# The Eightfold Way of Deliberation Dialogue

Peter McBurney
Department of Computer Science
University of Liverpool
Liverpool L69 7ZF, UK
p.j.mcburney@csc.liv.ac.uk
David Hitchcock
Department of Philosophy
McMaster University
Hamilton Ontario
Canada L8S 4K1
hitchckd@mcmaster.ca
and
Simon Parsons
Center for Co-ordination Science
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA 02142, USA
sparsons@mit.edu

June 29, 2002

**Abstract**

Deliberation dialogues occur when two or more participants seek to jointly agree an action or a course of action in some situation. We present the first formal framework for such dialogues, grounding it in a theory of deliberative reasoning from the philosophy of argumentation. We further fully articulate the locutions and rules of a formal dialogue game for this model, so as to specify a protocol for deliberation dialogues. The resulting protocol is suitable for dialogues between computational entities, such as autonomous software agents. To assess our protocol we consider it against various records of human deliberations, against normative principles for the conduct of human dialogues, and with respect to the outcomes produced by dialogues undertaken according to the protocol.

## 1 Introduction

In an influential typology, argumentation theorists Doug Walton and Erik Krabbe [65] classified human dialogues according to the objectives of the dialogue, the objectives of the participants (which may differ from one another) and the information which each participant had available at commencement of the dialogue. This classification resulted in six primary dialogue types, as follows: **Information-Seeking Dialogues** are dialogues where one participant seeks the answer to some question(s) from another participant, who is believed by the first to know the answer(s). In **Inquiry Dialogues** the participants collaborate to answer some question or questions whose answers are not known to any one participant. **Persuasion Dialogues** involve one participant seeking to persuade another to accept a proposition he or she does not currently endorse. In **Negotiation Dialogues**, the participants bargain over the division of some scarce resource. Here, each participant may be seeking to maximize his or her share of the resource, in which case the individual goals of the participants are in conflict. Participants of **Deliberation Dialogues** collaborate to decide what action or course of action should be adopted in some situation. Here, participants share a responsibility to decide the course of action, or, at least, they share a willingness to discuss whether they have such a shared responsibility. In **Eristic Dialogues**, participants seek to vent perceived grievances, and the dialogue may be act as a substitute for physical fighting.[1]

Formal models of several of these dialogue types have been developed in recent years. For example, models have been proposed for: information-seeking dialogues [30]; inquiry dialogues [43]; persuasion dialogues [4, 65]; and negotiation dialogues [6, 30, 42, 59]. Moreover, since most real-world dialogues are in fact combinations of primary types, models have been proposed for complex combinations of primary dialogues, for example, iterated, sequential, parallel and embedded dialogues [44, 55]. However, to our knowledge, no formal model has yet been proposed for deliberation dialogues, and it is the purpose of this paper to present such a model. For this model, we draw upon a model of deliberation decision-making from the philosophy of argumentation, and we use a dialogue-game formalism to define an interaction protocol. Our protocol effectively creates a public space in which multiple participants may interact to jointly decide on a course of action, with the structure and rules of the protocol defining the nature of these interactions.

The paper is structured as follows: Section 2 explores the features of deliberation dialogues which are specific to this type of dialogue. Section 3 presents our formal model of deliberation dialogues, drawing on work in the philosophy of argumentation. This is followed, in Section 4, with a dialogue game formalism for deliberation dialogues which accords with the general model presented in Section 3. The full syntax of the dialogue game locutions and the rules governing their use, however, are presented in an Appendix. We then consider, in Section 5, how we may assess our protocol. Here we consider it against various records of human deliberations, against normative principles for the conduct of human dialogues, and with respect to the outcomes produced by dialogues undertaken according to the protocol. This is followed, in Section 6, with an example of the use of our formalism. The paper ends with a discussion of related and future research in Section 7.

Before presenting our model, however, there is one aspect of our work which it is important to emphasize. Although our approach is motivated by human deliberation dialogues, we seek in this paper to define a model for deliberation interactions only between computational entities, as such autonomous software agents. We use the term *dialogue* to refer to such interactions because they are analogous to human dialogues, and because they may serve similar, or even identical, purposes to human dialogues.[2] However, we are *not* seeking to model human deliberations, or to provide models for

[1]Since Eristic dialogues are not generally rule-governed, formal models of them may be difficult to develop. However, recent work by Dov Gabbay and John Woods has looked at dialogues involving non-cooperation and hostility by the participants [20].

[2]The agents, for example, may be acting on behalf of human principals.

natural language explanation, generation or processing. Thus, this paper is not, and does not claim to be, a contribution to computational linguistics. This restricted focus has several implications for our work. Firstly, all utterances in a dialogue between agents conducted according to some protocol may be assumed to accord with the rules of that protocol; if one participant utters expressions invalid according to the rules of the protocol these will not be transmitted to the other participants. This is unlike the situation in human–human or human–machine interactions, where utterances which do not conform to the protocol syntax and combination rules may indeed be transmitted, resulting in considerable efforts being expended by listeners attempting to parse or to understand them. Secondly, we can assume that agents participating in a dialogue do so of their own free volition and may leave at any time. This contrasts with at least one model of human–human dialogues, that of Paul Grice [22, p. 48], where a conversation between two parties can only end when *both* parties agree to its termination. We believe our model is more appropriate for an open computational society of autonomous software agents. Thirdly, our assumption of agent autonomy leads us to assume that agents will not enter any dialogue unless and until they perceive this to be in their self-interest (however conceived by the agent concerned) to do so. In particular, agents will require, before entry, a statement of the intended topic of discussion in the dialogue. Unlike in many human dialogues, agents should not need to infer this from the utterances of others in the course of the dialogue.

## 2 Deliberation Dialogues

What distinguishes deliberation dialogues from other types of dialogue? A first characteristic arises from the focus of a deliberation, which concerns what is to be done in some situation by someone, either an individual or a group of individuals. This focus on action distinguishes deliberation dialogues from inquiry, information-seeking and eristic dialogues, although not from persuasion and negotiation dialogues; these latter two may also be about action. Moreover, information-seeking and inquiry dialogues involve a search for the true answer to some factual question, either by one participant or by all. In such a search for truth, appeals to value assumptions (goals, preferences, etc) would be inappropriate. However, this is not the case for deliberations, where a course of action may be selected on the basis of such considerations.

A second characteristic of deliberation dialogues is the absence of a fixed initial commitment by any participant on the basic question of the dialogue. Although the participants may express individual positions about what is to be done, the discussion is a mutual one directed at reaching a joint decision over a course of action; the actions under consideration, however, need not be joint, and may indeed be enacted by others not participating in the dialogue. A deliberation dialogue is not, at least not at its outset, an attempt by one participant to persuade any of the others to agree to an initially defined proposal. In this respect, deliberation dialogues differ from persuasion dialogues.

A third characteristic of deliberations relates to their mutual focus. Although the participants may evaluate proposed courses of actions according to different standards or criteria, these differences are not with respect to personal interests which they seek to accommodate in the resulting decision. In this respect, a deliberation dialogue differs from a negotiation dialogue, which must deal with reconciling potentially competing interests. In a negotiation, for example, it may be deleterious for a participant to share its information and preferences with others. But a sharing strategy should behoove participants in a deliberation; to the extent that agents are unwilling to share information or preferences, we would define their discussion to be a negotiation and not a deliberation.

These last two characteristics lead to an important observation about deliberation dialogues. An action-option which is optimal for the group when considered as a whole may be seen as sub-optimal from the perspective of each of the participants to the deliberation. This could be because a demonstration of optimality requires more information than is held by any one participant at the start of the dialogue, or because individual participants do not consider all the relevant criteria for assessment.[3] Similarly, an option for which the group has a compelling argument may be such that no one participant, on his or her own, has such an argument; only by pooling information or resources is the group able to construct a winning argument for the option. This characteristic means that an assumption of an individual sincerity condition on agent utterances (e.g., in the FIPA Agent Communications Language ACL [18]) may not be appropriate: with this condition the optimal option would never be proposed if no one participant has, on its own, a compelling argument for it. Moreover, real-life deliberations often benefit from whimsical or apparently-random proposals, which lead participants to discuss creative (*"off-the-wall"*) alternatives.

How do dialogues commence and proceed? Information-seeking dialogues, persuasions and inquiries each commence with a question or a statement by a participant and proceed by means of responses from other participants. Likewise, negotiation dialogues arise when a resource needs to be divided, and they can commence with a proposal by a participant to divide the resource in some manner, perhaps optimally for that participant. The negotiation will then proceed via responses to this proposal, including counter-proposals, and these responses, in the best case, converge on a mutually acceptable settlement. This is how auction and economic negotiation mechanisms, such as the monotonic concession protocol [58, 60], are conducted; one may view these as protocols for negotiation dialogues with limitations on the nature and content of the permitted utterances.

A deliberation dialogue arises with a need for action in some circumstance. In general human discourse, this need may be initially expressed in governing questions which are quite open-ended, as in, *Where shall we go for dinner this evening?* or *How should we respond to the prospect of global warming?* Proposals for actions to address the expressed need may only arise late in a dialogue, after discussion on the governing question, and discussion on what considerations are relevant to its resolution. When possible courses of action are proposed, they may be evaluated on a large number of attributes, including: their direct or indirect costs and benefits; their opportunity costs; their consequences; their practical feasibility; their ethical, moral or legal implications; their resourcing implications; their likelihood of realization or of success; their conformance with other goals or strategies; their timing, duration or location; etc. Negotiations over multi-attribute outcomes share the characteristic of multi-dimensionality with deliberations.

To achieve resolution of a deliberation dialogue, one or more participants must make a proposal for an appropriate course of action. But where do such proposals for action arise? And how do the participants know when they have identified all the possible alternatives, or at least all those alternatives worth considering? These are not

[3]For example, as William Rehg has noted [56], one benefit of public discussion of proposed Governmental actions is that participants to the discussion learn about the consequences of action-options for others of which they were not, or even could not have been, previously aware. For this reason, decision processes which incorporate public discussion may produce better quality outcomes than those which do not, as argued in [17].

easy questions, for human or for machine deliberators.

