# Syntax and Semantics of the *Fatio* Argumentation Protocol


Peter McBurney
Department of Computer Science
University of Liverpool
Liverpool L69 7ZF UK
p.j.mcburney@csc.liv.ac.uk

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



## Abstract

*Recent work in the design of agent interaction protocols has focused on applications involving dialectical argumentation — the giving and receiving of reasons for statements. Yet no widely-accepted standard set of locutions exist for such argument. Drawing on both speech act theory and the philosophy of communicative action, we first present a novel typology of speech acts for agent communications. We use this as the basis for proposing an interaction protocol comprising five locutions we consider necessary for argumentation. Both an axiomatic and an operational semantics are given for the protocol.*


## 1. Introduction

The last decade has seen considerable attention devoted to designing generic communication languages for agent interactions. The most widely adopted of these languages is the Foundation for Intelligent Physical Agents' Agent Communication Language (*FIPA ACL*) [6]. Although the leading standard, *FIPA ACL* has been criticized on several grounds, for example: that it requires co-operative agents and sincerity [17]; that its axiomatic semantics *SL* cannot be verified in open systems [24]; that specific locutions have inappropriate semantics [17]; and that it does not readily facilitate the expression of *self-transformation*, the process by which agents change their beliefs, preferences or intentions as a result of their interactions with one another [14].

This paper will address another criticism previously made of the *FIPA ACL*: that it encodes an impoverished theory of argumentation [14]. Agents participating in a dialogue using *FIPA ACL* have only limited means to question or contest information given to them by others; indeed, there are even limitations on what information may be confirmed [17]. Moreover, the semantic language defines the post-conditions of utterances only in terms of their effects on the mental states of hearers; there are no rules concerning combinations of locutions, or the dialectical obligations of the participants, such as requiring questions to be answered, or requiring that assertive statements be justified when challenged. In part, this weakness is due to the implicit assumption of the participants being co-operative; fully co-operating agents, presumably, do not lie, obfuscate or prevaricate. In part also, the lack of a rich argumentation structure is due to a semantics defined without reference to time and only in terms of single locutions, rather than in terms of conversations or protocols [17].

From the perspective of a designer of an open agent system comprising intelligent, autonomous and self-interested agents the absence of a sophisticated and explicit argumentation theory for the agent interaction protocol is a serious obstacle. In open systems, the beliefs of agents may not coincide, and so their interactions will require dialogues involving information seeking, mutual inquiry and persuasion; similarly, their intentions may also not coincide, and so their interactions will require dialogues involving persuasion, commands, negotiation and deliberation [23]. This need explains the recent attention given in the multiagent systems literature to the design of protocols for argument-based interaction, eg, [1, 10, 11]. Indeed, even in models of decision-making by a single agent the essential role of argument-based reflection and evaluation has recently been emphasized, eg [20].

But no standard locutions or protocols yet exist for the requesting and providing of reasons for beliefs and intentions between agents. The main contribution of this paper is to propose the first such set of locutions. These are presented in the form of an interaction protocol, the *Fatio* protocol (Section 3), which may be used by a system designer as a stand-alone protocol or may be incorporated into other protocols and ACLs, such as *FIPA ACL*. The definition of the protocol is undertaken within a novel typology of speech acts, which we propose in Section 2. An axiomatic and an operational semantics for the protocol are given in Section 4. An example application is presented in Section 5, and the paper concludes with a discussion of the value of the protocol in Section 6.

## 2. Types of Speech Acts

Before presenting our argumentation protocol, we first present a syntactic classification of agent speech acts, based on the earlier classifications of John Austin [2], John Searle [19] and Jürgen Habermas [7]. Both Austin and Searle developed their classifications at a time when the prevailing approach to the semantics of propositions in the philosophy of (human) language was truth-conditional. Under this semantics, due originally to Gottlob Frege and Alfred Tarski, if an agent A claims to believe some proposition $P$, then the relevant semantic question is: *"Is P true?"* This view of meaning has been criticized, most prominently by Michael Dummett [4] and Crispin Wright [25], on the basis that for most assertive statements we cannot answer this question definitively; for example, almost all propositions about the past or the future have an inherent uncertainty. At best, we can find evidence for believing $P$ rather than for holding some contrary belief. A better question, therefore, would be: *"Can agent A justify his belief in P?"* This is the essence of verificationist semantics.

