Desiderata for Agent Argumentation Protocols

Peter McBurney, Simon Parsons and Michael Wooldridge
Department of Computer Science
University of Liverpool
Liverpool L69 7ZF United Kingdom
{p.j.mcburney,s.d.parsons,m.j.wooldridge}@csc.liv.ac.uk

ABSTRACT

Designers of agent communications protocols are increasingly using formal dialogue games, adopted from argumentation theory, as the basis for structured agent interactions. We propose a set of desiderata for such protocols, drawing on recent research in agent interaction, on recent criteria for assessment of automated auction mechanisms and on elements of argumentation theory and political theory. We then assess several recent dialogue game protocols against our desiderata, revealing that each protocol has serious weaknesses. For comparison, we also assess the FIPA Agent Communications Language (ACL), thereby showing FIPA ACL to have limited applicability to dialogues not involving purchase negotiations. We conclude with a suggested checklist for designers of dialogue game protocols for agent interactions.

KEYWORDS: Argumentation, Agent Communication Languages, Dialogue Games, Foundation for Intelligent Physical Agents (FIPA), Interaction Protocols.

1. INTRODUCTION

Formal dialogue games are games in which two or more participants "move" by uttering locutions, according to certain pre-defined rules. They have been studied by philosophers since the time of Aristotle, most recently for the contextual modeling of fallacious reasoning [14, 23] and as a proof-theoretic semantics for intuitionistic and classical logic [22]. Outside philosophy, dialogue games have been used in computational linguistics, for natural language explanation and generation, and in artificial intelligence (AI), for automated software design and the modeling of legal reasoning, e.g., [4]. In recent years, they have found application as the basis for communications protocols between autonomous software agents, including for agents engaged in: negotiation dialogues, in which participating agents seek to agree a division of some scarce resource [3, 17, 27, 30]; persuasion dialogues, where one agent seeks to persuade another to endorse some claim [2, 7, 8]; information-seeking dialogues, where one agent seeks the answer to some question from another [17]; inquiry dialogues, where several agents jointly seek the answer to some question [24]; and deliberation dialogues, where participants seek to jointly agree a course of action in some situation [16].[1]

Why have agent protocol designers turned to dialogue games from argumentation theory? It is reasonable to assume that a rational agent would only change its beliefs or its preferences after receiving new information, i.e., not on the basis of whim or malice, say. For an agent to acquire new information in a dialogue, it needs a way to probe or challenge the statements of other agents; thus, locutions which enable utterances to be questioned or contested are required, along with locutions which enable appropriate responses to these. In order that this process takes place in an orderly and efficient fashion, we require rules which govern what can and cannot and must be said in a dialogue, and when. Dialogue games provide a framework for the design of such structured discourses, drawing on specific theories of argumentation. We would expect that the greater the amount of relevant information passed between participants, the greater is the likelihood of successful resolution of the dialogue; increased likelihood of resolution is therefore the expected payoff of these games when compared with more parsimonious interaction protocols, such as auctions.

However, despite this recent interest in dialogue-game protocols for multi-agent systems, we know of no discussion of appropriate design principles. As these protocols proliferate, designers and users will require means to assess protocols and to compare one with another. In this paper, we therefore propose the first list of desiderata to govern the design and assessment of dialogue game protocols. To do this, we have drawn upon: the criteria recently proposed for assessment of automated auction and negotiation mechanisms in, e.g., [31]; theories of deliberative decision-making from argumentation theory [1, 15] and political theory [6, 10, 12]; and recent studies of agent communications languages and interaction protocols [13, 20, 33, 36]. We believe our list of desiderata will be an initial step towards the development of formal design and assessment criteria for agent argumentation protocols.

2. PROPOSED DESIDERATA

We begin by assuming that agents engaged in dialogues are autonomous, willing and free participants, able to enter and withdraw from dialogues as and when they see fit. Within each dialogue, they remain autonomous, and are not compelled to accept or reject any proposition. These assumptions have implications for some of the desiderata, as explained below. We also assume that the specification of a dialogue game protocol consists of: (a) a set of topics of discussion (which may be represented in some logical language);

[1] This typology of dialogues is from [35], and we use it throughout this paper; note that *negotiation* and *deliberation* dialogues are defined more precisely than is usually the case within AI.

AAMAS'02, July 15-19, 2002, Bologna, Italy.

(b) the syntax for a set of defined locutions concerning these topics; (c) a set of rules which govern the utterance of these locutions; (d) a set of rules which establish what commitments, if any, participants create by the utterance of each locution; and (e) a set of rules governing the circumstances under which the dialogue terminates. We refer to such a specification as a *Dialectical System*.[2] We now list our desiderata with a brief explanation for each:

1. **Stated Dialogue Purpose:** A dialectical system should have one or more publicly-stated purposes, and its locutions and rules should facilitate the achievement of these. For example, the stated purpose of a system for negotiation may be an agreement on the division of a particular scarce resource; negotiation over a different resource will result in a different purpose. Likewise, a discussion about the same resource which is not a negotiation over its division constitutes a different purpose, e.g., it may be an information-seeking dialogue. The dialogue purposes need to be stated, so that all participating agents are aware of them in advance of entering the dialogue. Successful resolution of a dialogue will occur when its stated purposes are achieved.