## 3 A Formal Model of Deliberations

Guided by the considerations of the previous section, we now present a formal, high-level model for deliberation dialogues. Our work adopts a similar structure to the idealized, five-stage model for negotiation dialogues proposed by Joris Hulstijn [30].[4] We also draw on a domain-specific decision theory, the retroflexive argumentation model for non-deductive argument of Harald Wohlrapp [67]. This model talks of a matter-in-question, equivalent to a governing question or a proposal for action, being considered from a number of different *frames* or *perspectives*; we use the latter term, to avoid confusion with Reed's *Dialogue Frames* [55]. As mentioned above, perspectives may be factors such as moral implications, opportunity costs, etc. An argument for or against a particular option is a partial understanding of that option from one or more, but rarely all, perspectives. Having heard an argument for or against an option, Wohlrapp argues, one proceeds by re-examining the underlying assumptions or modifying the action proposal, in the light of that argument. Thus, an argument against a law permitting euthanasia may be that such practices are open to abuse of ill patients by malicious relatives. A retroflexive response to this argument is to modify the proposed law by adding restrictions which inhibit or preclude such abuses, such as a requirement that the patient be of sound mind and give prior consent to the act of euthanasia.

With Wohlrapp's model in mind, we assume that the subject-matter of dialogues can be represented in a propositional language, with propositions and propositional functions denoted by lower-case Roman letters, e.g. $p, q, \ldots$. We define the following types of propositions:

**Questions:** A question is a proposition, or a propositional function with one or more free variables (possibly conjoined with the proposition that exactly one sequence of objects satisfies the function), denoted by a lower-case Roman letter followed by a question-mark, e.g. $p$? A governing question is the overall issue or issues which motivated the participants to convene the particular deliberation dialogue.

**Actions:** An action is a proposition representing a deed or an act (possibly a speech act) which may be undertaken or recommended as a result of the deliberation dialogue. The purpose of the deliberation dialogue is to decide on an answer to the governing question, which will be some (course of) action. Possible actions are also called *action-options*.

**Goals:** A goal is a proposition representing a future world state (external to the dialogue), possibly arising following execution of one or more actions and desired by one or more participants. Goals express the purpose(s) for which actions are being considered in the dialogue.

**Constraints:** A constraint is a proposition expressing some limitation on the space of possible actions.

**Perspectives:** A perspective is a proposition representing a criterion by which a potential action may be evaluated by a participant.

[4] Hulstijn calls these negotiation dialogues *Transactions*.

**Facts:** A fact is a proposition expressing some possible state of affairs in the world external to the dialogue.

**Evaluations:** An evaluation is a proposition expressing an assessment of a possible action with respect to a goal, constraint or perspective.

These types are mutually exclusive. With these elements defined, we now present a formal model of the dialogue itself, a model which consists of eight stages:

**Open:** Opening of the deliberation dialogue, and the raising of a governing question about what is to be done.

**Inform:** Discussion of: (a) the governing question; (b) desirable goals; (c) any constraints on the possible actions which may be considered; (d) perspectives by which proposals may be evaluated; and (e) any premises (facts) relevant to this evaluation.

**Propose:** Suggesting of possible action-options appropriate to the governing question.

**Consider:** Commenting on proposals from various perspectives.

**Revise:** Revising of: (a) the governing question, (b) goals, (c) constraints, (d) perspectives, and/or (e) action-options in the light of the comments presented; and the undertaking of any information-gathering or fact-checking required for resolution. (Note that other types of dialogues, such as information seeking or persuasion, may be embedded in the deliberation dialogue at this stage.)

**Recommend:** Recommending an option for action, and acceptance or non-acceptance of this recommendation by each participant.

**Confirm:** Confirming acceptance of a recommended option by each participant. We have assumed that all participants must confirm their acceptance of a recommended option for normal termination.

**Close:** Closing of the deliberation dialogue.

This is a model of an ideal dialogue. The stages may occur in any order, and may be entered by participants as frequently as desired, subject only to the following constraints:

- The first stage in every dialogue is the **Open** stage. Once a second participant enters the dialogue, the dialogue is said to be *open.*
- The **Open** stage may occur only once in any deliberation dialogue. All other stages may occur more than once.
- The only stages which must occur in every dialogue which terminates normally are **Open** and **Close**.
- At least one instance of the **Inform** stage must precede the first instance of every other stage, excepting **Open** and **Close**.
- At least one instance of the **Propose** stage must precede the first instance of the **Consider, Revise, Recommend** and **Confirm** stages.
- The **Confirm** stage can only be entered following an instance of a **Recommend** stage.

- Upon successful completion of an instance of the **Confirm** stage, the dialogue must enter the **Close** stage.
- The last stage in every dialogue which terminates normally is the **Close** stage.
- Subject only to the constraints expressed in these rules and constraints expressed in the locution-combination rules (articulated below), participants may enter any stage from within any other stage at any time.

Some comments are appropriate on the rules constraining the order of stages. Firstly, the participants may enter a **Close** stage more than once in a particular dialogue. As the locution rules below will demonstrate, participants are required to indicate publicly that they wish to leave the dialogue. Whenever a participant does this, the dialogue enters a **Close** stage. However, the **Close** stage remains unconcluded, and the dialogue remains *open*, as long as there are at least two participants who wish to continue speaking. It is therefore possible for the **Close** stage, as with all the other stages except the **Open** stage, to be entered multiple times in any one dialogue.

Secondly, we have assumed for simplicity in this initial model that unanimity of the participants is required for a decision on a course of action to be made. It would be quite possible for the participants to adopt a different procedure for confirmation, such as majority voting or consensus procedures, as modeled formally in [31]. If alternative voting procedures were to be adopted, it would be useful to announce the results of any votes formally to the participants, with a statement of the group's decision, just as the minutes of human meetings usually record these. For this reason, we have demarcated a separate stage, **Confirm**, to record final commitments to action. In addition, the requirement that participants once again assert their endorsement for a particular course of action reinforces their commitment to this course as the group's decision. Once all participants have confirmed their acceptance of a recommended action, the dialogue must end, and any further discussion relevant to the same governing question can only occur by commencement of a new deliberation dialogue.

Apart from the constraints listed here, the order of stages is not fixed and participants may return to different stages multiple times in any one dialogue. Thus, a dialogue undertaken according to this model may cycle repeatedly through these stages, just as human dialogues do. In this way, the protocol here gives practical effect to Wohlrapp's model of retroflexive argumentation. The model is also quite general; we have not specified the nature of the governing questions, goals, constraints, facts, action-options, perspectives or evaluations. Nor have we specified here any particular mechanisms for producing, revising or accepting action-options.[5]

# 4 Locutions for a Deliberation Dialogue Protocol

We now articulate the locutions of a formal dialogue game which enables a deliberation dialogue to be conducted according to the eight-stage model just presented. Dialogue

[5] Wohlrapp's model of retroflexive argumentation and the formalization of it presented here have some similarities with Imre Lakatos' theory of mathematical discovery [36]. According to Lakatos, mathematicians work by proposing statements they believe may be theorems and then seeking proofs for these. In doing so, a counter-example to the proposed theorem may be found, which leads the mathematician to modify the proposal. A new attempt at seeking a proof is then undertaken, with the process repeated until such time as a theorem is identified for which a proof can be found. The theories of Lakatos and Wohlrapp may be seen as describing (in part) arguments which proceed by *precization*, in the terminology of Arne Naess [46].

games are interactions between two or more participants who "move" by uttering locutions, according to certain rules. They were first studied by Aristotle [7] and have been used in modern philosophy to understand fallacious arguments [27, 41] and to provide a game-theoretic semantics for formal logical systems [38]. Over the last decade they have been applied in various areas of computer science and artificial intelligence: for the specification of software systems with multiple stakeholders [16]; for the design of man-machine interfaces [8, 30]; for the analysis of complex human reasoning [54]; and for the design of interaction protocols for autonomous software agents [4, 12, 42].[6]

A dialogue game may be specified by listing the legal locutions, together with the rules which govern their use, and the commencement and termination of dialogues [44]. In this Section, we present only the locutions, and not also the necessary preconditions for, and the consequences of, their utterance; these conditions are presented in detail in the Appendix. We continue to assume that the subject-matter of dialogues can be represented in a propositional language by lower-case Roman letters, and, as before, we denote participating agents by $P_1, P_2$, etc. Since the work of Charles Hamblin [27], it has been standard to define a public store, called a *commitment store* for each participant in a dialogue game. We denote the store of agent $P_i$ by $CS(P_i)$. This store contains the propositions to which the participant is publicly committed, and the rules of the dialogue game may also define the circumstances under which propositions may be inserted or deleted from the commitment stores. The store for an agent contains the various propositions which that agent has publicly asserted or preferences he or she has declared; entries in the store are thus of two forms: (a) 2-tuples of the form $(type, t)$, where $t$ is a valid proposition instance of type *type*, with $type \in \{question, goal, constraint, perspective, fact, action, evaluation\}$; and (b) 3-tuples of the form $(prefer, a, b)$, where $a$ and $b$ are proposition actions. Each store can be viewed by all participants.

Before proceeding, it is worth noting that different authors have used different notions of commitment. For Hamblin [27, p. 257], commitment is purely dialectical, expressing only a willingness by the participant who has made the commitment to defend the proposition if it is attacked; in particular, commitments need not correspond to the participant's real beliefs. For Walton and Krabbe [65, Chapter 1], however, commitments are obligations to (execute, incur or maintain) a course of action. These actions may be utterances in a dialogue, as when a speaker is forced to defend a proposition he has asserted against attack from others; for these authors, propositional commitment is a special case of action commitments [65, p. 23]. For Munindar Singh, writing of agent communications languages [63], commitments may refer to either beliefs or actions, and may be private or expressed publicly, in which case they are termed *social commitments*. For our purposes, it is sufficient for commitments to be defended inside the dialogue, if and when attacked by other participants. We do not require that commitments correspond to the participants' real beliefs, nor that they indicate an intention to undertake some actions outside the world of the dialogue.

With this introduction, we are able to articulate the permissible locutions in the dialogue game:

**open_dialogue**($P_i, q?$)**:** Participant $P_i$ proposes the opening of a deliberation dialogue to consider the governing question $q$? A dialogue can only commence with this move.

[6] Dialogue games have also been used in computational linguistics to model natural language conversations, e.g., [37], although this work appears unaware of their far longer use in philosophy.

