Retraction and revocation in agent deliberation dialogs

Peter McBurney
Department of Computer Science
University of Liverpool
Liverpool L69 3BX
UK
p.j.mcburney@csc.liv.ac.uk

and

Simon Parsons
Department of Computer & Information Science
Brooklyn College
City University of New York
Brooklyn NY 11210
USA
parsons@sci.brooklyn.cuny.edu

Abstract
We present a generic denotational semantic framework for protocols for dialogs between rational and autonomous agents over action which allows for retraction and revocation of proposals for action. The semantic framework views participants in a deliberation dialog as jointly and incrementally manipulating the contents of shared spaces of action-intention tokens. The framework extends prior work by decoupling the identity of an agent who first articulates a proposal for action from the identity of any agent then empowered to retract or revoke the proposal, thereby permitting proposals, entreaties, commands, promises, etc, to be distinguished semantically.

KEY WORDS: Agent Communications, Deliberation Dialogs, Dialog Games, Interaction Protocols.

1. Introduction

The rise of distributed computing, as exemplified by the growth of the Internet and the WorldWideWeb, has created many research and engineering challenges for computer scientists. A key challenge has been – and remains – the design of artificial languages and protocols by which different computers, and computational entities, may communicate with one another. This area, known within computer science as *agent communications*, draws on linguistic theory, the philosophy of language, and argumentation theory, in addition to methods from artificial intelligence and software engineering. A key influence has been the classification of human dialog types presented by Erik Krabbe and Doug Walton in Walton and Krabbe (1995). By considering dialogs in terms of the possibly-different beliefs of the participants at the outset of the dialog, and the possibly-different goals they seek to achieve from

participation in the dialog, Walton and Krabbe were able both to distinguish several different types of dialog from one another, and also to provide a means by which additional dialog types may be identified.

Although its dialog classification has been influential, the book by Walton and Krabbe (1995) was primarily concerned with understanding commitments made in dialogs, both their statement and their retraction. This was a theme taken up at greater depth by Erik Krabbe in Krabbe (2001), where he discusses the subtleties involved for a dialog system designer in deciding how permissive to be in allowing (or not allowing) retractions of commitments. If retractions are unconstrained, then malevolent or whimsical or bug-ridden participants may wreak havoc on a dialog, or delay resolution to the interaction. If, on the other hand, retractions are not permitted at all, rational participants may see no value in engaging in dialog with one another using the protocol, since there may be no possibility of other participants being able to admit to a change of belief or intention.

A designer of a dialog system may, of course, leave the decision as to the possibility of retractions of utterances to the participants themselves. One way to enable this would be to allow for *meta-dialogs*, dialogs about dialogs, in which the participants discuss with each other what rules are appropriate for retraction and revocation of utterances in the ground dialogs. In McBurney and Parsons (2002), we presented a generic framework enabling participants to combine and invoke such meta-dialogs from within, or alongside, a ground dialog. Thus, given such a dialectical system for such meta-dialogs, it should be a straightforward matter to combine it with systems for ground dialogs in a coherent manner. However, designing a dialectical system to allow such exchanges between software agents would be a challenging undertaking at the present time, since it would require an understanding of the reasons agents may have for seeking or not seeking particular rules for revocation and nullification of different types of utterances, and how these reasons may relate to one another. Considerably more research on the dialectical consequences of particular revocation rules would be needed before such understanding would be possible.

An alternative approach, which we adopt in this paper, is to consider the issue of retraction in terms of the identity of the participant able to retract (or revoke or cancel) a prior commitment. Our attention only concerns *deliberation dialogs* (Walton and Krabbe 1995), those dialogs in which participants seek to reach a decision on what action or actions to take in some situation. Our focus throughout is on rational and autonomous participants, and so any decisions they make collectively in the course of dialog will only be reached through rational persuasion and argument. Argumentation provides the means by which participants evaluate proposals for action made by others, and the means by which they persuade others to adopt their own proposals. In these dialogs, agents may make many different types of utterances. In McBurney and Parsons (2005b), we articulated a comprehensive classification of agent speech acts in rational interactions, extending earlier classifications of Austin (1962), Searle (1969) and Habermas (1984):

- Factual statements, asserting some proposition as true in the world.

- Proposals to undertake some action.

- Expressions of preferences between two or more proposed actions.
- Promises by the speaker to the hearer to undertake some action.
- Requests or entreaties by the speaker urging the hearer to undertake some action.
- Commands by the speaker to the hearer to undertake some action.
- Arguments for or against proposals, promises, requests, and commands.
- Acceptances or rejections of proposals, promises, requests or commands.
- Retractions or revocations of previously uttered proposals, promises, requests or commands.
- Control statements, in which participants may enter the dialog, ask for statements to be repeated, withdraw from the dialog, etc.

With the exceptions of factual statements, of arguments and of control statements, the other types of statements listed here express some intention either to act or not to act, or to constrain actions in some way, for example, via statements of preference. The actor – the agent doing the action, were it to be executed – for these various types of statement may be different. Thus the action specified in a promise, if executed, is undertaken by the speaker, while that specified in a command is undertaken by the hearer. Similarly, who is empowered to retract or revoke an action-statement may differ by the type of statement. Commands, for example, if given lawfully, can usually only be revoked by the agent who issued the command, not by the executor of the action who heard it. In contrast, the executor of a promise, once it is accepted, may usually only be released from performing the stated action by its hearer, the recipient of the promise, not the agent who first uttered it.

One interpretation of these differences is summarized in Table 1. Here, a statement is uttered by agent A regarding action α, an action which may be executed either by agent A or by agent B; prior to execution, the statement may potentially be revoked or retracted by one or either agent. The table indicates which agent has the power to revoke the intention, at a time point after the action has been accepted by agent B, but before the action has been executed.