2. **Diversity of Individual Purposes:** A dialectical system should permit participating agents to achieve their own individual purposes consistent with the overall purpose of the dialogue. These individual purposes may conflict, as when parties to a negotiation each seek to maximize their individual utility in any outcome, or they may coincide, as when agents collectively seek to answer some unknown question.

3. **Inclusiveness:** A dialectical system should not preclude participation by any potential agent which is qualified and willing to participate. Because agents are autonomous entities, there is a sense in which all agents are deserving of equal respect. As with human beings [6], agents affected by decisions have a moral right to be included in deliberations leading to those decisions. In addition, inclusion of affected parties in decisions can improve the quality of the decision outcomes [10].

4. **Transparency:** Participants to a dialogue should know the rules and structure of the dialectical system prior to commencement of the dialogue. In particular, any reference from dialogues in a dialectical system to an external reality should be explicitly stated, and known to the participants before commencement, e.g., when commitments incurred inside a purchase negotiation dialogue imply subsequent real-world obligations to execute a particular commercial transaction.

5. **Fairness:** A dialectical system should either treat all participants equally, or, if not, make explicit any asymmetries in their treatment. For instance, it may be appropriate for participants to play different roles in a dialogue, such as sellers, buyers and auctioneers in a purchase transaction dialogue [32]. Agents in these different roles may have different rights and responsibilities, and these should be known to all.

6. **Clarity of Argumentation Theory:** A dialectical system should conform, at least at the outset, to a stated theory of argument, for example Hitchcock's Principles for Rational Mutual Inquiry [15] or the persuasion dialogue rules of [9]. The reason for this is so that all participants know, and adhere to, their dialectical obligations, agree on rules of inference and procedure, and have reasonable expectations of the responses of others. For example, an agent should know in advance of making an assertion that its statement may incur obligations to defend it upon contestation by others; likewise, agents contesting an assertion should know if they are entitled to receive a defence of it. The dialogue-game rules which embody a theory of argumentation ensure that such arguments are conducted in an orderly and efficient manner. If dialogue participants wish to change the argumentation-theoretic basis or the dialogical rules of the system in the course of using it for a particular dialogue, being free agents, they should be enabled to do so.[3]

7. **Separation of Syntax and Semantics:** The syntax of a dialectical system should be defined separately from its semantics. There are two reasons for this. Firstly, this approach enables the same protocol syntax to be used with multiple semantics. Secondly, the problem of semantic verification of an agent communications language is a thorny one [36], since it will always be possible for a sufficiently-clever agent to simulate insincerely any required internal state. Ensuring that the protocol syntax is defined separately from its semantics therefore enables the verification of conformity with protocol syntax, even if the protocol semantics cannot be completely verified.[4] The recent development of a *social semantics*, where agents first express publicly their beliefs and intentions relevant to an interaction, may be seen as an attempt to extend the domain of verifiability [33].

8. **Rule-Consistency:** The locutions and rules of a dialogue system should together be internally consistent; that is, they should not lead to deadlocks (where no participant may utter a legal locution), nor infinite cycles of repeated locutions.

9. **Encouragement of Resolution:** Resolution of each dialogue (normal termination) should be facilitated, and not precluded, by the locutions and rules of a dialectical system.

10. **Discouragement of Disruption:** Normally, the rules of a dialectical system should discourage or preclude disruptive behaviour, such as uttering the same locution repeatedly. However, as Krabbe notes with regard to retraction [18], achieving a balance between outlawing disruptive behaviour and permitting freedom of expression is not necessarily straightforward, and will differ by application.

11. **Enablement of Self-Transformation:** A dialectical system should permit participants to undergo self-transformation [12] in the course of a dialogue; e.g., participants to a negotiation should be able to change their preferences or their valuations of utility as a result of information they receive from others in the dialogue, or express degrees of belief in propositions. In particular, participants should have the right to retract commitments made earlier in the same dialogue, although not necessarily always unconditionally. If the protocol does not permit such transformation, then one agent would not be able to persuade another to change its beliefs or to adopt a proposal it had previously rejected; in such circumstances, there would be no point for the agents to engage in dialogue.

[2] This model is presented in [25]. Note that there is no consensus among philosophers over distinctions, if any, between the words "dialogical" and "dialectical" [5, p. 337].

[3] This last property is called *dialectification* in [15].

[4] Expressing the rules of dialogue in terms of observable linguistic behaviour is called *externalization* in [15].

12. **System Simplicity:** The locutions and rules of a dialectical system should be as simple as possible, consistent with the eleven criteria above. In particular, each locution should serve a specific and stated function in the dialogue, and the protocol rules should lead to efficient achievement of the dialogue purposes.

13. **Computational Simplicity:** A dialectical system should be designed to minimize any computational demands on its participants, and on the system itself, consistent with the twelve criteria above.

