# Beliefs, Reasons and Moves in a Model for Argumentative Dialogues

Sandra Zabala, Inés Lara and Héctor Geffner
Departamento de Computación
Universidad Simón Bolívar
Ap. 89000, Baruta, Caracas 1080-A, Venezuela
{ilara,szabala,hector}@ldc.usb.ve
Phone: (+582) 9063241/3242, Fax: (+582) 9063243

**Abstract**

This paper presents a model for dialogues in which the agents involved try to reach agreement or find grounds for disagreement regarding a particular claim. The model is based on the agents' belief systems, their arguments, and their interventions (moves). The agents **beliefs** include beliefs about the world and about the other agent's beliefs. They are expressed by means of facts, preferences and causal dependences. The agents' **arguments** are extracted from their beliefs and support or rebut the claim under discussion. The **moves** are knowledge communication acts made by the agents during the interaction aimed to solve differences between their positions using arguments. Our objective is to understand how and why agents give reasons for supporting or rebutting their positions and use such understanding to generate artificial dialogues automatically.

*Keywords*: Dialogue Modeling, Argumentation, Multiagents.

## 1 Introduction

In this paper we define a formal framework for modeling a particular class of dialogue in which the agents involved try to reach agreement or find grounds for disagreement regarding a particular claim. The type of claim that we deal in this paper is of the form *certain action is good* or *certain action is bad* For example, in Dialogue 1.1 (adapted from [12]) the claim under discussion is whether *capital punishment* is good or bad. One of the participants (Jim) is in favor of it, the other participant (Tom) opposes.

**Dialogue 1.1**

(1) Tom : This morning I heard an announcer describing the execution of a guy in Texas who raped and murdered a teen-aged couple.
(2) Jim : Well, he deserved it.
(3) Tom : Why?
(4) Jim : He didn't show much pity for his victims, did he?.
(5) Tom : Okay, but no matter what he did, capital punishment is really awful, barbaric. It's murder even if the State does it.
(6) Jim : No, I'd call it justice.
(7) Tom : I'm sorry but I think murder cannot be justice!
(8) Jim : Well, that's your opinion...

The proposed model for argumentative dialogues has three main features that distinguish it from other models of dialogues such as those proposed by Traum in [14] (the TRAINS-93 dialogue model), by Jokinen in [8] (Constructive Dialogue Model) and by Chu-Carroll and Carberry in [5] (a model based on Propose-Evaluate-Modify cycle of actions). First, it accommodates **beliefs** about the world and about the other agent. Second, it accommodates **arguments** extracted from such beliefs. Third, it provides a rationale for the dialogue moves by which the agents try to solve their differences.

**Beliefs** include beliefs about the world and beliefs about the other agent, and both are expressed by means of facts, preferences and causal dependences. In Dialogue 1.1, for instance, we might deduce from sentence (7) that Tom thinks that the action *capital punishment* leads to the negative goal *murder*, so this relation is assumed to be part of his belief system.

**Arguments** are extracted from the agents' beliefs and support or rebut the claim under discussion. We consider that arguments are composed of a claim and a support which is a slight modification to the model presented by Toulmin in [13]. In Dialogue 1.1, Tom has an argument whose claim *capital punishment is bad* is supported by the belief *capital punishment is murder*. In addition, arguments have a polarity, determining whether the argument supports or refutes the claim, and a strength, given by the strength of the 'weakest' support in the argument.

**Moves** are based on the agents' arguments and they are aimed to: (1) explain the differences between the agents' positions; (2) seek to understand these differences; (3) accept or reject a given explanation; and (4) provide information. In our framework moves are expressed in a meta-language that represents the structure of the agents' interventions. The meta-language is composed of expressions such as INFORM <*belief*> and <*claim*> BECAUSE <*support*>, which can be used by the agents for both supporting their positions or rebutting the other's positions. Figure 1 shows the structure of Dialogue 1.1 using this meta-language.

Our goal with this model of dialogue is to **understand** how and why agents give reasons for supporting their positions and use such understanding to **generate** artificial dialogues automatically.

The paper is organized as follows. First we present the representation of the agents' beliefs (Section 2), arguments (Section 3) and moves (Section 4). Then, we describe a prototype we have implemented (Section 5), discuss relations to other works and point to the main contributions (Section 6).