	A does α	**B does α**
A can revoke	A offers to do α	A commands B to do α
B can revoke	A promises B to do α	A entreats B to do α
A or B can revoke	A proposes that A do α	A requests or A proposes that B do α

Table 1: Types of Action Statement

Table 1 presents one consistent interpretation of these speech acts but other interpretations are possible. For example, after *entreating* B to do α, and having heard B agree to do this, A may usually only revoke the entreaty with a consequent loss of reputation; whether this loss is important or not depends on the social context in which the two agents are undertaking their dialog. However, whatever interpretations of the specific verbs are adopted, Table 1 demonstrates that part of the meaning – the semantics and pragmatics – of speech acts about actions concerns the agreed circumstances regarding their issuance and revocation. Who has permission to make an utterance and who the power to revoke or withdraw it differs greatly by the type of utterance. The specific rules for particular speech acts will differ from one culture to another, according to social conventions and norms. They may also differ, within any culture, from one type of dialog to another, or even, within any dialog, from one dialectical context to another.

Given this variety, the design of dialectical frameworks for meta-dialogs about the rules of revocation is, as we noted earlier, a major undertaking. Instead, we could seek to embed meta-dialogical discussion about powers of issuance and revocation of speech acts in the structure of the speech acts themselves, and this is the approach adopted here. In doing so, our work differs from the two main approaches to the semantics of agent communications languages, the semantics based on internal mental states of, for example, the Agent Communications Language FIPA ACL of the Foundation for Intelligent Physical Agents (FIPA)[1], and the social semantics of Singh (1999) and Colombetti and colleagues (Colombetti and Verdicchio 2002). Although the FIPA ACL includes many speech acts for actions – indeed, 10 of the 22 FIPA ACL locutions relate to actions – the semantics of FIPA ACL ignores these issues of revocation (see FIPA 2002). In contrast, social semantics overcomes this by treating utterances in agent dialogs as attempts at manipulation of the social relationships between the participants, but this seems too high a level of abstraction. Since it is not specific to deliberation dialogs, social semantics does not allow one to readily formulate rules of interaction in specific protocols for deliberation.

How are we to understand agent dialogs over action? In particular, how are we to represent speech acts in deliberation dialogs in a manner which can distinguish between promises and commands, etc. This paper is concerned with the questions: *How should we think about agent dialogs over action? Can we conceive of these dialogs in a way which enables a unified treatment of different types of speech acts? Can we also do so in a way which facilitates implementation of systems for agent dialogs over action?* The contribution of this paper is to answer these questions positively, by presenting a novel semantic framework for dialogs over action which incorporates different types of speech acts, and which is readily implementable using recent multi-agent technologies. Our focus is only on machine-to-machine dialogs, and not human-to-human or human-to-machine dialogs.

The paper is structured as follows. In Section 2, we articulate several key principles which motivate and guide the work of the paper. Following these, Section 3 presents the syntax of the protocols we consider, by specifying their legal locutions and the rules governing the combination of these locutions. Section 4 then presents our denotational semantics for these protocols, in the form of a trace semantics extending that of McBurney and Parsons (2005a). The paper ends, in Section 5, with a discussion of related and future work.

2. Guiding Principles

We begin by articulating certain principles of our conception of agent dialogs over action, and their semantics. These principles will guide the development of the semantic framework we propose in the paper. Firstly, the agent interaction is assumed to be an open one, with agent participation being voluntary and willing, and with any agreements being reached without external coercion. This assumption is in accord with other rules of rational dialog, such as Hitchcock's *Principles of Rational Mutual Inquiry* (Hitchcock 1991). Furthermore, the assumption still permits agents to legally issue instructions to other agents, provided the prior social relationship existing between the respective agents, such as an employment contract, was entered into without coercion.

Secondly, we assume that participants in dialogs over action themselves assume that the other participants enter the dialog with the intention of seeking joint agreement to undertake, or not undertake, actions. Participants assume that their fellow-participants are not engaged in whimsy, or malice, or in an insincere simulation of deliberation. This assumption means that utterances in dialogs over action are understood by their hearers as statements intended to change the world, or to constrain its change, in some way. Thus, these utterances are not primarily about communication of information or beliefs, although that may be an incidental consequence of their utterance. Rather, they are understood by their audience as statements of intentionality – of preferences, desires, proposals for action, and/or of intentions – by the agent who utters them.

Thirdly, we assume that the semantics of a dialog over action is something constructed jointly and incrementally by the participants in the course of the dialog. Articulating and constraining the possible actions is the purpose of a deliberation dialog, and this is something achieved in and through the dialog by the participants themselves. A shared space of possible actions is not something which exists before the dialog commences, and it is not something static.[2] Of course, the agents in a dialog may have separate, pre-existing spaces of possible actions, and these individual spaces may remain static throughout a dialog.

Putting these principles together, we are led to treat a statement about action in a deliberation dialog as an attempt by the speaker to manipulate a shared space of tokens, where the tokens represent not merely actions, but agent intentionality concerning actions.[3] Depending on the nature of the token and the identity of the speaker, a speaker may or may not be able to manipulate it, as for example, in having permission to revoke a command. Note that, as explained above, our notion of "statements of intentionality" refers to preferences, desires and proposals, in addition to requests, promises, commands, etc. Thus, these "action-intention tokens" may not necessarily indicate firm commitments by some agent to undertake some action, or may only do so following the occurrence of appropriate dialogical events, such as acceptance by another agent. In the sections which follow, we develop the syntax and a semantics for deliberation dialogs viewed in this way.