It is important to note two criteria we have not included here. We have not specified that dialectical systems should be realistic representations of some human dialogue, as we see no reason why agent interactions should necessarily adopt human models of interaction. Indeed, dialectical systems may be applied to agent dialogues which humans do not, or, even, could never undertake, such as simultaneous negotiations over multiple products with hundreds of participants. Secondly, we have not stated that the rules of a dialectical system should require that the participants use particular rules of inference (such as Modus Ponens), particular logics or particular decision-making procedures, or that the participating agents satisfy some criterion of rational behaviour, such as acting to maximize expected utility. Insisting on such rules and criteria is contrary to the notion of agent autonomy we assumed at the outset.

In addition to the thirteen principles listed above, there may be further desiderata appropriate for specific types of dialogue. For instance, for dialogues undertaken to negotiate a division of a scarce resource, it may be considered desirable that outcomes are Pareto optimal, i.e., that any other outcome leaves at least one participant worse off [31]. Because we assume agents are free and willing participants in a dialogue, acting under no duress, then any agreed outcome to a negotiation dialogue will satisfy this particular criterion, if certain of the above desiderata are met. We present the result formally, so as to make clear the assumptions needed:

Proposition: *Suppose two or more agents, each of which is purely self-interested and without malice, engage freely and without duress in a negotiation dialogue, i.e., a dialogue to agree a division of some scarce resource. Suppose these agents use a dialogue protocol which satisfies desiderata 2, 4, and 5, and that this dialogue is conducted with neither time constraints nor processing-resource constraints. Suppose further that their negotiation dialogue achieves resolution, i.e., they agree on a division of the resource in question. Then the outcome reached is Pareto Optimal.*

Proof. Suppose that the outcome reached, which we denote by X, is not Pareto Optimal. Then there is another outcome, Y, which leaves at least one agent, say agent a, better off, while all other agents are no worse off. Then, it behooves agent a to suggest Y rather than X for agreement by the participants, since agent a (like all participants) is self-interested. If Y is suggested by agent a, the other agents will at least be indifferent between Y and X, because they are no worse off under Y and may be better off; so, being without malice, the others should support proposal Y over X in the dialogue. Now, the only reasons agent a would not suggest Y would be because of: resource-constraints precluding the identification of Y as a better outcome; time-constraints precluding the making of the suggestion of Y in the dialogue; constraints imposed on agent a by the protocol itself, e.g., rules precluding that particular agent making suggestions; or social pressures exerted by other agents on agent a which prevent the suggestion of Y being made. Each of these reasons contradicts an assumption of the proposition, and so X must be Pareto Optimal. □

Of course, if the agents are not purely self-interested, or not free of duress, or if they enter the discussion under constraints such as resolution deadlines, then any agreement reached may not be Pareto optimal. Since most agents in most negotiation dialogues will be subject to resource- and time-constraints, Pareto optimality may be seen as a (mostly) unachievable ideal for agent negotiation dialogues. An interesting question would be the extent to which any given negotiation dialogue outcome approximates a Pareto optimal outcome.

Finally, it is important to note that these desiderata, particularly numbers 6 *(Clarity of Argumentation Theory)* and 11 *(Enablement of Self-Transformation)*, express a particular view of joint decision-making by autonomous entities. Political theorists distinguish *rational-choice* or *marketplace* models from *deliberative democracy* models of social and public decision-making [6]. Rational-choice models assume that each participant commences the decision-process with his or her beliefs, utilities and preferences fully formed and known (at least to him/herself); each participant then chooses between (i.e., votes for or against) competing proposals on the basis of his or her own beliefs and preferences. Such a model does not allow for beliefs and preferences to be determined in the course of the interaction, nor for participants to acquire a *group* view of the issues involved in the decision, for instance, the wider social consequences of individual actions [28]. In contrast, deliberative democracy models of joint decision-making emphasize the manner in which beliefs and preferences are formed or change through the very process of interacting together, with participants undergoing what has been called self-transformation [12, p. 184]. In a rational decision-process, this transformation occurs by the sharing of information, by challenging and defending assertions, by persuasion, and by joint consideration of the relevant issues — i.e., by argument and debate. Because we assume that software agents are autonomous, then such argument will be required to convince other agents to adopt specific beliefs and to commit to specific intentions; we believe, therefore, that a society of autonomous agents is best viewed as a deliberative democracy, and not as simply a marketplace.

Comparison with Game Theoretic Models

At this point it is worth discussing the relationship of argumentation protocols to work on game-theoretic approaches to negotiation, of which perhaps the best known examples are [19, 29]. An example of the kind of issue investigated in this work is how agents with tasks to carry out in some environment can divide the tasks amongst themselves to their mutual betterment — the task oriented domains of [29, pp.29–52]. The key abstraction in this work is that the utility of possible deals in the domain of negotiation (whatever agents are negotiating over) can be assessed for any individual agent.

Perhaps the greatest attraction of game-theoretic approaches to negotiation is that it is possible to prove many desirable features of a given negotiation protocol. Examples of such properties include [31, p.204]:

1. **Maximising social welfare**. Intuitively, a protocol maximises social welfare if it ensures that any outcome maximises 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 agent received in the outcome, then a protocol that maximised social welfare would maximise the *total* amount of money "paid out."