## 2 Modeling Beliefs

Beliefs are the basis of our model because arguments are extracted from them and dialogue moves are based on arguments. The belief systems include the agents' beliefs (their **own models**) and the perception the agent has about the other's beliefs (the **other's model**). For simplicity, we assume that initially both

(1) Tom : INFORM *Texan's execution testimony* is believed
(2) Jim : *capital punishment is good*
BECAUSE *justice is a strong reason for it*
(3) Tom : WHY *is justice a strong reason for capital punishment?*
(4) Jim : *justice is a strong reason for it*
BECAUSE *horrible crime* is believed.
ISN'T *horrible crime* believed?
(5) Tom : YES, *horrible crime* is believed
BUT *capital punishment implies murder* is believed
INFORM *capital punishment is bad*
(6) Jim : INFORM *capital punishment implies murder* is disbelieved.
INFORM *capital punishment applied to a terrible killer implies justice* is believed
(7) Tom : INFORM *capital punishment applied to a terrible killer implies justice* is disbelieved
INFORM *murder implies not justice* is believed
(8) Jim : INFORM *murder implies not justice* is disbelieved
DISAGREE

Figure 1: Structure of Dialogue 1.1

models are equal, i.e. agents assume that everybody shares their beliefs. We could discard this assumption by incorporating previous knowledge that one agent has about others; for example, one person in favor of abortion rights might initially believe that the other opposes it because he/she is Catholic.

Beliefs are represented by conditions, testimonies, actions, goals. These elements and their dependences are represented by means of a **Cognitive Map**. A cognitive map is a directed acyclic signed graph comprised of nodes (variables) and hyperlinks (links that go from a set of variables to another variable or link). Our cognitive maps are an extension of the cognitive maps presented by Axelrod in [2]. For the semantics of cognitive maps in terms of qualitative decision-theoretic models see [7] and [16].

**Nodes** represent boolean variables that can stand for either conditions, testimonies, goals or actions. **Conditions** are extracted from the initial view the agent has about the world. **Testimonies** usually represent a case that provides evidence concerning the value of a given element. Conditions and testimonies have associated a *belief measure* that can be *believed*, *plausible* or *disbelieved*. Their value is *plausible* by default. In Dialogue 1.1, the condition *terrible killer* and the *Texan's case* testimony are both *believed* by Tom and Jim. **Goals** represent the agent's preferences. They have associated a *polarity* and a *priority* indicating the kind of goals they are (*positive* or *negative*) and their level of importance (a non negative integer). In our example, *murder* represents a negative goal and *justice* a positive one. Regarding the single **action**, under discussion it does not have a predetermined value (good or bad) because it is deduced from its relation to the other elements in the cognitive map.

**Hyperlinks**, or simply links, represent the causal dependences among the variables. Links have *signs*: a positive link $A \xrightarrow{+} B$ means that if A is true then B is true and that if A is false then B is false; a negative link $A \xrightarrow{-} B$ means that if A is true then B is false and vice versa. By default, links are *believed* by all agents and they are positive. There are two kinds of links: causal and evidential. The distinction is made to avoid chaining causal links with evidential ones. The negative implications of such chaining are explained by Pearl in [11].

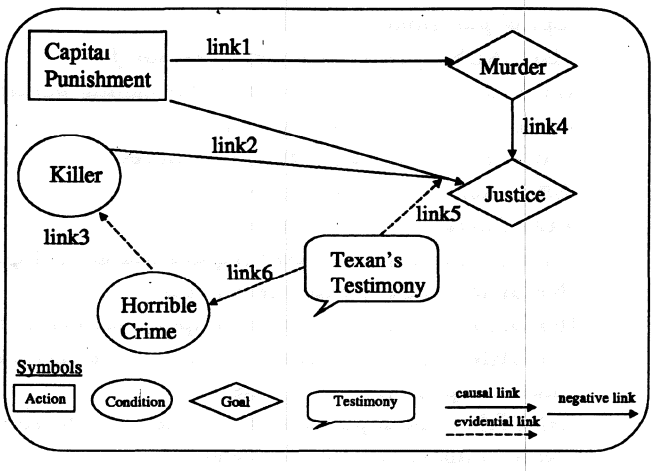


Figure 2: Cognitive Map of Dialogue 1.1

The beliefs underlying Dialogue 1.1 can be represented by the cognitive map presented in Figure 2. Links in that map can be expressed in natural language as follows:

Link 1 : Capital punishment implies murder
Link 2 : Capital punishment applied to a terrible killer implies justice
Link 3 : Only terrible killers commit horrible crimes
Link 4 : Murder implies injustice
Link 5 : Only if Link 2 is valid, justice would be achieved in the Texan's case
Link 6 : A horrible crime was committed in the Texan's case

Tom and Jim's belief systems can be completed by the information provided in Figure 3. It is important to note that in this case both agents share the map structure. However, one agent's map might have nodes or links which are not in the other.