The semantics we present translates utterances in an agent deliberation dialog into mathematical entities, specifically objects and arrows in certain categories. It is

therefore an example of a *denotational semantics* in the abstract theory of computer programming languages (Gunter 1992). This theory distinguishes several types of semantics for programme languages: *axiomatic semantics* provide the pre- and post-conditions of well-formed syntactical statements in the language; *operational semantics* treat statements as commands altering the overall state of a virtual computer, and articulate the state-transition functions for each statement; and *denotational semantics* translate statements into mathematical entities in order that the properties of the language or of programs written in the language may be studied through mathematical reasoning over these entities. To date, most agent communications protocols have been given an axiomatic semantics, as in Amgoud *et al.* (2000), and Bench-Capon *et al.* (2000); similarly, the FIPA ACL has been given an axiomatic semantics defined in terms of the mental states (beliefs, desires and intentions) of the speakers and hearers of utterances, using a modal logical formalism (FIPA 2002). Some protocols have also been given an operational semantics (McBurney *et al.* 2003) or a denotational semantics (McBurney and Parsons 2005a).

This paper is part of a long-term research effort by the authors to develop an appropriate denotational semantics for agent dialog protocols, in order to have a sound basis for comparison of different protocols (Johnson *et al.* 2003) and for exploration of their properties (McBurney and Parsons 2005a). A key motivation for this research effort is the desire to ensure that different software agents, possibly created by different (human or agent) design teams, which are using a particular protocol share the same understanding of the protocol, and of dialogs undertaken using it. While their beliefs and their immediate goals may be different, effective dialog between multiple agents requires at least a shared understanding of the protocol and the utterances within it.

3. Protocol Syntax

In prior work in agent deliberation dialogs, we presented a denotational semantics, called a *trace semantics*, for two specific classes of deliberation dialogs (McBurney and Parsons 2005a). Protocols in these classes allowed agents to make proposals for action, to express preferences between two proposals, and to accept or reject proposals. This earlier work implicitly assumed that only agents who uttered a proposal for action were empowered to revoke it, and thus could not represent all the types of speech acts indicated in Table 1. Here, we build on this earlier work in order to represent deliberation dialogs in which the identities of the agents uttering speech acts concerning action and the identities of those uttering revocations may be de-coupled. For reasons of space, we do not present syntax and semantics of all the types of locution listed earlier in Section 1, but only sufficient of these to illustrate our approach. For the same reason, we also assume that all actions considered in a dialog are to be executed only by the participants, and not by anyone outside the dialog. We use the same formalism as in (McBurney and Parsons 2005a), which is summarized here. The locutions PROPOSE, ACCEPT and PREFER are adapted from that earlier work.

3.1 Speech Acts

We assume that time is continuous, and isomorphic to the positive real numbers, but that utterances occur only at integer values, with precisely one utterance made at each integer time-point. We further assume that these protocols are specified as dialog games, in accordance with current research in agent communications protocols, e.g., McBurney *et al.* (2003). In this approach, the syntax of legal utterances comprises two layers, with the lower, content layer being wrapped in a higher, speech-act locution. Generic (uninstantiated) speech-acts are denoted with just the wrapper as, for example, in WITHDRAW(.), while instantiated locutions are denoted with both wrapper and contents shown, as in WITHDRAW(t, Pi). We denote participating agents by Pi, for i a positive integer in some finite set I indexing the set of all participating agents A = {Pi | i ε I }. The contents of locutions are denoted by lower-case Greek letters, and L = { α, β, . . . } denotes this collection of locution contents; each element of L represents an action or plan of action to be undertaken following agreement by the dialog participants. Although not strictly necessary, for ease of presentation, we assume the first field in the content of utterances is the integer time t of the utterance, and the second field in the content is an identifier Pi of the agent uttering the locution.

General Locutions

We assume the protocol contains control locutions for participants to initiate, enter and withdraw from the protocol, such as those defined in other recent dialog game protocols, e.g., McBurney *et al*. (2003). We assume the syntax of the withdrawal illocution is WITHDRAW(t, Pi).

Specific Locutions

The protocol contains locutions of the following form:

[L1] PROPOSE(t, Pi, α, E, Pj, Π), which enables the speaker, agent Pi, to propose the action α be undertaken by agent Pj (possibly Pi itself) upon achievement of state E (which may be null), with the utterance being revokable by any of the agents listed in finite set Π, a subset of the set of agents {Pk | k ε I }. Variable E is a proposition, or well-formed propositional formula, describing some state of the world. We allow state E to indicate a clock-tick: as in: *"The variable Time has value u, for some specified u > t."* We further assume that utterance of PROPOSE(t, Pi, α, E, Pj, Π) by a speaker expresses a willingness of the speaker Pi itself to accept the proposal α at the time t of utterance. Once accepted, the proposal can only be revoked by an agent included in the set Π.

[L2] PREFER(t, Pi, α, E, β, F), which indicates to any hearers that the speaker, agent Pi, prefers proposed action β, undertaken upon achievement of state F, to proposed action α, undertaken upon achievement of state E, at time t.

[L3] ACCEPT(t, Pi, α, E, LOC), which indicates to the hearer that the speaker, agent Pi, wishes to indicate agreement to the action α being undertaken upon achievement of state E, which has been the subject of the prior utterance LOC which must be of the form: PROPOSE(s, Pk, α, E, Pj, Π), for s < t, and for some values of k, j and some set of agents Π.

[L4] REJECT(t, Pi, α, E, LOC), which indicates to the hearer that the speaker, agent Pi, wishes to indicate disagreement to the action α being undertaken upon achievement of state E, which has been the subject of the prior utterance LOC which must be of the form PROPOSE(s, Pk, α, E, Pj, Π), for s < t, and for some values of k and j and some set of agents Π.