2. **Pareto efficiency**. As discussed above.

3. **Individual rationality**. A protocol is said to be individually rational if following the protocol — "playing by the rules"

— is in the best interests of negotiation participants. Individually rational protocols are essential because, without them, there is no incentive for agents to engage in negotiations.

4. **Stability**. A protocol is *stable* if it provides all agents with an incentive to behave in a particular way. The best-known kind of stability is Nash equilibrium.

5. **Simplicity**. A "simple" protocol is one that makes the appropriate strategy for a negotiation participant "obvious". That is, a protocol is simple if using it, a participant can easily (tractably) determine the optimal strategy.

6. **Distribution**. A protocol should ideally be designed to ensure that there is no "single point of failure" (such as a single arbitrator), and so as to minimise communication between agents.

It is worth comparing our desiderata with these. We do not explicitly state maximising social welfare, as there is no general notion of this in dialogue games. As we demonstrated above, our criteria imply Pareto optimal outcomes. Individual rationality amount to our criterion of individual purpose: an agent cannot be forced to participate. We do not assume stability, but we do assume that there in an incentive to resolve the dialogue, i.e., that it is in an agent's interests to participate in the successful conclusion of the dialogue. We explicitly assume simplicity. Finally, we do not explicitly consider distribution.

One of the best-known results in the area of game-theoretic negotiation is Nash's *axiomatic approach to bargaining*, an attempt to axiomatically define the properties that a "fair" outcome to negotiation would satisfy — a desideratum, in effect, for negotiation [29, pp.50–52]. The properties he identified were: (i) individual rationality (a participant should not *lose* from negotiation); (ii) Pareto optimality; (iii) symmetry; (iv) invariance with respect to linear utility transformations (e.g., if one agent counts utility in cents, while the other counts it in dollars, it should make no difference to the outcome); and (v) independence of irrelevant alternatives. Nash proved that *mechanisms that guarantee an outcome that maximises the product of the utilities of participant agents satisfy these criteria, and moreover, are the* only *mechanisms that satisfy these criteria*. We have begun working on the formalisation of dialogue games [25, 26], with the goal of making these desiderata formal. As a next step, it would be interesting to determine the extent to which Nash's results transfer to our dialogue game framework.

3. DIALOGUE GAME PROTOCOLS

In this section, we examine three recent proposals for dialogue game protocols against the desiderata presented above. The three protocols, which are representative of the literature, have been selected because they concern three different types of agent interactions. However, not all elements of these protocols are fully specified, thus making it impossible to assess them against some of the desiderata. In these cases, we write: *Unable to assess*.

3.1 A negotiation dialogue

We first consider a dialogue-game protocol for agent negotiation dialogues proposed by Amgoud, Parsons and Maudet in [3], drawing on the philosophical dialogue game *DC* of [23].[5] This agent interaction protocol comprises seven distinct locutions: *assert, question, challenge, request, promise, accept* and *refuse*, and these can be variously instantiated with: single propositions; arguments for propositions (comprised of sets of propositions); or certain types of implication. For example, the locution *promise*($p \Rightarrow q$) indicates a promise by the speaker to provide resource q in return for resource p. Arguments may be considered to be tentative proofs, i.e., logical inferences from assumptions which may not all be confirmed. The syntax for this protocol has only been provided for dialogues between two participants, but could be readily extended to more agents.

Following [14], when an agent asserts something (a proposition, an argument, or an implication), this something is inserted into a public commitment store accessible to both participants. Thus, participants are able to share information. In [3], the protocol was given an operational semantics in terms of a formal argumentation system. In this semantics, an agent can only utter the locution *assert*(p), for p a proposition, if that agent has an acceptable argument for p in its own knowledge base, or in its knowledge base combined with the public commitment stores. (Acceptable arguments are those which survive attack from counter-arguments in a defined manner.) The semantics provided, however, is not sufficient for automated dialogues.

1. **Stated Dialogue Purpose:** The protocol is explicitly for negotiation dialogues, but the syntax does not require the participants to state the purpose(s) of the specific negotiation dialogue undertaken.

2. **Diversity of Individual Purposes:** This is enabled.

3. **Inclusiveness:** There do not appear to be limitations on which agents may participate.

4. **Transparency:** The protocol rules are transparent, and the authors present the pre-conditions and post-conditions of each locution.

5. **Fairness:** Locutions are only given for one participant (*Proponent*), with an implicit assumption that they are identical for the other (*Contestor*).

6. **Clarity of Argumentation Theory:** The definitions of protocol syntax and semantics assume an explicit theory of argumentation.

7. **Separation of Syntax and Semantics:** The syntax is defined in terms of the argumentation theory semantics, but could be readily defined separately.

8. **Rule-Consistency:** The rules appear to be consistent.

9. **Encouragement of Resolution:** The protocol does not appear to discourage resolution of the negotiation.