## 3 Arguments

Once the agents' beliefs are defined, the next step is to show how they are used to build arguments. The definition of arguments is based on the model presented in [3] as a qualitative framework for decision making. We have adapted this model to build arguments supporting claims such us *Capital punishment is good* or *Abortion is bad.* Given a positive link $A \xrightarrow{+} B$, we say that:

- $A$ provides a **strong** support for $B$ if (1) the link is believed, (2) $A$ is believed or there is a strong support for $A$ and (3) $A$ does not give a strong support for $\neg B$.
- $A$ provides a **weak** support for $B$ if (1) it does not provides a strong support for $B$, (2) the link is believed or plausible, (3) $A$ is believed or plausible or there is a *strong* or *weak* support for $A$ and

| Tom | |
|---|---|
| **Element** | **Value** |
| **Conditions** | |
| killer | bel |
| horrible-crime | bel |
| **Testimonies** | |
| Texan's | bel |
| **Links** | |
| Link 1 | bel |
| Link 2 | disbel |
| Link 3 | bel |
| Link 4 | bel |
| Link 5 | disbel |
| **Goals** | |
| murder | -1 |
| justice | +1 |

| Jim | |
|---|---|
| **Element** | **Value** |
| **Conditions** | |
| killer | bel |
| horrible-crime | bel |
| **Testimonies** | |
| Texan's | bel |
| **Links** | |
| Link1 | disbel |
| Link 2 | bel |
| Link 3 | bel |
| Link 4 | disbel |
| Link 5 | bel |
| **Goals** | |
| murder | -1 |
| justice | +2 |

Figure 3: Initial Belief Systems of Dialogue 1.1

(4) $A$ does not give a strong or a weak support for $\neg B$.

- $A$ provides an **empty** support for $B$ if it does not provides a strong nor a weak support for it.

If the link is negative, $A \dashrightarrow B$, definitions are analogous but the support is for $\neg A$. In general, links are chained in a simple way as detailed in [3].

In our case, agents are discussing the goodness or badness of a particular action. This is why it is important to define the reasons that might justify such judgement. We consider goals as such reasons. From the framework presented in [3], we define the strength of the reason that a goal G gives for an action A as the strength of the support that A gives for G.

An action A is deemed *good* by an agent if and only if for each support that $A$ provides to a negative goal $G^-$, there is a support for a positive goal $G^+$ of higher priority, that is as strong as the support for $G^-$. Similarly, $A$ is deemed *bad* when the roles for the positive and negative goals $G^+$ and $G^-$ are exchanged. If neither set of conditions hold, then $A$ is neither good nor bad.

From these definitions and the information given in Figure 3 we may build the arguments supporting the position of the agents participating in Dialogue 1.1. For example, in Jim's belief system the positive goal justice is a strong reason for *Capital punishment* and there is no negative goal offering a strong or weak reason against it. As a result *Capital punishment* deemed a good action by Jim.

## 4 Interaction

In this section we show how the arguments defined in the previous section are used during the dialogue process to support the agent's position or to rebut the other's position.

The dialogue is constituted by a sequence of moves which are knowledge communication acts based on the agents' arguments. We consider that each agent has to follow certain principles for moving and he/she also assumes that the other is following them too. The principles we are assuming are: (1) Honesty, which implies that the agent has to believe what he/she is saying; (2) Conciseness, which means that agents

can only say few things at a time and those things must carry maximum information; (3) Collaboration, which means that both agents work together to accomplish the dialogue task by answering questions, giving solicited supports and so on.