[L5] REVOKE(t, Pi, α, E, LOC), which indicates to the hearer that the speaker, agent Pi, wishes to revoke or cancel the prior utterance LOC of the form PROPOSE(s, Pk, α, E, Pj, Π), for s < t, and for some values of k and j and some set of agents Π containing Pi.

The generic form of the PROPOSE(.) locution allows different types of speech acts to be represented, depending on how this locution is instantiated. Some of these different types are shown in Table 2, in which agent Pi is the speaker of the locution in every case.

Instantation of locution PROPOSE(.)	*Speech Act*	*Who acts*	*Who revokes*
(t, Pi, α, E, Pi, { Pi, Pj } }	Propose	Pi	Pi or Pj
(t, Pi, α, E, Pi, { Pj } }	Promise	Pi	Pj
(t, Pi, α, E, Pj, { Pj } }	Entreat	Pj	Pj
(t, Pi, α, E, Pj, { Pi } }	Command	Pj	Pi

Table 2: Locution Types for Instantiation of *PROPOSE(.)*

3.2 Combination and Termination Rules

The locutions listed above are subject to the following combination rules (C1—C5) and a termination rule (C6). For simplicity, we assume that any deliberation dialog concerns only one issue for which only one action (or one course of action), need be agreed. Once such agreement is reached, the dialog terminates.

[C1] The instantiated locution

ACCEPT(t, Pi, α, E, PROPOSE(s, Pj, α, E, Pk, Π))

may only be uttered legally if there has been a prior utterance of

PROPOSE(s, Pj, α, E, Pk, Π),

by some agent Pj at some time s < t.

[C2] The instantiated locution

PREFER(t, Pi, α, E, β, F)

may only be legally uttered if there have been prior instantiated utterances of

PROPOSE(s, Pj, α, E, Pk, Π)

and

PROPOSE(r, Pl, β, F, Pm, Σ)

in which actions α and β have each appeared, for r ≠ s. These utterances do not need to have been made by agent Pi.

[C3] The instantiated locution

REVOKE(t, Pk, α, E, PROPOSE(s, Pi, α, E, Pj, Π))

may only be uttered legally if there has been a prior utterance of

PROPOSE(s, Pi, α, E, Pj, Π)

by some agent Pi at time s < t, and provided the set of agents Π contains Pk.

[C4] Expressed participant preferences are transitive, i.e., utterance of the following two instantiated locutions at any times t and t+k in a dialog

PREFER(t, Pi, α, E, β, F)

and

PREFER(t+k, Pi, β, F, γ, G)

entitles a hearer to infer the relationship represented by the following speech act:

PREFER(t+k, Pi, α, E, γ, G).

[C5] Participant preferences are reflexive, i.e., for every action α and pre-condition E, every speaker Pi is able to utter:
PREFER(t, Pi, α, E, α, E).

[C6] The protocol has a voting rule indicating when an agreement is reached on an action, and this results in the termination of the dialog and execution of the action. For example, for unanimous agreement, the rule could be as follows:

If there is a proposal α such that all participants Pi have uttered either

PROPOSE(t, Pi, α, E, Pj, Π) or

ACCEPT(t, Pi, α, E, PROPOSE(t, Pk, α, E, Pj, Π)),

then the dialog ends immediately, with the participants agreeing to execute the action or action plan represented by α upon achievement of state E.

In the remainder of this paper, we will assume that unanimous agreement is required for an action to be agreed in the dialog. Rules C4 and C5 are required for the resulting mathematical structure to be a category. Note that we do not assume that every participant is always able to express a preference between any two proposals. At any given time, a participant in a dialog may prefer one proposal to a second, or may prefer the second to the first, or may be indifferent between the two proposals, or the participant may not yet have determined its preference between the two proposals.

3.3 Protocol Class DA

Definition 1: Class DA: Dialog-over-Action Protocols:
An agent interaction protocol is a member of the class of Dialog-over-Action Protocols (denoted *DA*) if it permits the general and specific speech acts L1—L5 and these are subject to the combination rules C1—C5 and the termination rule C6.

4. Protocol Semantics

We now define a denotational *trace semantics*, as in McBurney and Parsons (2005a), for deliberation dialogs conducted under protocols in *DA*, using concepts from Category Theory (Mac Lane 1998).[4] Assume G is a deliberation protocol in class *DA*. Let A = { P1, . . . , Pn } be a finite set of n distinct agents, engaged in a deliberation dialog conducted in accordance with protocol G, with the set L = { α, β, . . . } being the topics of the dialogs (i.e., the substantive contents of locutions) and each representing an action or plan of action. We let g1, g2, . . . denote dialogs – sequences of instantiated locutions – conducted by agents in A under protocol G. We denote the agent index set {1, . . . , n } by I.

We now assume the existence of the following sequences of mathematical categories:

- For each agent Pi we assume there exist n^2 time-indexed sequences of categories, each category denoted $C^t_{i, j, k}$, for time t a non-negative integer and j, k elements of I. For each agent Pj (including Pi) and for each time t, the category, $C^t_{i, j, k,}$ contains objects corresponding to the utterances made by agent Pi up to and including time t in the dialog, concerning actions to be executed by agent Pj, and such that the utterance may be revoked by agent Pk. These categories are called the *public proposal stores* of agent Pi.

- We next form the time-indexed sequence of categories C^t, with each category formed from the union of the objects and the identity arrows of the n^3 categories $C^t_{i, j, k}$, for i, j, and k elements of I, and time t a non-negative integer. We call each of these categories the *shared proposal space at time t*, and the collection of all of them, the *shared proposal space*.