10. **Discouragement of Disruption:** Disruption is not discouraged, as there are no rules preventing or minimizing this behaviour. For instance, there are no rules precluding the repeated utterance of the same locution by an agent, although there is such a condition in the argumentation semantics given for the dialogue protocol.[6]

11. **Enablement of Self-Transformation:** Self-transformation is not enabled. Agents may *add* to their knowledge base from the commitment stores of other participants, but there appears to be no mechanism for their knowledge base to change or to diminish. Because transformation is not enabled, there are no retraction locutions.

[5] This dialogue game was designed to enable persuasion dialogues which would preclude circular reasoning.

[6] The dialogue game *DC* also lacks such a rule [23].

12. **System Simplicity:** There do not appear to be extraneous locutions.

13. **Computational Simplicity:** Unable to assess. The computational complexity of the semantic argumentation mechanism may be high.

We believe the key weakness of this protocol is the absence of self-transformation capability. The protocol also makes several implicit assumptions, which may limit its applicability. Firstly, the protocol assumes the interaction is between agents with fixed (although possibly different) knowledge bases and possibly divergent interests. Secondly, the absence of rules precluding disruptive behaviour and rules for termination conditions, of an explicit statement of objectives and of formal entry and exit locutions suggest an implicit assumption that the participants are rational and share some higher goals. Thirdly, although the semantic argumentation framework allows agents to hold internally preferences regarding arguments, the dialogue protocol does not allow for these to be expressed in the dialogue; nor are degrees of belief or acceptability in propositions and arguments expressible. Allowing such expression should increase the likelihood of successful resolution of a negotiation dialogue. For example, this protocol does not permit the making of tentative suggestions — propositions uttered for which the speaker does not yet have an argument.

3.2 A persuasion dialogue

We next consider the protocol proposed by Dignum, Dunin-Kęplicz and Verbrugge [8] for the creation of collective intention by a team of agents. The protocol assumes that a team has already been formed, and that one agent, an *initiator* or *proponent*, seeks to persuade others (*opponents*) in the team to adopt a group belief or intention. For this dialogue, the authors adapt the rigorous persuasion dialogue-game of [35], which is a formalization of a rigorous persuasion dialogue in philosophy. Such dialogues involve two parties, one seeking to prove a proposition, and one seeking to disprove it.[7] The protocol presented by Dignum *et al.* includes seven locutions: *statement, question, challenge, challenge-with-statement, question-with-statement* and *final remarks*; these last include: *"quit"* and *"won"*. The statements associated with challenges and questions may be concessions made by the speaker.

1. **Stated Dialogue Purpose:** The protocol is explicitly for a persuasion dialogue when an initiating agent seeks to *"establish a collective intention within a group"* [8, p. 313]. The syntax requires the initiator to state explicitly the intention it desires the group to adopt.

2. **Diversity of Individual Purposes:** The protocol assumes a conflict of objectives by the participants, but not agreement.

3. **Inclusiveness:** There do not appear to be limitations on which agents may participate.

4. **Transparency:** The protocol rules are transparent to the participating agents. However, they are not yet fully specified, since the authors do not articulate the pre-conditions and post-conditions of each utterance, or all the rules governing their use.

5. **Fairness:** Following [35], the protocol rules are asymmetrical: the initiator has different rights and obligations from opponents. However, these differences are known to the participants.

[7] Note that the persuasion dialogues of [35] deal only with beliefs and not intentions.

6. **Clarity of Argumentation Theory:** The critical persuasion dialogues for which the dialogue-game formalism [35] was developed are idealizations of human dialogues, used by philosophers to study fallacious reasoning. This underlying argumentation theory is not stated explicitly in [8], nor is it self-evidently appropriate for agent interactions.

7. **Separation of Syntax and Semantics:** The locutions and syntax of the dialogue are not fully articulated. A partial operational semantics is provided in terms of the beliefs and intentions of the participating agents. To the extent that the syntax and semantics are specified, they appear to be defined separately.

8. **Rule-Consistency:** Unable to assess this, as the rules are not fully articulated.

9. **Encouragement of Resolution:** The argumentation theory underlying the protocol assumes the participants have contrary objectives, which is not necessarily the case. By assuming antagonism where this is none, the protocol may discourage resolution.

10. **Discouragement of Disruption:** Disruption is not discouraged, as there are no rules preventing or minimizing this behaviour. For instance, there are no rules precluding the repeated utterance of the same locution by an agent.

11. **Enablement of Self-Transformation:** Self-transformation is enabled. However, because the syntax and semantics are not fully articulated, it is not clear how this is achieved.[8] In addition, this protocol does not permit degrees of belief or acceptability to be expressed, nor does it permit retractions of prior statements.

12. **System Simplicity:** There do not appear to be extraneous locutions. However, following [35], participants may only speak in alternating sequence, and the rules of the dialogue are quite strict.

13. **Computational Simplicity:** Unable to assess. The computational complexity of the semantic mechanism may be high, as the authors concede.