**enter_dialogue**($P_j, q?$)**:** Participant $P_j$ indicates a willingness to join a deliberation dialogue to consider the governing question $q?$ All intending participants other than the mover of **open_dialogue(.)** must announce their participation with this move. Note that neither the **open_dialogue(.)** nor the **enter_dialogue(.)** move implies that the speaker accepts that $q?$ is the most appropriate governing question, only that he or she is willing to enter into a discussion about it at this time.

**propose**($P_i$, *type*, $t$)**:** Participant $P_i$ proposes proposition $t$ as a valid instance of type *type*, where *type* $\in$\{ *question, goal, constraint, perspective, fact, action, evaluation*\}.

**assert**($P_i$, *type*, $t$)**:** Participant $P_i$ asserts proposition $t$ as a valid instance of type *type*, where *type* $\in$ \{ *question, goal, constraint, perspective, fact, action, evaluation*\}. This is a stronger locution than **propose(.)**, and results in the tuple ($type, t$) being inserted into $CS(P_i)$, the Commitment Store of $P_i$.

**prefer**($P_i, a, b$)**:** Participant $P_i$ indicates a preference for action-option $a$ over action-option $b$. This locution can only be uttered following utterance (possibly by other participants) of **assert**($P_j, evaluation, e$) locutions of at least two evaluations $e$, one of which has $a$ as its first argument, and one $b$. This combination rule ensures that preferences expressed in the dialogue are grounded in an evaluation of each action-option according to some proposed goal, constraint or perspective, and thus contestable. This locution inserts ($prefer, a, b$) into $CS(P_i)$, the Commitment Store of $P_i$.

**ask_justify**($P_j, P_i$, *type*, $t$)**:** Participant $P_j$ asks participant $P_i$ to provide a justification of proposition $t$ of type *type*, where $t \in CS(P_i)$.

**move**($P_i, action, a$)**:** Participant $P_i$ proposes that each participant pronounce on whether they assert proposition $a$ as the action to be decided upon by the group. This locution inserts $(action, a)$ into $CS(P_i)$.

**retract**($P_j$, *locution*)**:** Participant $P_j$ expresses a retraction of a previous locution, *locution*, where *locution* is one of three possible utterances: **assert**($P_j$, *type*, $t$) or **move**($P_i, action, a$) or **prefer**($P_i, a, b$) locution. The retraction locution deletes the entry from $CS(P_i)$ which had been inserted by *locution*.

**withdraw_dialogue**($P_i, q?$)**:** Participant $P_i$ announces her withdrawal from the deliberation dialogue to consider the governing question $q?$.

The locution **ask_justify**($P_j, P_i$, *type*, $t$) is a request by participant $P_j$ of participant $P_i$, seeking justification from $P_i$ for the assertion that proposition $t$ is a valid instance of type *type*. Following this, $P_i$ must either retract the proposition $t$ or shift into an embedded persuasion dialogue in which $P_i$ seeks to persuade $P_j$ that proposition $t$ is such a valid instance. One could model such a persuasion dialogue with a formal dialogue-game framework consistent with the deliberation framework presented here, drawing, for example, on the dialogue game models of persuasion proposed by Walton and Krabbe [65] or by Prakken [54].

The **move(.)** locution requests that participants who agree with a particular action being decided upon by the group should utter an **assert(.)** locution with respect to this action. Participants who do not agree that the particular action should be the decision of the group, or who wish to abstain from pronouncing on the issue, are free not to utter anything in response to the **move(.)** locution. Because in this model we have assumed unanimity of decision-making, the **Recommend** stage is only concluded successfully, and hence the dialogue only proceeds to the **Confirm** stage, in the case when all participants respond to the **move(.)** locution with the appropriate **assert(.)** locution.

We next show that our dialogue game framework implements the eight-stage model for deliberation dialogues proposed in the previous Section.

***Proposition 1:*** *Each of the eight stages of the formal model of deliberation dialogues presented in Section 3 of this paper can be executed by judicious choice of these locutions.*

***Proof.*** We consider each stage in turn:

- A dialogue opens with the locution **open_dialogue**($P_i, q?$) and at least one utterance of **enter_dialogue**($P_j, q?$), for $P_j$ and $P_i$ distinct participants.
- The **Inform** stage consists of utterances of **propose(.), assert(.), retract(.)** and **ask_justify(.)** for some or all of the types *goal, constraint, perspective*, and *fact*.
- The **Propose** stage consists of one or more utterances of **propose**($P_i, action, t$).
- The **Consider** stage consists of utterances of locutions **assert**($P_i, evaluation, e$), **prefer**($P_j, a, b$) and **ask_justify(.)**.
- In the **Revise** stage, a revision $a2$ to an action $a1$ proposed earlier may be proposed by means of the locution **propose**($P_i, action, a2$).
- The **Recommend** stage consists of an execution of **move**($P_i, action, a$), possibly followed by utterances of **assert**($P_j, action, a$), for $P_j$ and $P_i$ distinct participants.
- The **Confirm** stage only occurs following a **Recommend** stage where all participants have indicated acceptance of the recommended action-option. It then consists of the utterance of **assert**($P_j, action, a$) by every participant $P_j$, including the speaker of **move**($P_i, action, a$).
- The **Close** stage occurs whenever a participant utters **withdraw_dialogue**($P_i, q?$). A dialogue closes only when there remain two participants who have not uttered this locution, and one of them does so. □

# 5 Assessment of the Protocol

How may we assess this model of a deliberation dialogue and the associated dialogue game protocol? In other words, is this a good protocol or not? There are several ways to approach this question, and in the next three subsections we consider three of these. Firstly, we compare our protocol with actual human deliberation dialogues; secondly, we consider the protocol from the perspective of the deliberation processes it implements; and thirdly, we consider the outcomes, if any, which deliberation dialogues conducted under the protocol achieve.

## 5.1 Human Dialogues

Although we intend our framework to support only interactions between computational entities, its motivation and structure derive from consideration of human deliberation dialogues. Therefore, one approach to the assessment of the framework would be to ask whether it provides a good model of actual human deliberation dialogues. However, in doing so, it is important to realize that our framework is an idealization of human dialogues in at least two respects. Firstly, the framework presupposes cognitive abilities on the part of the participants which probably exceed those of most human deliberators, for instance, maintaining conformity with the pre-conditions of locution utterance; adhering to the rules regarding the order of dialogue stages; and keeping track of the contents of commitment stores of all participants as the discussion proceeds. Secondly, actual human dialogues undoubtedly contain more irrelevancies, rigidities, interruptions, and transitions to other types of dialogues which are not functionally embedded, than does our framework. Given this reality, there are two features of actual human dialogues which could lead us to revise our framework: the presence of constructive moves in human dialogues which our framework is unable to accommodate, or the absence of constructive components of a type of move presently included in our framework.

On the latter, most readers will have experienced human deliberation dialogues in which instances of the various locutions we have proposed have been used. For instance, if a group of friends decide to have dinner together and jointly seek to agree on a restaurant, often one or more participants will make proposals on which restaurant to select. Some may even propose that criteria for selection be established first, for example, that the restaurant be within walking distance, or provide food of a certain cuisine, or be within a certain price range, etc. Similarly, once suggested, such proposals may be subject to requests for justification, statements of preference, or suggestions that a particular option be selected. In the case where there are many dinner participants having conflicting preferences, there may even be a vote taken to make the final restaurant selection. Although everyday human deliberation dialogues are typically not as formal or as structured as is our framework, we believe they typically incorporate some or all of the ideal stages and constructive locutions we have identified.

What of more important human deliberation dialogues, such as those to decide great matters of state or of public policy? While perhaps a majority of such decisions involve deliberation dialogues, we have found few examples giving full accounts or transcripts of the dialogues themselves. Typical studies of Governmental decision-making, such as [28], an account of the decision-making processes in seven public policy domains in post-Independence Zimbabwe, reconstruct the major options considered and the arguments for and against them, but not in a sufficiently detailed manner to reveal the structure of deliberation dialogues used to reach decisions. However, we have found two examples of human deliberation dialogues in public policy domain from which we may infer the structure of dialogues.

The first example concerns the discussions within the leadership of the Chinese Communist Party (CCP) at the time of the pro-democracy student demonstrations in the Northern spring and summer of 1989. Here the deliberation dialogues concerned what to do, if anything, about the demonstrations. In the end, the CCP leadership decided to impose martial law, and order soldiers from the Chinese People's Liberation Army to remove the demonstrators forcibly, an action which led to killings of demonstrators in Tiananmen Square in Beijing and elsewhere in China in June 1989. Recently, documents purporting to be the minutes of some of the relevant CCP meetings have been smuggled out of China and published [48]. Although their authenticity has not been verified, three eminent Western scholars of Chinese politics found in them nothing to indicate that they were not genuine [48].

However, as instances of deliberation dialogues, these records are not very informative. The relative political power of the participants appears to have greatly influenced what they say to one another, and there is little substantive discussion of the consequences of alternative courses of action, or their relative advantages and disadvantages. For such a major decision, there is (at least in these documents) remarkably little debate or substantive analysis. For example, once Deng Xiaopeng, the most powerful participant in the discussions, had decided upon martial law, all but two of the other participants, the brave Zhao Ziyang and Hu Qili, also supported it. The nature of this support appears to have mostly been political point-scoring and scape-goating, primarily directed against Zhao; in reading these transcripts, one has the impression that the speakers expressing such views were articulating positions they already knew Deng to support. Moreover, these dialogues do not provide an example of retroflexive argumentation, because the one proposal considered, imposing martial law, is not modified in the light of the few objections raised to it. Because our framework is intended to be a general one, we have not explicitly modeled any power relationships between the participants. This would be possible, and has, indeed, been done in other work on modeling coordination and negotiation in AI, e.g., in [50].

The participants in a second example of human deliberation dialogues in public policy domains were more equal than the CCP leadership appears to have been during the Tiananmen crisis. This example involved the discussions in the British War Cabinet in May 1940, when, following the appointment of Winston Churchill as Prime Minister, the members of the War Cabinet discussed various proposals regarding the conduct of the war with Germany [40]. One of the proposals considered was to seek to negotiate a peace agreement with Germany, and thus end the conflict quickly. Some of the participants, notably Churchill, had previously been strongly opposed to this option, but (according to the reconstruction by John Lukacs in [40]), Churchill felt his political support at this time within the Parliamentary Conservative Party and within the Cabinet was not strong. He therefore (according to Lukacs) pretended to entertain the proposal seriously, so as to strengthen his support with key Ministers and backbenchers, and so as not to provide his enemies with political ammunition against him at this time.