Habermas, in his philosophy of communicative action, extended verificationist ideas to statements other than those concerning factual propositions [7, 8], for example, to expressions of preference and to imperative statements. In doing so, he asked how such statements may be justified. One may justify a command, for example, by reference to some social relationship in which the speaker is superior to the hearer. This viewpoint led Habermas to revise Searle's typology of speech acts [7, pp. 325–326], and we have used this revised typology as the basis for our own classification.[1] Our typology is based on, firstly, the entities referenced by the utterance, for example: states of the real-world, internal mental states of the speaker, or the social relationships between participants; and secondly, on the nature of attacks which may be made on an utterance of each type. These attributes are described below and summarized in Table 1.

1. **Factual Statements:** Statements which claim to represent the state of the external world, and so may be objectively-verified. Examples are statements of belief about factual matters. For these statements, reasons for belief may be requested and provided. Contesting such a statement means denying that it is a true description of the reality external to the dialogue. Defending such a statement involves providing objective verification for it, or the provision of means by which it may be objectively verified.

2. **Expressive Statements:** Statements which seek to represent the state of the speaker's internal world, i.e., they aim to reveal publicly a subjective preference, a value assignment, or an intention. As with factual statements, the reasons for a speaker having the mental states revealed by a value statement may be requested and provided. However, value statements cannot be objectively verified or refuted; instead, only the sincerity of the speaker may be challenged. Sincerity of a speaker's internal states may be assessed by the consistency of these; if a speaker knowingly expresses conflicting intentions, for example, a hearer would be entitled to conclude that the speaker is not sincere about one or more intentions.

3. **Social Connection Statements:** Statements which assert some social or other relationship between different

| No. | Type | Referants | Basis of Attack |
|---|---|---|---|
| 1 | Factual Statements | Real-world states | Verifiability |
| 2 | Expressive Statements | Internal mental states of speaker | Sincerity |
| 3 | Social Connection Statements | Social relations | Normative rightness |
| 4 | Commissives | Social relations<br>Internal mental states of Speaker | Sincerity<br>Feasibility<br>Consequences<br>Efficacy<br>Etc |
| 5 | Directives | Social relations<br>Internal mental states of Speaker and Hearer | Normative rightness<br>Feasibility<br>Consequences<br>Efficacy<br>Etc |
| 6 | Inferences | Consequences of earlier utterances | Inferential validity |
| 7 | Argumentation Statements | Content of utterances | Dialogical rightness |
| 8 | Control Statements | Form of utterances | |

**Table 1. Typology of Speech Acts**

1 The structural basis of our typology differs from that of Habermas, as we consider both the referants of locutions and the basis on which they may be attacked. We therefore include commissives as an explicit category, and not a sub-category of expressive statements. We have also divided communicative acts into two categories, Argumentation Statements and Control Statements. Finally, our labels also differ in some cases.

participants. Examples of such relationships include: employer and employee; customer and supplier; peers; etc. Again, the reasons for a speaker asserting a particular connection may be requested and provided; contestation of these utterances takes the form of challenging the normative rightness of the relationship.

4. **Commissive Statements:** Here, the speaker desires that the world be in a particular state, and so commits to the hearer to undertake some action or course of action to establish or maintain this world state. Promises and vows are examples. They are stronger than merely expressive statements because, if accepted, they create obligations on the speaker in the world beyond the dialogue. Accordingly, they make reference to both the internal mental states (eg, desires, intentions) of the speaker and to the social relations existing between speaker and hearers. Because they refer to internal states, they may be challenged on the grounds of sincerity. They may also be challenged on substantive grounds, eg: their direct or indirect costs and benefits; their opportunity costs; their consequences; their practical feasibility; etc.

5. **Directive Statements:** Here, the speaker desires that the world be in a particular state, and so seeks that the hearer commits to undertaking some action or course of action to establish or maintain this world state. Requests, commands, warnings and entreaties are examples of these statements. As with commissives, they are stronger than merely expressive statements because, if accepted, they create obligations on the hearer to the speaker in the world beyond the dialogue. They make reference to the internal mental states of both the speaker (eg, desires) and the hearers (eg, intentions) and to the social relations existing between speaker and hearers. They may be challenged on all the substantive grounds for which commissives may be challenged, in addition to contestation of the normative rightness of the social connections required for their valid utterance.

6. **Inference Statements:** Statements which refer to the content of earlier statements in a dialogue, drawing inferences from them or assessing their implications. Contestation of such statements can take the form of questioning the appropriateness or the validity of the inferences made.