- Finally, for each agent Pi in A we assume there exists a time-indexed sequence of categories, denoted M^t_i, with t a non-negative real number. Each of these categories is called the *private proposal store* of agent Pi at time t. Agent Pi is assumed to commence the deliberation dialog with private proposal store M^0_i, which may be empty. This store contains tokens for possible actions which agent Pi is considering at time t (for execution by itself or by other agents), but may not yet have been revealed to the dialog. The presence in these private stores of objects representing possible actions does not indicate any commitment on the part of the respective agents whose private stores they are to the actions. Indeed, as their name implies, the contents of a private proposal store are only observable by the agent with which the store is associated.

The objects and arrows in these categories are action-intention tokens, with the objects and arrows inserted and deleted as a result of utterances in the dialog. Although defined in terms of sequences of categories for each agent Pi, we may think of the shared proposal space at time t, C^t, as being partitioned into the separate categories $C^t_{i, j, k.}$

These categories are constructed by the following trace-semantics rules, linking dialog statements to objects and arrows in the appropriate categories. In all categories, we label those objects corresponding to possible actions with lower-case Greek letters, while certain other objects have mnemonic labels; arrows are labelled with

lower-case Roman letters, corresponding to the agent whose preferences they represent. An object labelled θ^E may be understood as the action (or course of action) θ to be agreed and executed upon achievement of state E (which may, as before, be a clock-tick or null). Condition-stamping of objects in this way allows us to model an agent's preferences with respect to the same action to be undertaken at different times or with different pre-conditions. Arrows are used to indicate preferences, with the arrow pointing from the less-preferred object towards the more-preferred object. We first list the rules for the public stores:

[TS1:] Each agent Pi begins the dialog with public proposal stores $C^0_{i, j, k}$ which are empty.

[TS2:] An utterance of PROPOSE(t, Pi, α, E, Pj, Π) by an agent Pi at integer time t results in an object labelled α^E, corresponding to the execution of α upon achievement of state E, being inserted into the public proposal store $C^t_{i, j, k}$ of Pi, for each k such that Pk is an element of Π.

[TS3:] An utterance of the locution

ACCEPT(t, Pj, α, E, PROPOSE(s, Pi, α, E, Pk, Π))

by an agent Pj at integer time t results in an object labelled α^E, corresponding to the execution of α upon achievement of state E, being inserted in the public proposal store $C^t_{j, k, l}$ of Pj, for each l such that Pl is an element of Π.

[TS4:] For each agent Pi and for all times $t \geq 0$, every object θ^E in the public proposal store $C^t_{i, j, k}$ of Pi (and therefore in the shared proposal space C^t) has associated to it an identity arrow $id(\theta^E)$: $\theta^E \rightarrow \theta^E$. This identity arrow is in both categories $C^t_{i, j, k}$ and C^t. This rule encodes Combination Rule C5.

[TS5:] An utterance of the locution PREFER(t, Pi, α, E, β, F) by an agent Pi at integer time t results in an arrow, from the object corresponding to α to the object corresponding to β, and with the arrow labelled by Pi, being inserted into the shared proposal space at time t, C^t. A subsequent utterance of the locution PREFER(u, Pi, β, F, α, E) by the same agent Pi at time $u > t$ deletes the arrow, from the object corresponding to α to the object corresponding to β, in C^u inserted by the utterance of Pi and inserts in C^u an arrow, from the object corresponding to β to the object corresponding to α, again with the arrow labelled by Pi.[5]

[TS6:] An utterance of PREFER(s, Pi, α, E, β, F) by an agent Pi at integer time s following at a later integer time t by an utterance of the locution PREFER(t, Pi, β, F, γ, G) results in an arrow from the object corresponding to α to the object corresponding to γ being inserted into the shared proposal space at time t, C^t. If at a later time $u > t$, the same agent Pi utters the locution PREFER(u, Pi, γ, G, β, F), both the arrow from the object corresponding to β to the object corresponding to γ and the arrow from object corresponding to α to the object corresponding to γ are deleted from the shared proposal space at time u, C^u. This rule encodes Combination Rule C4.

[TS7:] An object α inserted at time s in a public proposal store remains in the store for times $t \geq s$, unless and until an agent Pk, an element of Π, power to revoke the utterance which created the object utters the locution

REVOKE(t, Pk, α, E, PROPOSAL(s, Pi, α, E, Pj, Π)).

Provided this utterance complies with Combination Rule C3, then the utterance results in the object α being deleted from every private proposal store $C^t_{i,j,l}$, such that Pl is an element of Π.

[TS8:] An arrow a from object α to object β inserted in the shared proposal space at time s and labelled by agent name Pi, remains in the space for all times $t \geq s$ unless and until either (a) an arrow b from object β to object α is inserted through a subsequent utterance by Pi, or (b) one of the objects α or β is deleted. The presence of an arrow a: $\alpha \rightarrow \beta$ between two distinct objects α and β and labelled by Pi in the shared proposal space at time t means there is no arrow b: $\beta \rightarrow \alpha$ with the same label in that space.

We now list the rules for the private stores:

[TS9:] Each agent Pi begins the dialog with a private proposal store M^0_i (which may be empty).

[TS10:] An utterance of PROPOSE(t, Pi, α, E, Pj, Π) by an agent Pi at integer time t means that there exists $\varepsilon > 0$ such that an object corresponding to α^E is in the private proposal store $M^{t-\varepsilon}_i$ of Pi at time t-ε.

[TS11:] An utterance of PROPOSE(t, Pi, α, E, Pj, Π) by an agent Pi at integer time t results in an object corresponding to α^E being inserted in the private proposal store M^t_l of agent Pl, for every $l \neq i$.

[TS12:] For each agent Pi and each time $t \geq 0$, every object θ^E in the private proposal stores M^t_i of Pi has associated to it an identity arrow $id_{\theta E}$: $\theta^E \rightarrow \theta^E$.