Our framework does not differentiate between sincere and insincere expressions of beliefs or preferences in a dialogue. Perhaps no computational framework can ever deal with this issue adequately, because any semantic requirement could always be simulated insincerely by a sufficiently-clever agent [68]. In other words, it is hard to see how a framework could represent dialogues in which statements are made to create the impression that the speaker supports a position he really does not, or to provoke other participants to reveal their true positions prematurely, so that these may be countered or rebutted, or to jockey for influence with third parties, both present and absent. All of these features are to be found in human deliberation dialogues, particularly when important public policy decisions are to be made. Even Singh's notion of a social semantics [62] — a commitment store involving a public express of beliefs and intentions by each participant at the outset of a dialogue — will only enable statements in the subsequent dialogue to be verified for *consistency* with the expressed beliefs and intentions, not the degree of sincerity with which these beliefs and intentions are held.

## 5.2 Deliberation Process

A second approach to assessment of our framework is to measure it against normative principles for deliberation. We know of only three such sets of principles. The first set are criteria for public decision processes in environmental matters, identified by Thomas Webler, Seth Tuler and Rob Krueger [66]. These principles were derived from a statistical multi-variate factor analysis of the interview responses of participants in recent environmental public consultation exercises in the USA. The five resulting principles are pitched at a very abstract level; for example, the second principle is that the process should promote a search for common values. Although certainly useful for designers of public policy decision processes, the abstraction of these principles makes them unsuitable for assessment of our framework.

#### Alexy's Rules for Discourse Ethics

The second set of normative principles are Robert Alexy's rules for discourse ethics [1]. These were designed as principles for rational discussion over ethical norms between free and consenting participants, building on Jürgen Habermas' philosophy of discourse ethics [24]. Habermas sought to understand how rational, free people could engage in reasoned discussion and reach agreement over moral and ethical questions, and Alexy articulated a set of rules for such discussions [1]. We list the rules here, using Alexy's categorization, naming and numbering (apart from an initial **A** for each rule); for simplicity we use only the masculine gender.

**A1. Basic Rules**

- **A1.1** No speaker may contradict himself.
- **A1.2** Each speaker may only assert what he himself believes.
- **A1.3** Each speaker who applies a predicate $F$ to an object $a$, must also be prepared to apply $F$ to any other object which is similar to $a$ in all respects.
- **A1.4** Different speakers may not use the same expression with different meanings.

**A2. Rules of Reason**

- **A2 (General Rule of Justification):** Every speaker must justify what he asserts upon request, unless he can provide grounds which justify avoiding giving a justification.
- **A2.1** Anyone who can speak may take part in discourse.
- **A2.2 (a)** Anyone may render any assertion problematic.
- **A2.2 (b)** Anyone may introduce any assertion into the discourse.
- **A2.2 (c)** Anyone may express his opinions, wishes and needs.
- **A2.3** No speaker may be prevented by constraint within or outside the discourse from making use of his rights established in 2.1 and 2.2.

**A3. Rules of the Burden of Argumentation**

- **A3.1** Whoever wishes to treat a person $A$ differently from a person $B$ is obliged to justify this.

**A3.2** Whoever attacks a statement or norm that is not the object of discussion must provide a reason for doing so.

**A3.3** Whoever has put forward an argument is only committed to further arguments in the case of a counterargument.

**A3.4** Whoever introduces an assertion or a statements concerning his opinions, wishes, or needs into the discourse, which as argument is not related to a previous statement, has to justify upon request why he has introduced this assertion or this statement.

**A4. Forms of Argument**

*Under this heading, Alexy proposes six normative models for the structural form of arguments concerning ethical values and norms, forms which depend upon the reasons advanced for such values and the perceived consequences of adopting them. We do not present or discuss these here, as they are specific to arguments over ethical values.*

**A5. Rules of Justification**

**A5.1.1** Everyone must be able to accept the consequences of the rule — presupposed in his normative statements — regarding the satisfaction of the interests of each individual person even for the hypothetical case in which he finds himself in the situation of this person.

**A5.1.2** The consequences of every rule for the satisfaction of the interests of each and every individual must be capable of being accepted by all.

**A5.1.3** Every rule must be openly and universally teachable.

**A5.2.1** The moral rules that form the basis of the moral conceptions of the speakers must be able to withstand scrutiny in a critical, historical genesis. A moral rule does not withstand such a scrutiny

- **(a)** if it was indeed originally justifiable rationally but in the meantime has lost its justification, or
- **(b)** if it was already originally not justifiable rationally and if no sufficient new reasons for it can be found.

**A5.2.2** The moral rules that form the basis of the moral conceptions of the speakers must be able to withstand the scrutiny of their individual history of emergence. A moral rule does not withstand such a scrutiny if it is only accepted on the basis of conditions of socialization that are not justifiable.

**A5.3** The factually given limits of realizability are to be observed.

**A6. Rules of Transition**

**A6.1** It is possible at all times for any speaker to switch to a theoretical (empirical) discourse.

**A6.2** It is possible at all times for any speaker to move to a linguistic-analytical discourse.

**A6.3** It is possible at all times for any speaker to move to a discourse on discourse theory.

Habermas' theory of discourse ethics has subsequently been applied to legal and political philosophy [25, 26], and to a philosophical assessment of electronic democracy [14]. Despite these examples of wider application, however, some of Alexy's rules appear very specific to ethical discussions and not applicable to generic deliberation dialogues. For instance, *Rule 4: Forms of Argument*, consists of six normative models for the structural form of arguments concerning ethical values and norms. Similarly specific to discourse ethics are Rules A2.1, A3.1, A5 and A6.3. The other rules have applicability to wider deliberation dialogues, and accordingly, we can assess our framework against them. The next proposition presents the result of this assessment.

***Proposition 2:*** *The deliberation dialogue framework presented in Sections 3 and 4 satisfies Alexy's rules for discourse ethics to the following extent:*

*Rules fully satisfied: Rules A2, A2.2, A3.2, A3.3, A3.4, A6.1 and A6.2.*

*Rules partly satisfied: Rule A2.3.*

*Rules not satisfied: Rules A1.1, A1.2, A1.3 and A1.4.*

*Rules specific to ethical discussions: Rules A2.1, A3.1, A4, A5 and A6.3.*

***Proof.*** We consider each rule in turn. Rule A1.1 is not satisfied: participants using our framework may contradict themselves, as seen by examining the pre-conditions for the locutions given in the Appendix. Rules A1.2, A1.3, and A1.4 are not satisfied: our framework is defined purely in terms of observable linguistic behaviour, and has no requirements that participants are sincere, or express consistent statements.

Rule A2 (General Rule of Justification) is satisfied, via the **ask_justify(.)** locution. The three parts of Rule A2.2 are satisfied, by means of the **ask_justify(.)**, **assert(.)** and **prefer(.)** locutions, respectively. Rule A2.3 is satisfied within the dialogue by means of the pre-conditions of the locutions given in the Appendix. Our framework makes no assumptions concerning any relationship between the parties external to the dialogue, and so the framework cannot be assessed with regard to constraints on speakers imposed outside the dialogue.

Rule A3.1 is specific to ethical discussions. Rule A3.2 is satisfied: participants may only attack a statement via the **ask_justify(.)** locution, which has as a pre-condition the requirement that a prior **assert(.)** locution has been uttered concerning the same statement. Rules A3.3 and A3.4 are both satisfied, by the definition of the **ask_justify(.)** locution. Rules A4 and A5 are specific to ethical discussions. Rule A6.1 and A6.2 are satisfied: Although the types of propositions, the locutions and the combination rules in our framework are specific to deliberation dialogues, the framework permits shifts to functionally-embedded dialogues of different types, such as inquiry dialogues or persuasion dialogues. These may concern theoretical, empirical or linguistic-analytical matters. Rule A6.3 is specific to ethical discussions. □

In assessing our framework against Alexy's normative rules, we note that three of his rules which are not satisfied, A1.2, A1.3 and A1.4, concern the relationship between what is uttered in the dialogue and what the speaker truly believes. As noted in the previous sub-section, our framework does not distinguish between sincere and insincere utterances, and makes no requirements that speakers express only their true beliefs or preferences.

### Hitchcock's Principles for Rational Mutual Inquiry

A third set of normative principles are the Principles of Rational Mutual Inquiry developed by one of us a decade ago [29]. These were intended for human dialogues whose primary purpose was defined as being *"to secure rational agreement by the participants on the answer to a specified question. A subsidiary purpose, if they do not come to agree on an answer, is to secure agreement on why they have not succeeded in answering their question."* [29, p. 237]. These human dialogues are called *mutual inquiries*; in terms of the typology of Walton and Krabbe [65], this definition was formulated with inquiry dialogues primarily in mind, but also covers deliberation dialogues. It is therefore appropriate to consider them as principles against which our deliberation dialogue protocol may be measured. We begin by summarizing these Principles, numbered H1 through H18; the linguistic labels are those of the original.

**H1 Externalization:** The rules should be formulated in terms of verifiable linguistic behaviour.

**H2 Dialectification:** The content and methods of dialogue should be subject to the agreement of participants, without any prior imposition.

**H3 Mutuality:** No statement becomes a commitment of a participant unless he or she specifically accepts it.

**H4 Turn-taking:** At most one person speaks at a time.

**H5 Orderliness:** One issue is raised at a time and is dealt with before proceeding to others.

**H6 Staging:** An inquiry dialogue should proceed by a series of stages, from initial clarification of the question at issue and of the methods of resolving it, through data gathering and intepretation, to formation of arguments.

**H7 Logical Pluralism:** Arguments should permit both deductive and non-deductive forms of inference.

**H8 Rule-consistency:** There should be no situation where the rules prohibit all acts, including the null act.

**H9 Semantic Openness:** The rules should not force any participant to accept any statement, even when these follow by deduction from previous statements.