7. **Argumentation Statements:** Statements which refer to the contents of prior speech acts, for example, questions, challenges, and requests for justification. These acts may be attacked on the basis of inappropriateness, timing or dialectical validity. They may also create dialectical obligations on the part of the speaker and/or the hearers.

8. **Control Statements:** Statements which refer to speech itself, aiming to synchronize communication. Examples of such statements are requests to repeat an utterance, or acknowledgments that an utterance was received.

Both commissive and directive statements concern commitments to undertake an action (or a course of action) to create or maintain a state of the world in which specified propositions are true. Because of these commitments, the allowable attacks for these statements are stronger than are those for expressive statements. Thus, for example, promising someone to undertake a specific action is a stronger statement than expressing an intention to undertake it. It is for this reason that we consider commissives as a separate class of speech act, rather than being considered as a subclass of expressive statements.

| **Locution Type** | **FIPA ACL Locutions** |
|---|---|
| Factual Statements | confirm<br>disconfirm<br>failure<br>inform<br>inform-if<br>inform-ref<br>query-if<br>query-ref |
| Expressive Statements | inform |
| Social Connection Statements | inform |
| Commissives | accept-proposal<br>agree<br>propose<br>refuse<br>reject-proposal |
| Directives | cancel<br>cfp<br>request<br>request-when<br>request-whenever |
| Inferences | inform |
| Argumentation Statements | |
| Control Statements | not-understood<br>propagate<br>proxy<br>subscribe |

**Table 2. Classification of FIPA ACL locutions**

As an aid for understanding, we have classified the 22 locutions of *FIPA ACL* according to our typology (Table 2).

As can be seen, making expressive statements, social connection statements or inferences requires use of the *inform* illocution, the same illocution used to make factual statements. This conflation creates problems for the designer of argumentation-theoretic combination rules for locutions, for such rules would need to examine the content of an *inform* utterance to determine what locutions are valid in response.[2] Moreover, Table 2 also shows that none of the *FIPA ACL* locutions relate to argumentation; this supports the criticism, made in [14], of an impoverished argumentation theory underlying *FIPA ACL*.

It would be possible to design a new generic Agent Communications Language using this typology. In doing so, it would be sensible to build on the *FIPA ACL*, which (one might argue) is strong in illocutions for factual statements, for commissives and for directives. A new generic ACL could also build on the recent work in the agent communications literature on social semantics, which considers utterances in a dialogue as means to manipulate inter-participant commitments in some wider social institution [3, 21]. The work on social semantics therefore provides illocutions by which commissives and directives may be uttered, responded to, and modified, and a framework for viewing these illocutions. Finally, recent research has also considered illocutions for inter-participant synchronization, the category of acts we have termed Control Statements [16]. Combining all this prior work to design a new generic ACL would be very ambitious, and so for this paper, we consider only illocutions for argumentation.

## 3. Protocol Syntax

We now present a protocol for undertaking arguments over statements in a dialogue. The statements made in the argument may be any of the first six types of illocutions: factual statements, expressive statements, social connection statements, commissives, directives, and inferences. Our aim here is to define a concise, generic protocol which can be used on its own, or incorporated into other interaction protocols or ACLs (including *FIPA ACL*).[3] For ease of reference, we name the protocol *Fatio*.[4]

Our syntax for utterances will be:

*illocution*$(P_i, \phi)$ or *illocution*$(P_i, P_j, \phi)$

where *illocution* is an illocution, $P_i$ is an identifier for the agent making the utterance (the speaker), $P_j$ denotes an agent at whom the utterance is directed, and $\phi$ is the content of the utterance. It would also be possible to have an identifier for the intended recipient (the hearer) of the utterance, but for simplicity we assume all utterances are made to the entire group involved in the dialogue. For the content of the utterance, any agreed formal languages may be used. We will assume the content layer is represented in a propositional language, with lower-case Greek letters as propositions. We denote the set of these well-formed content formulae, closed under the usual connectives, as $\mathcal{C}$. These propositions may represent objectively-verifiable statements about the world, or internal preferences, or intentions, or commitments, etc. Because we wish to use the protocol to exchange justifications for claims, some utterances will also have content comprising arguments (eg, premises and inference-rules), which will be represented by upper-case Greek letters. We denote the set of these well-formed argument formulae, closed under the usual connectives, as $\mathcal{A}$. Note that $\mathcal{C}$ is a proper subset of $\mathcal{A}$. If $\phi \in \mathcal{C}$ is a proposition in the content language and $\Phi \in \mathcal{A}$ is a justification, we will write $\Phi \vdash^{+} \phi$ to indicate that $\Phi$ is an argument in support of $\phi$, and $\Phi \vdash^{-} \phi$ to indicate that $\Phi$ is an argument against $\phi$. Finally, we assume that time is discrete and may be represented by the natural numbers, and that precisely one utterance occurs on each time-step. The first utterance of each dialogue is made at time-step 1. For simplicity, we do not include a time stamp in the syntax.