[TS13:] For every agent Pi and every time $t > 0$, the private proposal store M^t_i has a distinguished object, called ND^t_i , intended to represent *"No Action"*.

[TS14:] For every agent Pi and every time $t > 0$, the private proposal store M^t_i has a distinguished object, called FP^t_i, an abbreviation for *"Future Prospects at t"*, intended to represent the valuation at time t by agent Pi of all possible future actions, allowing for the estimation by the agent of any uncertainty in their achievement.[6]

[TS15:] An utterance of the locution PREFER(t, Pi, α, E, β, F) by an agent Pi at integer time t means that there exists $\varepsilon > 0$ such that there is an arrow from the object corresponding to α to the object corresponding to β in the private proposal store $M^{t-\varepsilon}_i$ of Pi at time t-ε.

[TS16:] An utterance of the locution PREFER(s, Pi, α, E, β, F) by an agent Pi at integer time s following at a later integer time t by an utterance of PREFER(t, Pi, β, F, γ, G) means that there exists $\varepsilon > 0$ such that there is an arrow from the object corresponding to α to the object corresponding to γ in the private proposal store $M^{t-\varepsilon}_i$ of Pi at time t-ε.

[TS17:] For every agent Pi and every time $t \geq 0$, whenever there are arrows a: $\alpha \rightarrow \beta$ and b: $\beta \rightarrow \gamma$ in the private proposal stores M^t_i then there is also an arrow c: $\alpha \rightarrow \gamma$ in M^t_i.

[TS18:] The presence of an arrow a: $\alpha \rightarrow \beta$ between two distinct objects α and β and with a given label in a private proposal store means there is no arrow b: $\beta \rightarrow \alpha$ with the same label in that store.

The rules for the private stores (TS9—TS18) create a mathematical model of the private states of the participating agents. It is important to note that agents may not necessarily conform to this model in their actual decision processes when engaged in deliberation dialogs.[7] In any case, such conformance would in general be unverifiable (Wooldridge 2000). Rules TS9, TS12, TS16 and TS17 encode category-theoretic axioms. Rules TS10 and TS15 ensure that agents only propose actions or utter preferences which they have considered (however briefly or incompletely) privately. Rule TS11 ensures that the private proposal stores of agents contain (tokens for) all the publicly-expressed proposals of other agents. Rule TS13 means that agents can compare proposals for action at a particular time with the action of doing nothing at that time. As we showed in McBurney and Parsons (2005a), Rule TS14 allows agents to compare acceptance at the present time of a particular proposal for action with continuation of the dialog in the hope of obtaining a better dialog outcome than that proposal. Rule TS18 encodes the intended meaning of preference, as stated in Syntactic Rule L2. Rule TS17 corresponds to an assumption that the private preferences of each agent are transitive. Note that we make no assumption that an agent's preferences are fixed or pre-determined. Thus, objects may enter and leave the private proposal stores of the participants throughout a dialog, and arrows likewise may change. In other words, there is no assumed relationship between M^s_i and M^t_i, for $s \neq t$. We believe this captures nicely the notion that agents may have resource-constraints on their processing powers, and so they may not consider all action-options at all times throughout an interaction.

Using these rules, we now define a denotational semantics for dialogs conducted under protocols in class DA:

Definition 2: Given a finite set of agents A, a collection of locution contents L, and a deliberation dialog protocol G in class DA, we define the **Deliberation Trace Semantics**, or **Trace Semantics**, of a dialog g undertaken by A about topics in L according to protocol G by the pair:

$$< C, M >$$

where $C = \{ C^t_{i, j, k} \mid i, j, k \in I, t \in Z^+ \cup \{0\} \} \cup \{ C^t \mid t \in Z^+ \cup \{0\} \}$ is a collection of public proposal stores and shared proposal spaces for the agents in the dialog, created according to rules TS1—TS8, and $M = \{ M^t_i \mid i \in I, t \in R^+ \cup \{0\} \}$ is a collection of private proposal stores for each agent in the dialog, created according to Rules TS9—TS18. We also call $< C, M >$ a **deliberation trace** of A, L and G, denoted:

$$< C, M > \models (A, L, G).$$

Given this definition of the denotational semantics, it is easy to show:

Proposition 1: Each element of C and M is a category.

Proof: Straightforward from the definitions of the semantics given above and the definition of a category (Mac Lane 1998), using Rules TS4, TS5 and TS6, in the case of elements of C, and Rules TS12, TS15 and TS16, in the case of elements of M.

Now, as with the denotational semantics presented in McBurney and Parsons (2005a), it is an easy matter to demonstrate the consistency of the trace semantics with respect to deliberation dialogs in DA.

Proposition 2 [Consistency]: For any finite set of agents A, any collection of locutions L and any dialog protocol G in the class DA, there is a trace semantics $< C, M >$ such that $< C, M > \models (A, L, G)$.

Proof: The consistency of the trace semantics follows in a straightforward way from the rules of construction of the semantic framework given above.

We can also demonstrate completeness of the trace semantics with respect to deliberation dialogs in DA. For this, we must confine attention to collections of categories satisfying the properties implied by rules TS1—TS18. We therefore have:

Proposition 3 [Completeness]: Suppose the two collections of categories $< C, M >$, with $C = \{ C^t_{i, j, k} \mid i, j, k \in I, t \in Z^+ \cup \{0\} \} \cup \{ C^t \mid t \in Z^+ \cup \{0\} \}$ and $M = \{ M^t_i \mid i \in I, t \in R^+ \cup \{0\} \}$ have the following properties:

(a) I is finite, with cardinality n.