**H10 Realism:** The rules must make agreement between participants a realistic possibility.

**H11 Retraceability:** Participants must be free at all times to supplement, change or withdraw previous tentative commitments.[7]

**H12 Role reversal:** The rules should permit the responsibility for initiating suggestions to shift between participants.

**H13 Experiential Appeal:** The rules should permit direct mutual appeal to experience.

[7] Political theorists use the term *self-transformation* to refer to the changes in beliefs or intentions which participants may experience in the course of a discussion [19].

**H14 Openness:** There should be no restrictions on the content of contributions.

**H15 Tentativeness:** Participants should be free to make tentative suggestions as well as assertions.

**H16 Tracking:** The rules should make it possible to determine at any time the cumulative commitments, rights and obligations of each participant.

**H17 Termination:** There should be rules for the orderly termination of the dialogue. Hitchcock proposes that an inquiry terminate as soon as (a) a participant declares an intention to abandon it, (b) in two successive turns neither participant has a suggestion for consideration, or (c) there is agreement on the conclusion of the discussion.

**H18 Allocation of Burden of Proof:** The burden of proof remains with the participant who makes a suggestion, even after contestation by another Participant.

As the next Proposition reveals, our protocol satisfies most of Hitchcock's Principles.

***Proposition 3:*** *The deliberation dialogue framework presented in Sections 3 and 4 satisfies all but four of Hitchcock's eighteen Principles of Rational Mutual Inquiry. Principles H2 (Dialectification), H5 (Orderliness), H11 (Retraceability) and H14 (Openness) are only partly satisfied.*

***Proof.*** We consider each Principle in turn. Principle H1 (Externalization) is satisfied, as can be seen by an examination of the pre- and post-conditions of the locutions listed in the Appendix and the constraints on the order of dialogue stages given in Section 3.[8] Principle H2 (Dialectification) is only partly satisfied, since we do not permit participants to change the protocol framework itself. Principle H3 (Mutuality) is satisfied, as shown by the commitment store conditions for the locution **assert(.)**. Principle H4 (Turn-taking) will be satisfied in any computational application on a sequential processor. Principle H5 (Orderliness) is satisfied to the extent that each dialogue under the protocol concerns one governing question. However, there is nothing to stop issues related to this question being considered simultaneously in a manner contrary to this Principle. The next principle, H6 (Staging), is satisfied by the phased framework presented in Section 3. Principle H7 (Logical Pluralism) is satisfied, since there are no restrictions placed on the content of the justifications participants may advance for their statements. Principle H8 (Rule-consistency) is satisfied, as is shown by an examination of the post-conditions of each locution given in the Appendix.

Principle H9 (Semantic Openness) is satisfied, since no rules force a participant to accept any statement. Principle H10 (Realism) is satisfied, since the protocol readily permits participants to express their agreement to statements uttered in dialogues under it. Principle H11 (Retraceability) is satisfied up to the execution of the **Confirm** Stage, by means of the **retract(.)** locution. Utterances of acceptances in this stage can not be subsequently retracted. Principle H12 (Role reversal) is satisfied, since any participant may initiate suggestions in the dialogue. Principle H13 (Experiential Appeal) is satisfied, since participants may support their utterances in any way they wish. Principle H14 (Openness) is partly satisfied, since the contents of utterances are typed according to the types of proposition given in Section 3. However, apart from this typing, there

[8] In contrast, the definition of the syntax of the Agent Communications Language (ACL) of the Foundation for Intelligent Physical Agents (FIPA), an emerging standard for agent communications, requires agents to sincerely believe statements they make in dialogues [18], thus violating this principle.

are no restrictions on the content of contributions. Principle H15 (Tentativeness) is satisfied because the **propose(.)** locution permits participants to make tentative suggestions. Principle H16 (Tracking) is satisfied by means of the commitment stores established for each participant. Principle H17 (Termination) is satisfied by the rules governing the **Confirm** stage and the rules governing withdrawal from the dialogue. The protocol rules allow participants to withdraw at any time, and without giving reasons. Principle H18 (Allocation of Burden of Proof) is satisfied by the definition of the **ask_justify(.)** locution, which permits a participant to contest an earlier assertion by another participant, and requires that other participant to provide a justification for the earlier assertion. □

It is worth noting that there is some inconsistency within Hitchcock's collection of Principles. Principles H5 (Orderliness), H6 (Staging) and H17 (Termination) may conflict with Principle H2 (Dialectification), since the latter gives the participants complete freedom, including the freedom to change the rules of the protocol. Essentially, this inconsistency arises because of the need to meet two desirable, but conflicting, objectives in the design of a protocol: freedom for the participants and orderliness of the resulting dialogues. By the very act of defining a protocol for dialogues, we are constraining the freedom of the participants in some way, and are imposing some structure on the interactions between them. Because we seek to define a framework within which deliberation dialogues between computational entities can occur, our task, as designers, is to strike an appropriate balance between these conflicting objectives.[9] Our framework, while not maximally dialectical, is dialectical to a considerable extent, for instance, in leaving the participants free to agree on what factors to accept as relevant to the governing question, or even to change the governing question. The framework could be made more dialectical by providing for the opportunity to convene a *"loya jurga"* or "constituent assembly" to change the framework rules; such an assembly could, for example, change the requirement of unanimity of decision-making (in the definition of the **Confirm** stage given in Section 3) to a requirement that, say, only a two-thirds majority of acceptances is necessary for a decision to be adopted by the group.

## 5.3 Deliberation Outcomes

The previous section considered our deliberation dialogue protocol from the perspective of the *processes* it implemented. We could also assess a protocol in terms of the outcomes achieved, if any, of dialogues conducted under the protocol. For example, a protocol to support an inquiry dialogue could be assessed on whether or not dialogues conducted according to the protocol succeed in finding the answer to the question motivating the dialogue. In other words, is the outcome of an inquiry dialogue the true answer to the governing question? Since some questions may be undecidable or may require considerable time or significant resources for answers to be found, a more refined measure of the protocol may be whether it leads, on average, to the truth, or whether it would do so, given infinite time and unlimited processing resources. Two of us adopted this approach to study the formal properties of a dialogue game protocol we proposed for scientific inquiry dialogues, showing that, under some conditions, the probability that a dialogue under the protocol did not converge on the truth could be bounded away from 1 [43].

[9] Krabbe notes a similar conflict of design objectives in a discussion of retraction rules in dialogue games [35]. As an example, the dialogue game protocols of Amgoud and Parsons [4, 5, 6], particularly those in [52], are at the orderliness end of the freedom–orderliness spectrum.

In contrast to inquiry dialogues, deliberation dialogues have as their stated objective agreement on some course of action, rather than a search for truth. In this, they are similar to negotiation dialogues, where the stated objective is agreement on an action of a particular type, namely a division of a scarce resource. We mention negotiation dialogues here because this objective is shared by the auction and negotiation protocols studied in the branch of economics known as mechanism theory, and considerable attention has been devoted to assessment of the outcomes of these mechanisms, e.g., [53, 58, 60]. Among the usual criteria proposed are:

**Maximum social welfare:** Intuitively, a protocol maximizes social welfare if it ensures that any outcome maximizes the sum of the utilities of negotiation participants. If the utility of an outcome for an agent was simply defined in terms of the amount of money that the agent received in the outcome, then a protocol that maximized social welfare would maximize the *total* amount of money "paid out."

**Pareto-efficient:** An outcome is *Pareto-optimal* if any other outcome leaves at least one participant worse off, as measured by the utility of the outcome. A mechanism which achieves Pareto-optimal outcomes is said to be *Pareto-efficient*.

Many auction and economic negotiation mechanisms have been studied and shown to have these properties. We know of only one study of negotiation dialogues which considers properties such as these, recent work of two of us [45]. In this we showed, under assumptions concerning the absence of time constraints and of coercion on participants, that the outcomes of negotiation dialogues between self-interested and non-malicious participants conducted according to protocols with certain properties are Pareto-optimal [45, Proposition 1].

Adopting a similar approach to assess protocols for deliberation dialogues would mean considering whether dialogues conducted according to the protocol succeed in agreeing on a course of action, and considering the quality of this agreed course. But how to judge the quality of a course of action? We are not given antecedently a set of evaluative criteria (goals, constraints, considerations, etc) in terms of which one could theoretically determine, given all the relevant factual circumstances, what is the "best" answer to the governing question. In fact, even the governing question itself is subject to revision under our protocol if the participants agree to revise it. Furthermore, the protocol does not require all participants to agree at any point in the discussion on the evaluative criteria to be used, and so conflicting evaluative criteria may be supported throughout the dialogue.

For these reasons, it seems that the best one might do is to establish *conditional* results about outcomes of dialogues using the protocol. For example, such a result might be that, given agreement by the participants to a set of evaluative criteria and a set of factual propositions, then, if the participants use the protocol, they will reach agreement on an answer to the governing question which is optimal, provided those agreed evaluative criteria and factual propositions are valid and exhaustive of matters relevant to the governing question, and provided the participants undertake the dialogue free of time and processing constraints, and free of coercion or duress. However, to prove this formally we believe would require a "semantic" theory of actions akin to the standard account of propositional truth initiated by Wittgenstein and Tarski. We know of no such theory, and developing it would be a much larger undertaking than could be accommodated in this paper. We therefore leave the assessment of our protocol on the basis of the outcomes of dialogues conducted under it to another time and place.

# 6 Example

We now consider a simplified example of a dialogue undertaken according to our deliberation dialogue protocol. In this example, the deliberation concerns what action to take regarding potential health hazards from the use of cellular phones. The dialogue utterances are numbered sequentially from **U1**, and each is annotated.

**U1: open_dialogue**($P_1$, *Do what about mobile phone health risk?*)

This move is the first move in the **Open** stage of the dialogue.

**U2: enter_dialogue**($P_2$, *Do what about mobile phone health risk?*)

With the entry of a second participant, the dialogue may be said to commence.

**U3: enter_dialogue**($P_3$, *Do what about mobile phone health risk?*)

A third participant also enters the dialogue.

**U4: propose**($P_2$, *perspective, degree of risk*)

Participant $P_2$ proposes that *degree of risk* should be a perspective from which to consider the question. With this move, the dialogue enters an **Inform** stage.

**U5: propose**($P_3$, *perspective, economic cost*)

Participant $P_3$ proposes that *economic cost* should be a perspective from which to consider the question.