This protocol is difficult to assess against the desiderata because the locutions, rules of syntax and the semantics are not fully articulated. In addition, the authors present no case for using the rigorous persuasion dialogue game adapted from [35] in the agent domain. This game embodies an explicit theory of argumentation which is not necessarily appropriate for agent dialogues. In particular, the theory assumes participants are engaged in a critical persuasion, and thus have conflicting objectives (namely to prove or disprove a proposition); consequently the rules and locutions are stricter than most people would consider appropriate for an ordinary (human or agent) persuasion dialogue. Participants must speak in alternating sequence, for example. Moreover, the rigorous persuasion dialogue is based on the dialogue games of [22], originally designed as a constructive proof-theory for logical propositions. While such games can be used to construct, step-by-step, an argument for a proposed group intention, this would seem a singularly inefficient means of persuasion. Allowing agents to express a complete argument for a proposal in one utterance, as Amgoud *et al.* permit

[8] An agent asked by an initiator to adopt an intention which conflicts with an existing intention may challenge the initiator to provide a proof for the proposed intention, but the authors do not indicate when and how that proof leads to a revision of the existing intention [8, Section 4.2.4].

in the negotiation protocol assessed above, would seem far more efficient.

3.3 An inquiry dialogue

We now consider a dialogue game protocol proposed by McBurney and Parsons for inquiry dialogues in scientific domains [24]. This presents 30 locutions, which enable participants to propose, assert, question, accept, contest, retract, and refine claims, the arguments for them, the assumptions underlying and the rules of inference used to derive these arguments, and the consequences of claims. The protocol is based on a specific philosophy of science, due to Feyerabend and Pera, which stresses the dialogical nature of scientific knowledge-development. In addition to the protocol syntax, a game-theoretic semantics is presented, linking arguments in the dialogue after finite times with the long-run (infinite) position of the dialogue.[9]

1. **Stated Dialogue Purpose:** There is no stated dialogue purpose.
2. **Diversity of Individual Purposes:** This is enabled.
3. **Inclusiveness:** There do not appear to be limitations on which agents may participate.
4. **Transparency:** The protocol rules are transparent, and the authors present the pre-conditions and post-conditions of each locution, along with rules governing the combination of locutions.
5. **Fairness:** The rules treat all participants equally. Utterance of specific statements may incur obligations on the agent concerned.
6. **Clarity of Argumentation Theory:** The protocol conforms explicitly to a specified philosophy of scientific discourse, uses Toulmin's well-known model of an individual argument [34], and adheres to most of the principles of rational human discourse proposed by Alexy and Hitchcock [1, 15]. The conformance of the protocol is demonstrated formally.
7. **Separation of Syntax and Semantics:** These are defined separately.
8. **Rule-Consistency:** The rules appear to be consistent.
9. **Encouragement of Resolution:** The protocol does not appear to discourage resolution, but no rules for termination are provided. Because this is a model for scientific dialogues, it is assumed to be of possibly-infinite duration.
10. **Discouragement of Disruption:** The rules prohibit multiple utterances of the same locution, but not other forms of disruption.
11. **Enablement of Self-Transformation:** Self-transformation is enabled. Assertions may be retracted and qualified. In addition, degrees of belief in propositions and rules of inference may be expressed. However, the mechanisms of self-transformation internal to an agent are not presented.
12. **System Simplicity:** With 30 locutions, the protocol is not simple.
13. **Computational Simplicity:** Unable to assess.

[9]The purpose of this semantics is to assess to what extent a finite shapshot of a debate is representative of the long-run counterpart, in order to assess the likelihood of dialogues conducted under the protocol finding the answer to the question at issue.

The semantics provided for this protocol is not an operational semantics, and no assumptions are made concerning the internal architectures of the participating agents. Deciding what locutions to utter in a dialogue under this protocol may be computationally difficult for a participating agent, particularly if the number of participants is large and the topics discussed diverse. The absence of a stated dialogue purpose means that any topic may be discussed at any time. This is a significant weakness of the protocol.

4. THE FIPA ACL

Finally, by way of comparison, we consider the Agent Communications Language of FIPA, the Foundation for Intelligent Physical Agents [11]. The FIPA ACL standard essentially defines a standard format for labelled messages that agents may use to communicate with one-another. The standard defines 22 distinct locutions, and these have been provided with an operational semantics using speech act theory [20]. The semantics of the language is defined using pre- and post-condition rules, where these conditions define the mental state of participants of communication — their beliefs and intentions. This semantics links utterances in the dialogue to the mental states of the participants, both preceding utterance of each locution, and subsequent to it. All the locutions in the FIPA ACL are ultimately defined in terms of *inform* and *request* primitives.

The FIPA ACL standard is a generic agent communication protocol, and is not based on a dialogue game. However, the various dialogue-game protocols have all been proposed for agent interactions — negotiation, persuasion, etc — for which the FIPA ACL could potentially also be used. It is therefore of interest to see how the FIPA ACL compares to these protocols when assessed against the desiderata above.