Before proceeding, we note the differing usages of the word "commitment" in the agent communications and argumentation literatures.[5] In the early dialogue game literature in philosophy, commitments refer to dialectical obligations incurred by participants inside a dialogue, and may have no relationship to the true beliefs or actions of the participants [9]. More recently, philosophers of argumentation have defined commitments more broadly, to include actions intended to establish or maintain a particular state of the world, including states of the dialogue [23]. Commitments as actions external to the dialogue is closer to usage in agent communications literature, eg, [3, 21].

For the *Fatio* protocol, we reserve the word *commitment* to refer only to actions external to the dialogue: in other words, only commissive and directive statements make reference to commitments. We use the term *dialectical obligations* to refer to commitments inside the dialogue, for example, an obligation within a dialogue to defend an assertion against attack by another participant.

We now define the five legal locutions in *Fatio*:

**F1: assert**$(P_i, \phi)$**:** A speaker $P_i$ asserts a statement $\phi \in \mathcal{C}$ (a belief, an intention, a social connection, an external commitment, etc). In doing so, $P_i$ creates a dialectical

2 To be fair to the designers of *FIPA ACL*, they do not appear to have imagined that *inform* utterances would elicit attacks, further indication of an assumption of fully co-operative participants.

3 If used as a stand-alone protocol, additional locutions for entry to, withdrawal from, and termination of the dialogue would be necessary [12]. These could be taken from another protocol, eg, [10].

4 After Nicolas Fatio de Duillier (1664–1753), a Swiss mathematician and polymath, and famous disputant, on Newton's behalf, with Leibniz over who invented the differential calculus.

5 Note that we are not using the word in the sense of an agent's persistent intentions.

obligation within the dialogue to provide justification for $\phi$ if required subsequently by another participant.

**F2: question**($P_j, P_i, \phi$)**:** A speaker $P_j$ questions a prior utterance of *assert*($P_i, \phi$) by another participant $P_i$, and seeks a justification for $\phi$. The speaker $P_j$ of the question creates no dialectical obligations on himself by the question utterance.

**F3: challenge**($P_j, P_i, \phi$)**:** A speaker $P_j$ challenges a prior utterance of *assert*($P_i, \phi$) by another participant $P_i$, and seeks a justification for $\phi$. In contrast to a question, with this locution, $P_j$ also creates a dialectical obligation on himself to provide a justification for not asserting $\phi$, for example an argument against $\phi$, if questioned or challenged. Thus, *challenge*($P_j, P_i, \phi$) is a stronger utterance than *question*($P_j, P_i, \phi$).

**F4: justify**($P_i, \Phi \vdash^+ \phi$)**:** A speaker $P_i$ who had uttered *assert*($P_i, \phi$), and was then questioned or challenged by another speaker, is able to provide a justification $\Phi \in \mathcal{A}$ for the initial statement $\phi$ by means of this locution. The utterance *justify*($P_i, \Phi \vdash^- \phi$) is similarly defined.

**F5: retract**($P_i, \phi$)**:** A speaker $P_i$ who had uttered *assert*($P_i, \phi$) or *justify*($P_i, \Phi \vdash^+ \phi$) can withdraw this statement with the utterance f *retract*($P_i, \phi$) or *retract*($P_i, \Phi \vdash^+ \phi$), respectively. This removes the earlier dialectical obligation on $P_i$ to justify $\phi$ or $\Phi$ if questioned or challenged.

As part of the protocol, these locutions are subject to several combination rules [12]:

**CR1:** The utterance *assert*($P_i, \phi$) may be made at any time.

**CR2:** The utterances *question*($P_j, P_i, \phi$) and *challenge*($P_j, P_i, \phi$) may be made at any time following an utterance of *assert*($P_i, \phi$). Similarly, the utterances *question*($P_j, P_i, \Phi$) and *challenge*($P_j, P_i, \Phi$) may be made at any time following an utterance of *justify*($P_i, \Phi \vdash^+ \phi$).

