agents negotiate over semi-cooperative deals, they will agree on $(s, \Lambda, \frac{1}{2})$ (A is the empty joint plan that does nothing and costs 0). There does not exist an intermediate state (other than 8) to which the agents will agree to cooperatively bring the world. The utility for each agent from this deal is 0 $(\text{or } \frac{1}{2}\epsilon)$. However, this is not the only kind of deal that the agents can agree on in this situation. If the agents are using multi-plan deals, then they can agree on a better deal.

Definition 1 *A* multi-plan deal *is* (δ_A, δ_B, q), *where the Si are mixed joint plans, and* $0 \leq q \leq 1 \in \mathbb{R}$, *is the probability that the agents will perform 6A (they will perform* 8_B *with probability* 1 — *q).*

Negotiation over multi-plan deals will cause the agents to agree on $(\delta_A, \delta_B)_{\frac{1}{2}}$, where δ_i is the mixed joint plan in which both agents cooperatively achieve i's goal. The best joint plan for doing the reverse in either one of the slots costs 2 Pick Up/Put Down operations for each agent. Each agent's utility from this deal is $2 = (\frac{1}{2}(8-2)-\frac{1}{2}(2))$.

Agent A might lie and claim that his goal is to reverse the blocks in slot 1 and leave the blocks in slot 2 in their initial position (his real goal) *OR* to have W be alone in slot 2. It costs A 6 to achieve his apparent goal alone; to do the reverse alone would cost him 8, and thus to achieve the imaginary part of his goal is cheaper. The agreement will be $(\delta_A, \delta_B)_{\frac{4}{7}}$, where δ_i is again the mixed joint plan in which both agents cooperatively achieve i's goal. It turns out to be cheaper for both agents to cooperatively carry out A's real goal than it is to cope with A's imaginary alternative. A'S apparent utility will be $1\frac{3}{7} = \frac{4}{7}(6-2) - \frac{3}{7}(2)$; this is also B's utility. A'S actual utility, however, will be $2\frac{4}{7} = \frac{4}{7}(8-2) - \frac{3}{7}(2)$, which is greater than the unvarnished utility of 2 that A would get without lying. So even without a common subgoal, A had a beneficial lie, but only because of the multi-plan deal protocol. Here we have been introduced to a new type of lie, a kind of "interference decoy," that can be used even when the agents' goals are decoupled.

4 Conclusions

Incomplete information between negotiating agents does not require a new concept of "solution." We can continue to use the concept of solution associated with complete information negotiation.

The existence of beneficial lies during negotiation can be sensitive both to the domain, as well as to the type of negotiation protocol used. In the Postmen Domain, for example, beneficial lies exist when agents negotiate over pure deals, but are eliminated when a more general deal type (i.e., all-or-nothing mixed deals) is used. In the Blocks Domain, however, the lies that are facilitated by tightly coupled goals are not eliminated by using a more general type of deal. Moreover, using an extremely

[5]This is a borderline situation in which the agents are indifferent between achieving their goals and leaving the world in its initial state. It still seems reasonable to assume that an agent would prefer to achieve his goal even though technically his utility is 0. To overcome this problem, we can assume that the worth of a goal to an agent is equal to the cost of this goal plus E (an infinitesimally small number that is the same for all agents). An agent would prefer to achieve his goal in this case because his utility would be t instead of 0, and q will always be equal to $\frac{1}{2}$.

general negotiation over multi-plan deals actually opens up the space of lies; the new "interference decoy" kind of lie can be used with multi-plan deals even when the agents' goals are decoupled.

Using goal recognition techniques will not eliminate lying by agents. The best we can hope for from any such approach is to enable the agents to reach pareto optimal deals (exactly the same deals that they would have reached with complete information). Therefore, regardless of the protocol being used, an agent could always negotiate as if he had a different goal. If a beneficial lie exists, he can still benefit from the deception, even if it is an "implicit" as opposed to explicit deception. Since, however, it may be computationally difficult to find beneficial lies (or even impossible given symmetric incomplete information between the agents), our negotiation protocols may still be usable in real-world situations. Beneficial lies may exist, but be difficult or impossible to find.

5 Acknowledgments

This research has been partially supported by the Leibniz Center for Research in Computer Science, by the Israel National Council for Research and Development (Grant 032-8284), and by the Center for Science Absorption, Office of Aliya Absorption, the State of Israel.