**U6: propose**($P_1$, *action, prohibit sale of phones*)

Participant $P_1$ proposes *prohibition of sale of phones* as an action-option. With this move, the dialogue enters a **Propose** stage.

**U7: propose**($P_3$, *action, do nothing*)

Participant $P_3$ proposes *doing nothing* as an action-option.

**U8: assert**($P_1$, *evaluation, prohibit sale from a degree of risk perspective is lowest risk*)

Participant $P_1$ asserts that from the perspective of the degree of risk, prohibiting the sale of phones is the lowest risk action-option possible. With this move, the dialogue enters a **Consider** stage.

**U9: assert**($P_3$, *evaluation, prohibit sale from an economic cost perspective is high-cost*)

Participant $P_3$ asserts that from the perspective of economic cost, prohibiting sale is a high-cost option.

**U10: propose**($P_1$, *action, limit usage*)

Participant $P_1$ proposes limiting usage as an action-option, thus responding retroflexively to the previous two **assert(**$P_i, evaluation, e$**)** locutions. With this move, the dialogue enters a **Revise** stage.

**U11: propose**($P_2$, *perspective, feasibility*)

Participant $P_2$ proposes feasibility as a perspective from which to consider the question. With this move, the dialogue enters another **Inform** stage.

**U12: assert**($P_2$, *evaluation, limit usage from a feasibility perspective is impractical*)

Participant $P_2$ asserts that from the perspective of feasibility, limiting usage is not practical. With this move, the dialogue enters another **Consider** stage.

**U13: prefer**($P_1$, *prohibit sale, limit usage*)

Participant $P_1$ expresses a preference for the option of prohibiting the sale of phones over limiting their usage. The utterance is valid at this point, since each action-option has appeared as the first argument in a proposition $e$ of type *evaluation* in an **assert**($P_i, evaluation, e$) locution.

This example, although very simple, illustrates the usage of selected locutions, and demonstrates the way in which a dialogue may move between stages as it proceeds. Such cycling between stages is commonplace in human deliberations, where comments, arguments or preferences uttered by one participant are likely to provoke others to think of new goals, constraints, facts, perspectives or action-options.

# 7 Discussion

This paper has presented a dialogue-game protocol for deliberation dialogues between computational entities, with the syntax for the protocol being fully specified. In the terminology of computer programming theory, e.g., [13, 64], the protocol has been given an axiomatic semantics. The protocol was based on a model for deliberative reasoning taken from argumentation theory, namely Harald Wohlrapp's theory of Retroflexive Argumentation [67]. Moreover, we demonstrated that the protocol conforms to the majority of a set of normative principles proposed for rational mutual inquiries between humans. Further work is needed to assess the quality of outcomes achieved, if any, by dialogues conducted according to the protocol. In enabling participants to contribute to a joint discussion which may proceed iteratively and to view each other's commitment stores, our model has some similarities with *"blackboard"* architectures for intelligent systems in computer science [49].

### Related Work

Considerable research effort in AI over the last thirty years has concerned the task of designing robots so that, when given a specific goal, such as moving into the next room, they may determine a plan for achievement of this goal. Because this research, known as *AI Planning*, concerns the determination of an action or course of actions, it would seem amenable to the application of deliberation dialogues. However, the only research program known to us which combines AI Planning with models of dialogues is the *TRAINS* project [2], which constructed an intelligent computer assistant for a human rail-freight scheduler. For this project, actual human-human conversations in the specific domain were first recorded and analyzed as a basis for the design of machine-human interactions. Although the two participants in the TRAINS system, machine and human, discuss a course of action, and thus ostensibly engage in a deliberation dialogue, the design of the system assumes that the machine and the human-user each begin the dialogue with a privately-developed proposal for action, which they then present to one another. Each tries to persuade the other to adopt its proposal. Thus, in the terminology of Walton and Krabbe [65], their conversation is a persuasion dialogue, albeit two-way, rather than a true deliberation. In addition, the TRAINS system design assumes that the human user's goal is paramount, and that the machine participates in the dialogue to assist the human to find an effective plan for achievement of this goal. Thus, the model of dialogue assumes a specific relationship of inequality between the two participants. By contrast, the model of deliberation dialogue we have presented here is not limited in this way.

Other work in AI has also come close to developing a formal model of deliberation dialogues without yet doing so. The dialogue-game protocols proposed for developing collective intention by Frank Dignum and his colleagues [11, 12], assume, like the research in AI Planning, that the overall goal of the participating agents is predetermined. Moreover, these authors assume that one agent, an *Initiator*, undertakes a persuasion dialogue to convince the others to adopt some joint intention it has adopted. Although the task is a deliberative one, the dialogue model proposed for it is not that of a deliberation. The same comment is true of other recent research in multi-agents systems. The agent interactions in the work of Simon Parsons and his colleagues [51], for example, are deliberations mixed with persuasions, negotiations and information-seeking dialogues, as noted in [4]. However, they are modeled as persuasions, with one agent uttering an argument which the recipients try to counter. Similarly, the *SharedPlans* framework of Barbara Grosz and Sarit Kraus [23], for collaborative planning between agents, assumes that agents begin their interaction with a partial plan; this framework does not fully specify the mechanisms by which this partial plan is transformed into a full plan. Luke Hunsberger and Massimo Zancanaro [31], seeking to remedy this, have articulated mechanisms to enable SharedPlan participants to vote over contested elements of a possible plan. However, these mechanisms do not permit the expression of arguments for and against proposals, and so we would not call them models of deliberation dialogues. In the language of argumentation theory, e.g., [32], the inability to express reasons for statements renders these mechanisms *non-rational*. Another approach to representing deliberation interactions between autonomous agents is the work of Pietro Panzarasa, Nicholas Jennings and Timothy Norman [50], who propose a modal logic formalism to represent the mental states of, and interactions between, the participating agents. This framework, however, assumes [50, Section 8] that at least one participant begins the discussion with a suggested proposal for action; the resulting agent interaction to decide a course of action, although termed a *negotiation* by the authors, is therefore modeled, like the *TRAINS* system, as a persuasion and not a deliberation. Moreover, this model of persuasion, which the authors call *social mental shaping*, is one based on the exercise of social relationships between agents, such as that pertaining between a manager and her subordinates in a company.[10] Although such a model has wide applicability, it is not as general as the one we have presented above, which assumes nothing about the relationships between the participants; nor could social mental shaping be called an entirely *rational* model for deliberation, because essentially the only reason an agent can provide to another to adopt a proposed course of action is, *"Because I said so!"*

Within the area of computational dialectics specifically, several computer systems have been designed to support human deliberation dialogues. The *Zeno* system, for ex-

[10]Such relationships are readily captured in preference-based argumentation systems, such as [5].

ample, of Thomas Gordon and his colleagues [21, 34], was designed to support community participation in urban planning decisions. The model of argumentation used in this work was the *IBIS* system of Rittel and Webber [57], which provides a framework for connecting topics, issues and attributes in a multi-attribute decision domain. Later systems inspired by Zeno, such as the *Hermes* system for computer-supported collaborative human decision-making [33] and the *Demos* system to support human debate over issues of public policy [39], also use the IBIS framework. This framework connects utterances in a dialogue on the basis of their meanings (with respect to some decision problem), but does not specify or constrain the dialectical obligations of the participants. If a statement uttered by a participant in the dialogue challenges a previous statement by another participant, the IBIS framework provides a mechanism to represent the relationship between the two, but the framework has no rules or mechanisms for requiring such a challenge to be made, or for defending the earlier statement against such a challenge, or for resolving multiple conflicting statements.[11] The IBIS framework has no rules of the form described in our dialogue game protocol which require or preclude particular types of responses when statements are uttered. The same comment is also true of various systems using spatial representations of statements and their relationships in computer-supported human dialogues, as in [10, 47]. The resulting systems are thus capable of supporting human dialogues which are more free-wheeling than the agent dialogues our model supports, but, because of the absence of rules specifying dialectical obligations, we do not believe that any of these systems incorporates a formal model of deliberation dialogues.

### Future research

We are exploring a number of extensions of this work. Firstly, we seek to model and automate more general classes of deliberation dialogue. For example, many human deliberations exhibit strong disagreement between the participants over the relevance and importance of different perspectives. Our dialogue-game model may be extended to allow for similar arguments between agents over these. Secondly, we plan to enable discussion over confirmation procedures, so that, for example, majority or plurality voting may be used instead of the unanimity now required in the **Confirm** stage. If a group of agents were to engage regularly in deliberation dialogues using the same decision-procedures, these procedural discussions would not need to be undertaken in each dialogue but could be assumed constant. Systems for agent interactions with such pre-determined rules of encounter have been called *Institutions* in the AI literature, e.g., [61]. Thirdly, our explicit typing of propositions (into facts, goals, constraints, etc) may facilitate the mathematical representation of dialogues under this model by means of the $\lambda$-calculus [9], and thus the possible development of a denotational semantics for the protocol using enriched category theory, as has been achieved for monolectical argumentation in [3].

### Acknowledgments

This work was partly funded by the EU IST Programme through the Sustainable Lifecycles in Information Ecosystems (SLIE) Project (IST-1999-10948), and a PhD studentship from the British Engineering and Physical Sciences Research Council (EPSRC). An earlier version of this paper was presented at the May 2001 Meeting of the

[11] In the three systems mentioned, this task is left to a human mediator, possibly assisted by computer summarization.

Ontario Society for the Study of Argumentation (OSSA), in Windsor, Canada, and we thank the audience on that occasion for their comments. We also thank Joris Hulstijn for his comments on an earlier version of this work.

# 8 Appendix: Axiomatic Semantics

In this appendix, we define the pre-conditions for the legal utterance of locutions, and the post-conditions which occur upon their utterances, for each of the locutions of the deliberation dialogue protocol presented in Section 4. Such a presentation in terms of pre- and post-conditions is commonly known in AI as a *STRIPS-like* notation, following [15]. Within the theory of computer programming languages it is also called an *axiomatic semantics* for the language [13, 64].