(b) $C^0_{i, j, k} = \{ \}$, for all $i, j, k \in I$.

(c) $C^0 = \{ \}$.

(d) Each $C^t_{i, j, k}$ is isomorphic to a subcategory of M^t_i, for all $i, j, k \in I$, and for all $t \in Z^+ \cup \{0\}$.

(e) The only arrows in each $C^t_{i, j, k}$ are identity arrows, for all $i, j, k \in I$, and for all $t \in Z^+ \cup \{0\}$.

(f) Each category M^t_i has at most a countable number of objects, for all $i \in I$, and for all $t \in R^+ \cup \{0\}$.

(g) Every object and arrow of $C^t_{i, j, k}$ is also an object and arrow of C^t, for all $i, j, k \in I$, and for all $t \in Z^+ \cup \{0\}$, and C^t has no other objects beside these. (C^t may have other arrows.)

(h) There is at most one arrow between any two distinct objects in each category in the collection M.

(i) There are no more than min{t-2, n} arrows between any two distinct objects in each category $C^t_{i, j, k}$ and between any two distinct objects in each category C^t, for all i, j, k ε I, and for all t ε $Z^+ \cup \{0\}$.

(j) The total combined number of objects and arrows in the union of categories

$$\cup_I C^t_{i, j, k}$$

is at most t, for all i, j, k ε I, and for all t ε $Z^+ \cup \{0\}$.

(k) The total combined number of objects and arrows in each category $C^t_{i, j, k}$ is at most t, for all t ε $Z^+ \cup \{0\}$.

Then, there exists a dialog g undertaken by a finite set of agents A, about a collection of topics L according to a dialog protocol G, an element of the class DA, for which < C, M > is the trace semantics of (A, L, G).

Proof: The proof follows a similar argument to that for Proposition 3 in McBurney and Parsons (2005a), by counting and labeling the first appearances of the objects and arrows of the categories in $\cup_I C^t_{i, j, k}$, for successive integer points of time, and then using these labels to reconstruct a dialog between virtual agents in a finite set A, isomorphic to I, which uses locutions in a set L, instantiated with these labels. It is then possible to show that these utterances conform to a protocol in DA.

Thus all dialogs under all protocols in the class DA conform to the trace semantics.

5. Discussion

In this paper, we have presented a novel semantic framework for multi-agent dialogs over actions. The main contributions of our framework are, firstly, to view statements about actions as manipulating a shared space of action-intention-tokens, and, secondly, to represent formally, through the partitioned structure of this space, permissions to utter and revoke statements about actions. This second contribution means that we can distinguish semantically between different types of utterances about possible action, for example, *proposals, entreaties, promises*, and *commands*. Indeed, our semantic framework allows the agent who first makes an utterance about a possible action to specify not only which agents will execute this action, but which agents have the right to revoke or cancel the utterance. The framework thereby provides considerable flexibility to representing different speech acts concerning actions, and, moreover, provides this flexibility to agents participating in a dialog to decide revocation or retraction rights at run-time, rather than to protocol designers at design-time. In McBurney and Parsons (2005a), we presented a syntax and semantics for two classes of deliberation dialogs, which we have extended in this paper. That earlier framework assumes implicitly that only the speaker of a proposed action may revoke or retract it; hence, that framework does not deal with promises, commands or related locutions. The current paper is the first to consider deliberation dialogs in which the agent first making a proposal may not necessarily be the agent empowered to revoke or retract it.

Our approach differs from related work. The social semantics of Singh and Colombetti and their respective colleagues (Singh 1999, Colombetti and Verdicchio 2002) treats utterances in agent dialogs as manipulating the social relationships between the speakers. Our work, focused only on deliberation dialogs, and thus on statements about actions, is at a lower level of abstraction than social semantics. We assume that a deliberation dialog commences with two or more participants joining together with the shared intention of deciding what action or actions to take in some circumstance. There may already be prior social relationships between the participants, which could thereby allow, for example, commands to be uttered legally by one agent to another. However, once a deliberation dialog commences, we desire to understand how agreement is reached (or not reached) between participants in the dialog. Our focus is therefore on the short-term effects of utterances on the space of action-intention tokens, not their longer-term effects on the social relationships between the participants.

One could ask why the semantic differences of speech acts identified in Section 1 could not be captured by the notion of agent roles, as in a framework such as that of Wooldridge *et al.* (2000). The reason is that the role of revoker or retractor of an utterance is not usually fixed throughout an interaction; it potentially depends on: the nature of the utterance (promise, command, etc); the identities of the agent making the utterance, and the agent receiving it; and on the history of the dialog to that point. All of these may change through the course of a dialog, particularly if there are embedded dialogs or other complex combinations of dialogs, and so agent roles will usually be too rigid a framework for tracking this ability to revoke utterances.

Our notion of a shared space of action-intention tokens has some similarities to other work. For instance, Hamblin's dialog commitment stores (Hamblin 1970), are shared spaces tracking the propositions to which dialog participants have endorsed in a dialog. Similarly, the use of a shared deal space in negotiation dialogs was discussed informally in Jennings *et al.* (2001) and implemented in the negotiation system of Bratu *et al.* (2002). However, neither of these approaches explicitly considered intentionality, so the objects in the shared space represent actions (possible deals), rather than action-intentions. Moreover, neither work defines the shared space or its contents formally, for example as objects in a mathematical semantics for agent negotiation interactions.