References

[Axelrod, 1984] Robert Axelrod. *The Evolution of Cooperation.* Basic Books, Inc., New York, 1984.

[Conry *et ai,* 1988] Susan E. Conry, Robert A. Meyer, and Victor R. Lesser. Multistage negotiation in distributed planning. In Alan H. Bond and Leg Gasser, editors, *Readings in Distributed Artificial Intelligence,* pages 367-384. Morgan Kaufmann Publishers, Inc., San Mateo, California, 1988.

[Durfee, 1988] Edmund H. Durfee. *Coordination of Distributed Problem Solvers.* Kluwer Academic Publishers, Boston, 1988.

[Ephrati and Rosenschein, 1991] Eithan Ephrati and Jeffrey S. Rosenschein. The Clarke Tax as a consensus mechanism among automated agents. In *Proceedings of the Ninth National Conference on Artificial Intelligence,* Anaheim, California, July 1991.

[Gasser, 199l] Les Gasser. Social conceptions of knowledge and action: DAI foundations and open systems semantics. *Artificial Intelligence,* 47(1-3): 107-138, 1991.

[Kraus and Wilkenfeld, 1991] Sarit Kraus and Jonathan Wilkenfeld. Negotiations over time in a multi agent environment: Preliminary report. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence,* Sydney, Australia, August 1991.

[Kreifelts and von Martial, 1990] Thomas Kreifelts and Frank von Martial, A negotiation framework for autonomous agents. In *Proceedings of the Second European Workshop on Modeling Autonomous Agents and Multi-Agent Worlds,* pages 169-182, Saint-Quentin en Yvelines, France, August 1990.

[Kuwabara and Lesser, 1989] Kazuhiro Kuwabara and Victor R. Lesser. Extended protocol for multistage negotiation. In *Proceedings of the Ninth Workshop on Distributed Artificial Intelligence,* pages 129-161, Rosario, Washington, September 1989.

[Laasri *et* al., 1990] B. Laasri, H. Laasri, and V- R. Lesser. Negotiation and its role in cooperative distributed problem problem solving. In *Proceedings of the Tenth International Workshop on Distributed Artificial Intelligence,* Bandera, Texas, October 1990.

[Malone *et* al., 1988] T. Malone, R. Fikes, and M. Howard. Enterprise: A market-like task scheduler for distributed computing environments. In B. A. Huberman, editor, *The Ecology of Computation.* North-Holland Publishing Company, Amsterdam, 1988.

[Rosenschein and Genesereth, 1985] Jeffrey S. Rosenschein and Michael R. Genesereth. Deals among rational agents. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence,* pages 91-99, Los Angeles, California, August 1985,

[Smith, 1978] Reid G. Smith. *A Framework for Problem Solving in a Distributed Processtng Environment* PhD thesis, Stanford University, 1978.

[Sycara, 1988} Katia P. Sycara. Resolving goal conflicts via negotiation. In *Proceedings of the Seventh National Conference on Artificial Intelligence*, pages 245-250, St, Paul, Minnesota, August 1988.

[Zlotkin and Rosenschein, 1989] Gilad Zlotkin and Jeffrey S. Rosenschein. Negotiation and task sharing among autonomous agents in cooperative domains. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pages 912-917, Detroit, Michigan, August 1989.

[Zlotkin and Rosenschein, 1990a] Gilad Zlotkin and Jeffrey S. Rosenschein. Blocks, lies, and postal freight: The nature of deception in negotiation. In *Proceedings of the Tenth International Workshop on Distributed Artificial Intelligence,* Bandera, Texas, October 1990.

[Zlotkin and Rosenschein, 1990b] Gilad Zlotkin and Jeffrey S. Rosenschein. Negotiation and conflict resolution in non-cooperative domains. In *Proceedings of the Eighth National Conference on Artificial Intelligence,* pages 100-105, Boston, Massachusetts, July 1990.

[Zlotkin and Rosenschein, 1990c] Gilad Zlotkin and Jeffrey S. Rosenschein. Negotiation and goal relaxation. In *Proceedings of The Workshop on Modelling Autonomous Agents in a Multi-Agent World,* pages 115-132, Saint-Quentin en Yvelines, France, August 1990.