The dialogue task is to reach agreement or find the grounds for disagreement regarding a particular claim. This task is accomplished by a sequence of moves in which agents:

- **Explain the differences between their positions**: From the previous move the agent can deduce that there is an element in the other's beliefs which is different from his/hers. Then, the agent can give an argument in favor of his/her belief or against the other's belief. For example, if the previous move was *Capital punishment is bad*, a possible next move would be *Capital punishment is good because it leads to justice.*

- **Seek to understand differences**: The agent does not know how to fit the last move information into the other's model. Then, the agent proceeds to ask for a support or a belief measure. For example, if the previous move was *Capital punishment is bad*, a possible next move would be *Isn't the accused a terrible killer?.*

- **Provide information**: In response to a direct question. For example, the answer to the question *Why is capital punishment bad?* could be *Because it is murder.*

- **Accept or reject a given explanation**: In the previous move the agent gave a support which is accepted by the other but it is considered irrelevant in relation to the claim. Even if the agent accepts the support, the differences between the agents' positions still remain. Then, he/she gives an explanation to support his/her position or rebut the other's. For example, an answer to the question *Isn't he a terrible killer?* could be *Yes he is a terrible killer, but capital punishment is murder.* A special case of this intention is **expressing agreement or disagreement**: If the agents end up sharing the same position, agreement has to be expressed. If they do not have any more arguments to support their own position or to rebut the other's and there are no more questions to formulate, agents may consider that they established enough grounds for disagreement. In either case the dialogue is finished.

The precise form of the moves is detailed in Figure 4.

Once we have defined the set of moves, it is necessary to describe the interaction main processes: revision and selection, which are carried out until the dialogue is over.

The **revision** can be sketched as follows:

- Revising the belief system:

  The agent's own model only can be modified after a move which adds a new piece of information. It happens when the other agent's map has nodes or links which are not in the listener's belief system. The listener incorporates this new piece of information into his/her map and assigns a belief measure to it. The other's model can be modified when: (1) the last move added a new piece of information to the listener's belief system and the revision is similar to his/her own model revision; and (2) the listener detects a difference between the information transmitted in the last move and the information registered in the other's model and updates it with this information. Sometimes it is not possible to accommodate a given information into the other's model, because it is not clear how the speaker deduced such information. The agent will have to ask for more information to try to accommodate it in the model.

| Move | Description |
|---|---|
| WHY? | The agent asks for the support of an affirmation given by the other |
| IS/ISN'T? | The agents attempts to confirm if the other has the belief that he/she thinks he/she has |
| YES/NO | The agents answer to a confirmation seeking |
| INFORM | The agent gives some variable values or claims without giving support for them |
| BECAUSE | The agent gives an argument (a claim with its support) |
| YES-BUT | The YES-BUT move, accepts a given information, but at the same time implicitly says that there are stronger reasons that support a different position |
| AGREE/ DISAGREE | This moves are related to the acceptance of the other's interventions. It also indicates how the task of the dialogue was achieved: reaching agreement or finding grounds for disagreement |

Figure 4: Moves considered in the model

- Rebuilding arguments:

  It happens only if there was a change in the belief system. Arguments are reconstructed from the belief system new state following the definitions given in the previous section.

The **selection** of a move depends on the listener's belief system, his/her arguments and pending request for information.

- Analyzing positions:

  If both agents have the same position the next move will be to express agreement.

- Analyzing arguments:

  (1) If the agents don't share a position and there are no more arguments to support their own position or to rebut the other's, and there are no more questions to ask, the next move is to express that enough grounds for disagreement; (2) If there are elements (beliefs or arguments) in the other's model that could not be accommodated by the revision process, the next move might be to seek information.

- Analyzing the last move:

  (1) If it was to seek information, the listener looks for an answer in his/her beliefs or arguments, and the next move will provide that information; (2) Otherwise, the next move intention might be to explain differences between the agents' positions or accept/reject the speaker's explanation by using the strongest argument the listener has, which has not been used yet.

Let us illustrate how the revision and selection processes work after a YES-BUT move:

- If the revision process could not accommodate the BUT part in the other's model, the agent next move it to seek for information (WHY or IS/ISN'T).

| Previous Move | Next Move | Intention |
|---|---|---|
| Why | Because | Providing information |
| Yes-But | Why | Seeking information |
| | Is/Isnt | Idem |
| | Inform | Explicating differences |
| | Because | Idem |
| | Yes-but | Accepting/ rejecting |
| | Yes/No | Does not apply |
| | Disagree | Expressing disagreement/ |
| | Agree | agreement |
| Is-Isn't | Yes/No or Yes-But | Providing Information |

Figure 5: Possible Next Moves(for moves such as INFORM, BECAUSE and YES/NO, the table is analogous to the YES-BUT)

- If the revision process could accommodate the BUT part, and the listener does not share it because his/her related arguments have a different *polarity* or *strength*, he/she has to explicate this difference by supporting his/her own position or rebutting the other's (either INFORM or BECAUSE can be used to do this).
- If the revision process could accommodate the BUT part, the listener shares it but considers that there are stronger arguments that support a different position, he/she has to reject it (YES-BUT).
- If the revision process could accommodate the BUT part and it produced changes in the listener's own model which lead he/she to share the other's position, the next move is to express agreement.
- If the revision process could accommodate the BUT part, the listener still does not share the other's position but there are any more arguments or counterarguments to give and no more questions to ask, the next move is to express disagreement.