**L1:** The **open_dialogue(.)** locution:

> **Locution:** **open_dialogue**($P_i, q?$), where $q?$ is a question.
>
> **Preconditions:** No preconditions within the dialogue.
>
> **Meaning:** Participant $P_i$ proposes the opening of a deliberation dialogue to consider the governing question $q?$. A dialogue can only commence with this move.
>
> **Response:** No response required. Other intending participants may respond with the **enter_dialogue(.)** locution.
>
> **Commitment Store Update:** No effects.

**L2:** The **enter_dialogue(.)** locution:

> **Locution:** **enter_dialogue**($P_i, q?$), where $q?$ is a question.
>
> **Preconditions:** A participant $P_i$, where $P_i$ and $P_j$ are distinct, must previously have uttered the locution **open_dialogue**($P_i, q?$).
>
> **Meaning:** Intending participant $P_j$ indicates a willingness to join a deliberation dialogue to consider the governing question $q?$ All intending participants other than the speaker of **open_dialogue(.)** must announce their participation with this move.
>
> **Response:** No response required. This locution is a pre-condition for all locutions other than **open_dialogue(.)**, i.e., an intending speaker $P_k$ of any other locution must have previously uttered **enter_dialogue**($P_k, q?$)**.** As soon as one participant has uttered the **enter_dialogue**($P_j, q?$) locution, the dialogue is said to be *Open*.
>
> **Commitment Store Update:** No effects.

Since all the locutions listed below have a common precondition, namely that the speaker $P_j$ has previously uttered either the locution **open_dialogue**($P_j, q?$) or the locution **enter_dialogue**($P_j, q?$), we do not list this precondition under each locution; only those preconditions specific to the locution concerned are listed.

**L3:** The **propose(.)** locution:

> **Locution:** **propose**($P_i$, *type*, $t$), where $t$ is a proposition, and *type* is an element of the set {*question, action, goal, constraint, perspective, fact, evaluation*}.

**Preconditions:** No specific preconditions.

**Meaning:** Participant $P_i$ proposes proposition $t$ as a valid instance of type *type*.

**Response:** No response required.

**Commitment Store Update:** No effects.

**L4:** The **assert(.)** locution:

**Locution:** **assert(**$P_i$, *type*, $t$**)**, where $t$ is a proposition, and *type* is an element of the set $\{$*question, action, goal, constraint, perspective, fact, evaluation*$\}$.

**Preconditions:** No specific preconditions.

**Meaning:** Participant $P_i$ asserts proposition $t$ as a valid instance of type *type*.

**Response:** No response required.

**Commitment Store Update:** The 2-tuple $(type, t)$ is inserted into $CS(P_i)$, the Commitment Store of participant $P_i$.

**L5:** The **prefer(.)** locution:

**Locution:** **prefer(**$P_i, a, b$**)**, where $a$ and $b$ are propositions of type *actions*.

**Preconditions:** Some participants $P_j$ and $P_k$, possibly including $P_i$, must previously have uttered the locution **assert(**$P_j, evaluation, e$**)** and the locution **assert(**$P_k, evaluation, f$**)**, where $e$ and $f$ are propositions of type *evaluation* which refer respectively to action-options $a$ and $b$.

**Meaning:** Participant $P_i$ indicates a preference for action-option $a$ over action-option $b$.

**Response:** No response required.

**Commitment Store Update:** The 3-tuple $(prefer, a, b)$ is inserted into $CS(P_i)$, the Commitment Store of $P_i$.

**L6:** The **ask_justify(.)** locution:

**Locution:** **ask_justify(**$P_j, P_i$, *type*, $t$**)**, where *type* is an element of the set $\{$*question, action, goal, constraint, perspective, fact, evaluation*$\}$.

**Preconditions:** Participant $P_i$ has previously uttered the locution **assert(**$P_i$, ***type***, $t$**)**.

**Meaning:** Participant $P_j$ asks participant $P_i$ to provide a justification of proposition $t$ of type *type*, where $(type, t) \in CS(P_i)$.

**Response:** $P_i$ must either retract the proposition $t$ or seek to persuade $P_j$ in an embedded persuasion dialogue that proposition $t$ is a valid instance of type *type*.

**Commitment Store Update:** No effects.

**L7:** The **move(.)** locution:

**Locution:** **move(**$P_i, action, a$**)**, where $a$ is a proposition of type *action*.

**Preconditions:** Some participant $P_j$, possibly $P_i$, must previously have uttered either **propose(**$P_i, action, a$**)** or **assert(**$P_i, action, a$**)**.

**Meaning:** Participant $P_i$ proposes that each participant pronounce on whether they assert proposition $a$ as the action to be decided upon by the group.

**Response:** None required. Other participants $P_k$, distinct from $P_i$, who wish to support proposition $a$ as the action to be decided upon by the group can respond with locution **assert(**$P_k, action, a$**)**.

**Commitment Store Update:** The 2-tuple $(action, a)$ is inserted into $CS(P_i)$.

**L8:** The **retract(.)** locution:

**Locution:** **retract(**$P_i$**, *locution*)**, where *locution* is one of the locutions, **assert(.), move(.)** or **prefer(.)**.

**Preconditions:** Participant $P_i$ must have previously uttered and not subsequently retracted the locution *locution*.

**Meaning:** Participant $P_i$ expresses a retraction of a previous utterance *locution*, where *locution* is one of the following three locutions: **assert(**$P_i$, *type*, $t$**)**, **move(**$P_i, action, a$**)** or **prefer(**$P_i, a, b$**)**.

**Response:** No response required.

**Commitment Store Update:** Exactly one of: (a) the 2-tuple $(type, t)$; (b) the 2-tuple $(action, a)$; or (c) the 3-tuple $(prefer, a, b)$ is deleted from $CS(P_i)$, according to whichever of the three possible prior locutions is being retracted.

**L9:** The **withdraw_dialogue(.)** locution:

**Locution:** **withdraw_dialogue(**$P_i, q?$**)**, where $q?$ is a question.

**Preconditions:** No specific preconditions.

**Meaning:** Participant $P_i$ announces her withdrawal from the deliberation dialogue considering the governing question $q?$.

**Response:** No response required. If only two participants remain in a dialogue and one of these utters this locution, the dialogue terminates.

**Commitment Store Update:** No effects.

# References


[1] R. Alexy. A theory of practical discourse. In S. Benhabib and F. Dallmayr, editors, *The Communicative Ethics Controversy*, Studies in Contemporary German Social Thought, pages 151–190. MIT Press, Cambridge, MA, USA, 1990. Translation by D. Frisby of: Eine Theorie des praktischen Diskurses. In: W. Oelmüller, Editor, *Normenbegründung-Normendurchsetzung.* Paderborn, Germany: Schöningh, 1978.

[2] J. F. Allen, L. K. Schubert, G. Ferguson, P. Heeman, C. H. Hwang, T. Kato, M. Light, N. G. Martin, B. W. Miller, M. Poesio, and D. R. Traum. The TRAINS Project: a case study in building a conversational planning agent. *Journal of Experimental and Theoretical Artificial Intelligence*, 7:7–48, 1995.

[3] S. J. Ambler. A categorical approach to the semantics of argumentation. *Mathematical Structures in Computer Science*, 6:167–188, 1996.

[4] L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In E. Durfee, editor, *Proceedings of the Fourth International Conference on Multi-Agent Systems (ICMAS-2000)*, pages 31–38, Boston, MA, USA, 2000. IEEE Press.

[5] L. Amgoud and S. Parsons. Agent dialogues with conflicting preferences. In J-J. Meyer and M. Tambe, editors, *Pre-Proceedings of the Eighth International Workshop on Agent Theories, Architectures, and Languages (ATAL 2001)*, pages 1–14, Seattle, WA, USA, 2001.

[6] L. Amgoud, S. Parsons, and N. Maudet. Arguments, dialogue, and negotiation. In W. Horn, editor, *Proceedings of the Fourteenth European Conference on Artificial Intelligence (ECAI-2000)*, pages 338–342, Berlin, Germany, 2000. IOS Press.

[7] Aristotle. *Topics*. Clarendon Press, Oxford, UK, 1928. (W. D. Ross, Editor).

[8] T. J. M. Bench-Capon, P. E. Dunne, and P. H. Leng. Interacting with knowledge-based systems through dialogue games. In *Proceedings of the Eleventh International Conference on Expert Systems and Applications*, pages 123–140, Avignon, France, 1991.

[9] A. Church. A formulation of the simple theory of types. *Journal of Symbolic Logic*, 5(2):56–68, 1940.

[10] J. Conklin and M. L. Begeman. gIBIS: A hypertext tool for exploratory policy discussion. In *Proceedings of the Second Conferences on Computer-Supported Co-operative Work*, pages 140–152, New York, NY, USA, 1988. ACM Press.

[11] F. Dignum, B. Dunin-Kęplicz, and R. Verbrugge. Agent theory for team formation by dialogue. In C. Castelfranchi and Y. Lespérance, editors, *Proceedings of the Seventh International Workshop on Agent Theories, Architectures, and Languages (ATAL-2000)*, pages 141–156, Boston, USA, 2000.

[12] F. Dignum, B. Dunin-Kęplicz, and R. Verbrugge. Creating collective intention through dialogue. *Logic Journal of the IGPL*, 9(2):305–319, 2001.

[13] R. M. van Eijk. *Programming Languages for Agent Communications*. PhD thesis, Department of Computer Science, Utrecht University, The Netherlands, 2000.

[14] C. Ess. The political computer: Democracy, CMC, and Habermas. In C. Ess, editor, *Philosophical Perspectives on Computer-Mediated Communication*, pages 197–230. State University of New York Press, Albany, NY, USA, 1996.

[15] R. E. Fikes and N. J. Nilsson. STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2:189–208, 1971.

[16] A. Finkelstein and H. Fuks. Multi-party specification. In *Proceedings of the Fifth International Workshop on Software Specification and Design*, Pittsburgh, PA, USA, 1989. ACM Sigsoft Engineering Notes.

[17] D. J. Fiorino. Environmental risk and democratic process: a critical review. *Columbia Journal of Environmental Law*, 14:501–547, 1989.

[18] FIPA. Communicative Act Library Specification. Technical Report XC00037H, Foundation for Intelligent Physical Agents, 10 August 2001.

[19] J. Forester. *The Deliberative Practitioner: Encouraging Participatory Planning Processes*. MIT Press, Cambridge, MA, USA, 1999.