1. **Stated Dialogue Purpose:** The ACL is intended primarily for purchase negotiations, and use of the locution *cfp* — standing for *Call For Proposal* — can initiate a negotiation dialogue with a stated purpose. There does not appear to be means to state the purpose of other types of dialogue, e.g., information-seeking or persuasion dialogues.
2. **Diversity of Individual Purposes:** This is enabled.
3. **Inclusiveness:** There do not appear to be limitations on which agents may participate.
4. **Transparency:** The ACL rules are transparent, and the definitions present the pre-conditions and post-conditions of each locution.
5. **Fairness:** The rules treat all participants equally.
6. **Clarity of Argumentation Theory:** There is no explicit underlying argumentation theory for the FIPA ACL. Implicitly, the argumentation model is an impoverished one. Participating agents, for example, have only limited means to question or contest information given to them by others, i.e., via the *not-understood* locution. Moreover, the rules provide speakers uttering such challenges with no rights to expect a defence of prior assertions by those who uttered them.
7. **Separation of Syntax and Semantics:** The syntax is defined in terms of the semantics, so the two are not separated. For example, the pre-conditions of the *inform* locution include a sincerity condition: the speaker must believe the argument of this locution to be true before uttering the locution.

8. **Rule-Consistency:** The rules appear to be consistent.

9. **Encouragement of Resolution:** The FIPA ACL does not appear to discourage resolution, but no rules for dialogue termination are provided.

10. **Discouragement of Disruption:** There are no rules which explicitly preclude disruptive behaviour, although the rationality conditions incorporated in the semantics may limit such behaviour.

11. **Enablement of Self-Transformation:** Self-transformation is limited. The semantics imposes sincerity conditions on utterances, and so agents may only assert what they sincerely believe to be true. However, there are no locutions for retraction of prior assertions, and no means to express degrees of belief or to qualify assertions.

12. **System Simplicity:** The locutions include both substantive locutions, e.g., *accept-proposal*, and procedural locutions, e.g., *propagate*, which asks the recipient to forward the message contents to others. The language would be simpler if these were treated as different classes of locution.

13. **Computational Simplicity:** The FIPA ACL essentially just defines a standard format for messages, and so it is hard to assess in general the complexity of its use. However, because agents must check the sincerity condition of *inform* locutions before uttering these, they require an internal proof mechanism; this will be in the first-order modal logic of the FIPA ACL semantics [20], which is at best semi-decidable.

The implicit model of joint decision-making underlying the FIPA ACL is a rational choice one. As explained in Section 2, this model has no place for self-transformation in the course of the interaction, and hence gives no value to argumentation activities such as information-seeking, persuasion, inquiry or joint deliberation. The rational-choice model may be appropriate for agent purchase negotiation dialogues, although marketing theorists would argue differently, since their models of consumer behaviour typically assume that consumer preferences may only be finalized *during* the purchase decision process and not before, e.g., [21]. However, the rational-choice model, as we argued earlier, is not at all appropriate for other forms of agent dialogue, such as joint determination of plans of action or joint inquiries after truth.

5. DISCUSSION

Designer's Checklist

The experience of developing the list of desiderata presented above and applying them to the three dialogue game protocols have led us to formulate a list of guidelines for designers of such protocols, as follows:

G1 The protocol should embody a formal and explicit theory of argument.

G2 The rules for the protocol should ensure that the reason(s) for conducting the dialogue are stated within the dialogue at its commencement.

G3 The protocol should include locutions which enable participants to:

- **G3.1** formally enter a dialogue
- **G3.2** request information
- **G3.3** provide information
- **G3.4** request arguments and reasons for assertions
- **G3.5** provide arguments and reasons for assertions
- **G3.6** challenge statement and arguments
- **G3.7** defend statement and arguments
- **G3.8** retract previous assertions
- **G3.9** make tentative proposals
- **G3.10** express degrees of belief in statements
- **G3.11** express degrees of acceptability or preferences regarding proposals
- **G3.12** formally withdraw from a dialogue.

G4 The protocol syntax should be defined in observable terms, so this its conformance can be verified without reference to internal states or mechanisms of the participants.

G5 The rules of the protocol should seek to preclude disruptive behavior.

G6 The rules of the protocol should indicate circumstances under which a dialogue terminates.

G7 The rules of the protocol should identify any difference in formal roles and the rights and duties pertaining to these.

These guidelines may be viewed as a checklist for designers and users of agent interaction protocols involving argumentation. To our knowledge they are the first such design guidelines proposed for agent dialogue game protocols.

Conclusions

In this paper we have presented the first list of criteria by which to assess a dialogue-game protocol for agent interactions. These thirteen desiderata were developed after consideration of: economic and computational criteria recently proposed for the assessment of automated auction and negotiation mechanisms; theories of social and public decision-making in argumentation theory and political philosophy; and recent research in designing and studying agent communications languages and interaction protocols.

We have applied these thirteen desiderata to three dialogue game protocols from the literature, for agent dialogues involving negotiation, persuasion and mutual inquiry, respectively. Interestingly, each protocol was found to be weak in at least one important aspect. The negotiation protocol did not permit agents to express changes in their beliefs or preferences in the course of the dialogue (what political theorists call *self-transformation*), while the persuasion dialogue implicitly drew on a theory of argumentation we believe to be inappropriate for the agent domain. The inquiry dialogue was the only one to make explicit the argumentation theories upon which it is based, but it permitted discussion to range over any and all topics simultaneously, thereby limiting its practical usefulness.