Future work will include an implementation of this framework, and further study of its formal and operational properties. Implementation of the framework is likely to be facilitated by viewing the shared space of action-intention-tokens as a *co-ordination artifact*, manipulated by the participants through their utterances; we would thereby be able to draw on recent research on the theory and implementation of such artefacts (Viroli and Ricci 2004), a theory which itself generalizes blackboards, tuple spaces and similar frameworks for agent co-ordination. We also plan to extend the framework to allow for the addition of agent identifiers for agents not in the dialog (so that dialog participants may discuss action-options to be executed by others) and to allow for actions to be executed by more than one agent.[8]

Bibliography

Amgoud, L., N. Maudet, and S. Parsons: 2000, 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. IEEE Press.

Austin, J. L.: 1962, *How To Do Things with Words.* Oxford University Press, Oxford, UK.

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

Bratu, M., J. M. Andreoli, O. Boissier, and S. Castellani: 2002, A software infrastructure for negotiation within inter-organisational alliances. In J. Padget, D. C. Parkes, N. M. Sadeh, O. Shehory, and W. E. Walsh, editors, *AMEC-IV: Designing Mechanisms and Systems*, Lecture Notes in Artificial Intelligence 2531, pages 161—179. Springer, Berlin, Germany.

Colombetti, M., and M. Verdicchio: 2002, An analysis of agent speech acts as institutional actions. In C. Castelfranchi and W. L. Johnson, editors, *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002)*, pages 1157—1164, New York, USA. ACM Press.

FIPA: 2002, *Communicative Act Library Specification*. Standard SC00037J. Foundation for Intelligent Physical Agents.

Gell, A.: 1998, *Art and Agency: An Anthropological Theory*. Clarendon Press, Oxford, UK.

Gunter, C. A.: 1992, *Semantics of Programming Languages: Structures and Techniques*. MIT Press, Cambridge, MA.

Habermas, J.: 1984, *The Theory of Communicative Action: Volume 1: Reason and the Rationalization of Society.* Heinemann, London, UK. Translation by T. McCarthy of: *Theorie des Kommunikativen Handelns, Band I, Handlungsrationalitat und gesellschaftliche Rationalisierung*. Suhrkamp, Frankfurt, Germany, 1981.

Hamblin, C. L.: 1970, *Fallacies*. Methuen, London, UK.

Hitchcock, D.: 1991, 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 (ISSA 1991)*, pages 236—243, Amsterdam, The Netherlands. SICSAT.

Jennings, N. R., P. Faratin, A. R. Lomuscio, S. Parsons, M. Wooldridge, and C. Sierra: 2001, Automated negotiation: prospects, methods and challenges. *Group Decision and Negotiation*, 10(2):199—215.

Johnson, M. W., P. McBurney, and S. Parsons: 2003, When are two protocols the same? In M.-P. Huget, editor, *Communication in Multi-Agent Systems: Agent Communication Languages and Conversation Policies*, Lecture Notes in Artificial Intelligence 2650, pages 253—268. Springer, Berlin, Germany.

Kamp, H., and U. Reyle: 1993, *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory.* Kluwer, Dordrecht, The Netherlands.

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

Mac Lane, S.: 1998, *Categories for the Working Mathematician*. Springer, New York, USA.

McBurney, P., R. M. van Eijk, S. Parsons, and L. Amgoud: 2003, A dialogue-game protocol for agent purchase negotiations. *Journal of Autonomous Agents and Multi-Agent Systems*, 7(3): 235—273.

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

McBurney, P., and S. Parsons: 2005a, A denotational semantics for deliberation dialogues. In I. Rahwan, P. Moraitis, and C. Reed, editors, *Argumentation in Multi-Agent Systems*, Lecture Notes in Artificial Intelligence 3366, pages 162—175. Springer, Berlin, Germany.

McBurney, P., and S. Parsons: 2005b, Locutions for argumentation in agent interaction protocols. In R. M. van Eijk, M-P. Huget, and F. Dignum, editors, *Developments in Agent Communication*, Lecture Notes in Artificial Intelligence 3396, pages 227—244. Springer, Berlin, Germany.

Searle, J.: 1969, *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge, UK.

Singh, M. P.: 1999, An ontology for commitments in multiagent systems: toward a unification of normative concepts. *Artificial Intelligence and Law*, 7: 97—113.

Viroli, M., and A. Ricci: 2004: Instructions-based semantics of agent-mediated interaction. In N. R. Jennings, C. Sierra, E. Sonenberg, and M. Tambe, editors, *Proceedings of the Third International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2004)*, pages 102—109, New York, USA. ACM Press.

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

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

Wooldridge, M. J., N. R. Jennings, and D. Kinny: 2000, The Gaia methodology for agent-oriented analysis and design. *Journal of Autonomous Agents and Multi-Agent Systems*, 3(3): 285—312.

---

[1] www.fipa.org

[2] This view of the semantics of dialog is similar to that of Discourse Representation Theory in linguistics (Kamp and Reyle 1993).

[3] This view owes much to Alfred Gell's anthropological theory of art (Gell 1998), which views artistic artifacts as understood by their recipients as being tokens of intentionality (by an artist, a community, and/or a spiritual being).

[4] We assume the standard definition of a category, in which a collection of objects, arrows between some pairs of objects, and an identity arrow from each object to itself, obey certain composition and associativity rules.

[5] Note that Rule TS5 only permits an agent to utter a statement which deletes an arrow arising from a prior utterance by that same agent.

[6] Thus, for an agent engaged in utility-maximizing behaviour, FP(t, i) would represent its estimated maximum expected utility, evaluated at t, of all future actions believed by the agent Pi to be possible.

[7] Although the model provides a suitable framework for reasoning about what agents do if they engage in a dialog under a protocol from DA.

[8] We are grateful for financial support received from the Information Society Technologies (IST) programme of the European Commission through Project *ASPIC: Argumentation Service Platform with Integrated Components* (IST-FP6-002307). We also thank Jan Albert van Laar and the anonymous referees of this paper for their comments and careful reading.