[20] D. M. Gabbay and J. Woods. Non-cooperation in dialogue logic. *Synthese*, 127(1-2):161–186, 2001.

[21] T. F. Gordon and N. Karacapilidis. The Zeno argumentation framework. In *Proceedings of the Sixth International Conference on AI and Law*, pages 10–18, New York, NY, USA, 1997. ACM Press.

[22] H. P. Grice. Logic and conversation. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics III: Speech Acts*, pages 41–58. Academic Press, New York City, NY, USA, 1975.

[23] B. J. Grosz and S. Kraus. The evolution of SharedPlans. In M. J. Wooldridge and A. Rao, editors, *Foundations of Rational Agency*, volume 14 of *Applied Logic*. Kluwer, The Netherlands, 1999.

[24] J. Habermas. *Moral Consciousness and Communicative Action*. MIT Press, Cambridge, MA, USA, 1991. Translation by C. Lenhardt and S. W. Nicholsen of: *Moralbewusstsein un kommunikatives Handeln.* Suhrkamp, Frankfurt-am-Main, Germany, 1983.

[25] J. Habermas. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. MIT Press, Cambridge, MA, USA, 1996. Translation by W. Rehg of: *Faktizität und Geltung. Beiträge zur Diskurstheorie des Rechts und des demokratischen Rechtsstaats.* Published by Suhrkamp, Frankfurt am Main, Germany, 1992.

[26] J. Habermas. *The Inclusion of the Other: Studies in Political Theory*. MIT Press, Cambridge, MA, USA, 1998. Edited by C. Cronin and P. De Greiff.

[27] C. L. Hamblin. *Fallacies*. Methuen and Co Ltd, London, UK, 1970.

[28] J. Herbst. *State Politics in Zimbabwe*, volume 45 of *Perspectives on Southern Africa*. University of California Press, Berkeley, CA, USA, 1990.

[29] D. Hitchcock. Some principles of rational mutual inquiry. In F. van Eemeren, R. Grootendorst, J. A. Blair, and C. A. Willard, editors, *Proceedings of the Second International Conference on Argumentation*, pages 236–243, Amsterdam, The Netherlands, 1991. SICSAT: International Society for the Study of Argumentation.

[30] J. Hulstijn. *Dialogue Models for Inquiry and Transaction*. PhD thesis, Universiteit Twente, Enschede, The Netherlands, 2000.

[31] L. Hunsberger and M. Zancanaro. A mechanism for group decision making in collaborative activity. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI 2000)*, pages 30–35, Menlo Park, CA, USA, 2000. AAAI Press.

[32] R. Johnson. *Manifest Rationality: A Pragmatic Theory of Argument*. Lawrence Erlbaum Associates, Mahwah, NJ, USA, 2000.

[33] N. Karacapilidis and D. Papadias. Computer supported argumentation and collaborative decision making: the HERMES system. *Information Systems*, 26(4):259–277, 2001.

[34] N. Karacapilidis, D. Papadias, T. Gordon, and H. Voss. Collaborative environmental planning with GeoMed. *European Journal of Operational Research*, 102(2):335–346, 1997.

[35] E. Krabbe. The problem of retraction in critical discussion. *Synthese*, 127(1-2):141–159, 2001.

[36] I. Lakatos. *Proofs and Refutations: The Logic of Mathematical Discovery*. Cambridge University Press, Cambridge, UK, 1976.

[37] J. A. Levin and J. A. Moore. Dialogue-games: metacommunications structures for natural language interaction. *Cognitive Science*, 1(4):395–420, 1978.

[38] P. Lorenzen and K. Lorenz. *Dialogische Logik*. Wissenschaftliche Buchgesellschaft, Darmstadt, Germany, 1978.

[39] R. Luehrs, T. Malsch, and K. Voss. Internet, discourses, and democracy. In T. Terano, T. Nishida, A. Namatame, S. Tsumoto, Y. Ohsawa, and T. Washio, editors, *New Frontiers in Artificial Intelligence: Joint JSAI 2001 Workshop Post Proceedings*, Lecture Notes in Artificial Intelligence 2253, pages 67–74. Springer, Berlin, Germany, 2001.

[40] J. Lukacs. *Five Days in London: May 1940*. Yale University Press, New Haven, CT, USA, 1999.

[41] J. D. MacKenzie. Question-begging in non-cumulative systems. *Journal of Philosophical Logic*, 8:117–133, 1979.

[42] P. McBurney, R. M. van Eijk, S. Parsons, and L. Amgoud. A dialogue-game protocol for agent purchase negotiations. *Journal of Autonomous Agents and Multi-Agent Systems*, 2002. (*In press*).

[43] P. McBurney and S. Parsons. Representing epistemic uncertainty by means of dialectical argumentation. *Annals of Mathematics and Artificial Intelligence*, 32(1–4):125–169, 2001.

[44] P. McBurney and S. Parsons. Games that agents play: A formal framework for dialogues between autonomous agents. *Journal of Logic, Language and Information*, 11(3):315–334, 2002.

[45] P. McBurney, S. Parsons, and M. Wooldridge. Desiderata for agent argumentation protocols. In C. Castelfranchi and W. L. Johnson, editors, *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002)*, Bologna, Italy, 2002. ACM Press. *In press*.

[46] A. Naess. *Communication and Argument: Elements of Applied Semantics*. Allen and Unwin, London, UK, 1966. Translation of: *En del Elementaere Logiske Emner.* Universitetsforlaget, Oslo, Norway, 1947.

[47] K. Nakata. Enabling public discourse. In T. Terano, T. Nishida, A. Namatame, S. Tsumoto, Y. Ohsawa, and T. Washio, editors, *New Frontiers in Artificial Intelligence: Joint JSAI 2001 Workshop Post Proceedings*, Lecture Notes in Artificial Intelligence 2253, pages 59–66. Springer, Berlin, Germany, 2001.

[48] A. J. Nathan and P. Link, editors. *The Tiananmen Papers*. Little, Brown and Company, London, UK, 2001. Compiled by Zhang Liang, with an afterword by O. Schell.

[49] H. P. Nii. Blackboard systems (Part One): The blackboard model of problem solving and the evolution of blackboard architectures. *The AI Magazine*, Summer:38–53, 1986.

[50] P. Panzarasa, N. R. Jennings, and T. J. Norman. Formalizing collaborative decision-making and practical reasoning in multi-agent systems. *Journal of Logic and Computation*, 12(1):55–117, 2002.

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

[52] S. Parsons, M. Wooldridge, and L. Amgoud. An analysis of formal interagent dialogues. In C. Castelfranchi and W. L. Johnson, editors, *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002)*, Bologna, Italy, 2002. AAMAS. (*In press*).

[53] A. Pekeč and M. H. Rothkopf. Combinatorial auction design. Working Paper, July 2001. Presentation to Workshop on Electronic Market Design, Infonomics Group, University of Maastricht, Maastricht, The Netherlands.

[54] H. Prakken. On dialogue systems with speech acts, arguments, and counterarguments. In M. Ojeda-Aciego, M. I. P. de Guzman, G. Brewka, and L. M. Pereira, editors, *Proceedings of the Seventh European Workshop on Logic in Artificial Intelligence (JELIA-2000)*, Lecture Notes in Artificial Intelligence 1919, pages 224–238, Berlin, Germany, 2000. Springer.

[55] C. Reed. Dialogue frames in agent communications. In Y. Demazeau, editor, *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS-98)*, pages 246–253. IEEE Press, 1998.

[56] W. Rehg. The argumentation theorist in deliberative democracy. *Controversia*, 2002 (*In press*). Keynote Address at the Second Conference of the International Debate Education Association (IDEA), Prague, Czech Republic, 13 October 2001.

[57] H. W. J. Rittel and M. M. Webber. Dilemmas in a general theory of planning. *Policy Sciences*, 4:155–169, 1973.

[58] J. S. Rosenschein and G. Zlotkin. *Rules of Encounter: Designing Conventions for Automated Negotiation among Computers*. Artificial Intelligence Series. MIT Press, Cambridge, MA, USA, 1994.

[59] F. Sadri, F. Toni, and P. Torroni. Logic agents, dialogues and negotiation: an abductive approach. In M. Schroeder and K. Stathis, editors, *Proceedings of the Symposium on Information Agents for E-Commerce, Artificial Intelligence and the Simulation of Behaviour Conference (AISB 2001)*, York, UK, 2001. AISB.

[60] T. W. Sandholm. Distributed rational decision making. In G. Weiss, editor, *Multiagent Systems: A Modern Introduction to Distributed Artificial Intelligence*, pages 201–258. MIT Press, Cambridge, MA, USA, 1999.

[61] C. Sierra, N. R. Jennings, P. Noriega, and S. Parsons. A framework for argumentation-based negotiation. In M. P. Singh, A. Rao, and M. J. Wooldridge, editors, *Intelligent Agents IV: Agent Theories, Architectures, and Languages. Proceedings of the Fourth International ATAL Workshop*, Lecture Notes in Artificial Intelligence 1365, pages 177–192, Berlin, Germany, 1998. Springer.

[62] M. P. Singh. A social semantics for agent communications languages. In F. Dignum, B. Chaib-draa, and H. Weigand, editors, *Proceedings of the Workshop on Agent Communication Languages, International Joint Conference on Artificial Intelligence (IJCAI-99)*, Berlin, Germany. Springer. *(In press).*

[63] M. P. Singh. A conceptual analysis of commitments in multiagent systems. Technical Report 96-09, Department of Computer Science, North Carolina State University, Raleigh, NC, USA, 1996.

[64] R. D. Tennent. *Semantics of Programming Languages*. International Series in Computer Science. Prentice-Hall, Hemel Hempstead, UK, 1991.

[65] D. N. Walton and E. C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of New York Press, Albany, NY, USA, 1995.

[66] T. Webler, S. Tuler, and R. Krueger. What is a good public participation process? Five perspectives from the public. *Environmental Management*, 27(3):435–450, 2001.

[67] H. Wohlrapp. A new light on non-deductive argumentation schemes. *Argumentation*, 12:341–350, 1998.

[68] M. Wooldridge. Semantic issues in the verification of agent communication languages. *Journal of Autonomous Agents and Multi-Agent Systems*, 3(1):9–31, 2000.