It illustrates that moves cannot be chained in arbitrary order; for example it is meaningless to answer *yes* to *why capital punishment is bad?*. The possible move sequences are given in Figure 5.

## 5 The Prototype

We have implemented a prototype which illustrates some of the functionality of the proposed model. The prototype is implemented in Prolog (BinProlog 5.75) on a Solaris 2.6 SUN platform.

The prototype implements the artificial dialogue generation. The general process is illustrated in Figure 6. It follows the process described in section 4 as *interaction*. The system inputs are the agents' belief systems and the claim under discussion (whether a particular action is good or bad) and the initial

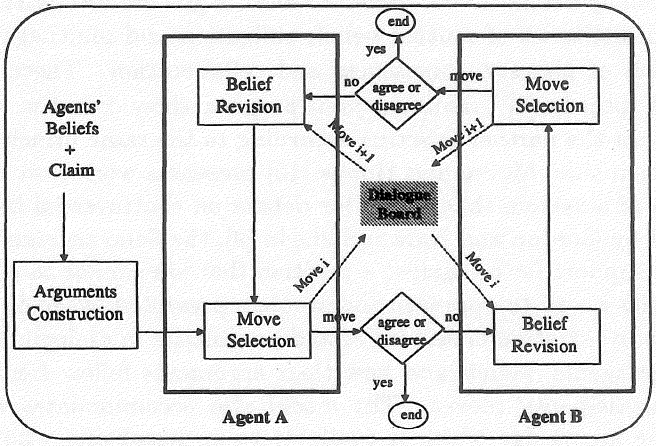


Figure 6: Dialogue Process in the Prototype

arguments are built from these beliefs. Then, one of the agents makes the first move expressing his/her position regarding the claim in the meta-language defined in the previous section. The other agent revises his/her belief system and selects his/her next move. From this moment on, the agents cycle revising their beliefs and selecting their moves until agreement or disagreement are established. The selected moves are registered in the *Dialogue Board* which keeps the dialogue history.

We have conducted tests generating dialogues regarding the claim *Capital punishment is good* from the belief systems we have shown in this paper. We have also tested the system with belief systems regarding the claim *Working full-time is good for women*, taken from [17]. Different dialogue structures have been generated for each topic, depending on who speaks first and which move is selected among the set of moves that are eligible in a given moment.

Given that we want to work with large cognitive maps, we are developing a system in Java with the belief systems stored in MiniSQL, what we hope is going to be more efficient than the Prolog version.

## 6 Discussion

Our framework combines features of the three approaches to modeling dialogues, discussed by Cohen in [6], section 6.3: dialogue grammars, plan-based models and joint action theories.

We have certain rules (like *dialogue grammars*) indicating what the next move in a dialogue might be. Our approach is *plan-based* in the sense that the agents share a common goal (to reach agreement or find grounds for disagreement) and given their belief system they select the move that best achieve it. Many proposals for dialogue modeling are goal-driven. One of these is the work presented by Bratman et al in [4] where they present the Beliefs/Desires/Intentions (BDI) architecture. Part of our model can be seen as a particular BDI architecture where conditions, testimonies and links are the agent's beliefs, goals and preferences represent desires and the achievement of the dialogue task is the common intention. In [15] and [14], Traum discusses the dialogue process in terms of the BDI architecture in which plans are revised

and repaired. It is functionally similar to our belief revision, argument reconstruction and move selection. The model includes social attitudes of mutual belief, obligation and multi-agent plan execution, which is related to our principles of honesty, conciseness and collaboration. These conversational principles also appear in a different form in [1]: assume a *joint purpose*, show cognitive consideration and ethical consideration and trust that the partner is acting according to the same principles.