We also considered the Agent Communications Language (ACL) of FIPA, for comparison purposes. The key weaknesses of FIPA ACL, relative to these desiderata, were found to be its limited support for formal argumentation and for self-transformation by participants. These findings were not surprising, since its designers did not seek to embody a theory of argumentation, and because it appears to express a rational-choice (or marketplace) view of agent society, rather than a deliberative democracy view. In our opinion, these weaknesses preclude the use of FIPA ACL beyond the purchase negotiation dialogues it was designed for.

In future work, we aim to formalize these desiderata and thus be in a position to prove formally the properties of dialogue game protocols, such as those assessed in Section 3.

6. REFERENCES

[1] R. Alexy. A theory of practical discourse. In S. Benhabib and F. Dallmayr, editors, *The Communicative Ethics Controversy*, pages 151–190. MIT Press, Cambridge, MA, USA, 1990. (Originally published in German in 1978.).

[2] L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In E. Durfee, editor, *Proc. 4th ICMAS*, pages 31–38, Boston, 2000. IEEE.

[3] L. Amgoud, S. Parsons, and N. Maudet. Arguments, dialogue, and negotiation. In W. Horn, editor, *Proc. 14th European Conf. on AI (ECAI 2000)*, pages 338–342, Berlin, 2000. IOS Press.

[4] T. J. M. Bench-Capon, T. Geldard, and P. H. Leng. A method for the computational modelling of dialectical argument with dialogue games. *AI and Law*, 8:233–254, 2000.

[5] J. A. Blair. The limits of the dialogue model of argument. *Argumentation*, 12:325–339, 1998.

[6] J. Bohman and W. Rehg, editors. *Deliberative Democracy: Essays on Reason and Politics*. MIT Press, Cambridge, MA, USA, 1997.

[7] F. Dignum, B. Dunin-Kęplicz, and R. Verbrugge. Agent theory for team formation by dialogue. In C. Castelfranchi and Y. Lespérance, editors, *Proc. 7th ATAL Workshop (ATAL-2000)*, pages 141–156, Boston, 2000.

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

[9] F. H. van Eemeren and R. Grootendorst. *Argumentation, Communication and Fallacies: A Pragma-Dialectical Perspective*. LEA, Mahwah, NJ, USA, 1992.

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

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

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

[13] F. Guerin and J. Pitt. Denotational semantics for agent communication languages. In J. P. Müller, E. Andre, S. Sen, and C. Frasson, editors, *Proc. 5th Intern. Conf. Autonomous Agents*, pages 497–504, New York, 2001. ACM Press.

[14] C. L. Hamblin. *Fallacies*. Methuen, London, 1970.

[15] D. Hitchcock. Some principles of rational mutual inquiry. In F. van Eemeren *et al.*, editor, *Proc. 2nd Intern. Conf. Argumentation*, pages 236–243, Amsterdam, 1991. SICSAT.

[16] D. Hitchcock, P. McBurney, and S. Parsons. A framework for deliberation dialogues. In H. V. Hansen *et al.*, editor, *Proc. 4th Conf. Ontario Soc. Study of Argumentation (OSSA-2001)*, Windsor, Canada, 2001.

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

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

[19] S. Kraus. *Strategic Negotiation in Multiagent Environments*. The MIT Press, 2001.

[20] Y. Labrou, T. Finin, and Y. Peng. Agent communication languages: The current landscape. *IEEE Intelligent Systems*, 14(2):45–52, 1999.

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

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

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

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

[25] P. McBurney and S. Parsons. Games that agents play: A formal framework for dialogues between autonomous agents. *J. Logic, Language and Information*, 11(3), 2002. *In press*.

[26] P. McBurney and S. Parsons. A geometric semantics for dialogue-game protocols for autonomous agent interactions. *Electronic Notes in Theoretical Computer Science*, 52(2), 2002. *In press*.

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

[28] W. Rehg. The argumentation theorist in deliberative democracy. *Controversia*, 2002 (to appear).

[29] J. S. Rosenschein and G. Zlotkin. *Rules of Encounter: Designing Conventions for Automated Negotiation among Computers*. The MIT Press, 1994.

[30] F. Sadri, F. Toni, and P. Torroni. Logic agents, dialogues and negotiation: an abductive approach. In M. Schroeder and K. Stathis, editors, *Proc. Sym. Info Agents for E-Commerce, AISB-2001*, York, UK, 2001. AISB.

[31] 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.

[32] 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*, LNAI 1365, pages 177–192, Berlin, 1998. Springer.

[33] M. P. Singh. A social semantics for agent communications languages. In F. Dignum, B. Chaib-draa, and H. Weigand, editors, *Proc. IJCAI-99 Workshop on Agent Communication Languages*, Berlin, 2000. Springer.

[34] S. E. Toulmin. *The Uses of Argument*. Cambridge University Press, Cambridge, UK, 1958.

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

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