From the argumentation side, Moore and Hobbs [10] present a work that is conceptually similar to ours, that takes the form of a system that allows for debate on controversial issues. Another interesting work is the one presented by Gordon and Karacapilidis in [9], the Zeno argumentation framework. It has messages which transmit support for or against a position that are similar to our moves .

Our framework provides a way to approach the problem presented by Scott and Kamp in [6], section 6.2, of achieving a genuine integration of semantics and pragmatics in dialogue modeling and generation. The representation of the agents' beliefs and how their arguments follow from them gives a semantics for the content of the participants' moves. The model also accommodates what they call *pragmatic information* regarding the knowledge which is available to an agent and what he/she wants, which is reflected by his/her intention when moving.

**Summary**

We consider that the presented model is useful to understand how and why agents may do moves during a particular class of dialogue. Beliefs are represented graphically and have associated qualitative plausibility measures. Arguments are extracted from the agents' belief systems and support or rebut the claim under discussion. Moves are aimed to solve differences between the agents' positions, using arguments and counterarguments.

The understanding of the dialogue elements and their use during the interaction allows the construction of systems to generate artificial dialogues, which is illustrated by the prototype we developed.

## References

[1] J. Alwood. Linguistic communication as action and cooperation. Technical Report Gothenburg Monographs in Linguistics 2, University of Göteborg, Department of Linguistics, 1976.

[2] R. Axelrod. *The Structure of Decision*, chapter 3. The analysis of Cognitive Maps. 1972.

[3] B. Bonet and H. Geffner. Arguing for decisions: A qualitative modeling for decision making. In *Proceedings of the Twelfth Conference Uncertainty in AI*, pages 98–105, Portland, OR., 1996. Morgan Kaufmann.

[4] M. Bratman and D. Israel. Plans and resource-bounded parctical reasoning. *Computer Intelligence*, 4(4):349–355, 1988.

[5] J. Chu-Carroll and S. Carberry. Conflict detection and resolution in collaborative planning. In *Proceedings of the 2nd Intl. Workshop on Agents Theories, Architectures and Languages*, Lecture Notes, pages 111–126. Springer-Verlag, 1996.

[6] H. Uszkoreit (ed). *Survey of the State of the Art in Human Language Technology*, chapter 6: Discourse and Dialogue. 1997. Available as http://www.cse.ogi.ed/CSLU/HLTsurvey/download.htm.

[7] H. Geffner. A formal framework for causal modeling and argumentation. In *Proceedings of Intl. Conf. on Formal and Applied Practical Reasoning*, Germany, 1996. Springer-Verlag.

[8] K. Jokinen. Goal formulation based on communicative principles. In *Proceedings of COLING'96*, Copenhagen, 1996.

[9] N. Karacapilidis, D. Papadias, and T. Gordon. An argumentation based framework for defeasible and qualitative reasoning. In D. Borges and C. Kaestner, editors, *Proceedings of the 13th Brazilian Symposium on Artificial Intelligence*, Lecture Notes in Artificial Intelligence, pages 1–10. Springer-Verlag, October 1996.

[10] D. Moore and D. Hobbs. "Leeds Argumentation System"(LAS): towards an ITS for educational debate. In *Proceedings of Complex Learning in Computer Environments (CLCE'94)*, Joensuu, Finland, 1994. Available as http://cs.joensuu.fi/ mtuki/www_clce.270296/Moore.html.

[11] J. Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann, San Mateo, CA., 1988.

[12] A. Rottenberg. *Elements of argument.* Bedford Book-St. Martin's Press, New York, second edition edition, 1988.

[13] S. Toulmin. *The Uses of Arguments.* Cambridge University Press, 1958.

[14] D. Traum. Conversational agency: The trains-93 dialogue manager. In *Proceedings of the Tweente Workshop on Language Technology: Dialogue Management in Dialogue Management Systems*, pages 1–11, June 1996.

[15] D. Traum. A reactive-deliberative model of dialogue agency. In J. Muller, M. Wooldridge, and N. Jennings, editors, *Proceedings of the Third Intl. Workshop on Agents Theories, Architectures and Languages*, Lecture Notes in Artificial Intelligence, Heildelberg, 1997. Springer-Verlag.

[16] M. Wellman. Inference in Cognitive Maps. *Mathematics and Computers in Simulation*, 36:137–148, 1994.

[17] A. Wolfe. The Moral Meanings of Work. *The American Prospect*, (34):82–90, September-October 1997.