**CR3:** Immediately following an utterance of *question*($P_j, P_i, \phi$) or *challenge*($P_j, P_i, \phi$), the speaker $P_i$ of *assert*($P_i, \phi$) must reply with *justify*($P_i, \Phi \vdash^+ \phi$), for some $\Phi \in \mathcal{A}$.

**CR4:** The utterances *question*($P_k, P_j, \phi$) and *challenge*($P_k, P_j, \phi$) may be made at any time following an utterance of *challenge*($P_j, \phi$).

**CR5:** Following an utterance of *question*($P_k, P_j, \phi$) or *challenge*($P_k, P_j, \phi$), the speaker $P_j$ of *challenge*($P_j, \phi$) must reply immediately with *justify*($P_j, \Delta \vdash^- \phi$), for $\Delta \in \mathcal{A}$.

**CR6:** The utterance *retract*($P_i, \phi$) may be made at any time following an utterance of *assert*($P_i, \phi$). The utterance *retract*($P_i, \Phi$) may be made at any time following an utterance of *justify*($P_i, \Phi \vdash^+ \phi$).

Some comments are in order. First, we may ask why each locution is needed. *assert(.)* is necessary in order that some assertion be made explicitly which can form the basis of argument. If these locutions are added to an existing ACL, then an existing illocution could be used for *assert(.)*, provided it is given an appropriate semantics. For instance, in *FIPA ACL*, *inform(.)* could be used, so long as it given the semantics assigned to *assert(.)* in Section 4. Then, given a claim, other participants need illocutions to request justifications for the claim from the original speaker. Both *question(.)* and *challenge(.)* enable this. Once questioned or challenged, an illocution is needed to enable the original speaker to present her justification for the claim; this is made possible with *justify(.)*. In the course of an argument, a rational speaker may change her view on the matter under discussion, and therefore an illocution is needed to enable the speaker to express this *self-transformation* [14]; the illocution *retract(.)* enables this.

Another question here is why both *question(.)* and *challenge(.)* are needed. The answer lies in the semantics of the two locutions. On the basis of the informal definitions above (and ignoring any complications due to the timing of utterances), one could say that the utterance:

*challenge*($P_j, P_i, \phi$)

is equivalent in effect to the sequence of two utterances:

*question*($P_j, P_i, \phi$)
*assert*($P_j, \neg\phi$).

However, this equivalence only holds if the propositions concern beliefs and the Law of Excluded Middle (LEM) is assumed. Participant $P_j$ may have an argument against $\phi$, but no argument in favor of $\neg\phi$; if LEM is not assumed, then having a negative argument against one is not necessarily the same as having a positive argument for the other. Similarly, if, for example, the propositions concern intended actions, then attacking a proposed action $\phi$ is not at all the same as supporting a conflicting action included in $\neg\phi$. Accordingly, we retain both *question(.)* and *challenge(.)*.

Another question is whether these five locutions are sufficient for argumentation-based interactions. An obvious absence is a specific illocution to enable a hearer $P_j$ to endorse or support an earlier utterance of *assert*($P_i, \phi$) by a speaker $P_j$. For some application domains, such as the scientific inquiries of [11], hearers may desire to indicate partial or qualified acceptance, and so specific locutions are valuable. However, for the generic protocol, participant $P_j$ can indicate acceptance with an utterance of *assert*($P_j, \phi$).

## 4. Protocol Semantics

We now present both an axiomatic and an operational semantics for the locutions in the *Fatio* protocol. The axiomatic semantics is useful for understanding the intended meaning of each locution, while the operational semantics better reveals the impacts of utterances on hearers. In addition, because the axiomatic semantics we present is similar to the modal semantics of *FIPA ACL* [6], our work provides a formal basis for an agent system designer to use the five *Fatio* locutions within *FIPA ACL*. However, since each semantics makes assumptions about the internal mental states of the participants, neither is verifiable [24].

### 4.1. Axiomatic Semantics

An **axiomatic semantics** for a programming language defines the pre-conditions and post-conditions for each command [22]. We give a formal, axiomatic semantics for the locutions of *Fatio* in terms of the beliefs and desires of the participating agents, when the content of utterances under the protocol refer to beliefs. This approach could equally well apply for other types of utterance content, such as intentions, social connections, commitments, etc. The classes $\mathcal{C}, \mathcal{A}$ are as before.

Central to the axiomatic semantics are publicly-viewable stores to record dialectical obligations of the participants. Following [9], we define a private-write, public-read store for each participant $P_i$ in the dialogue, called a *dialectical obligations store*, $DOS(P_i)$, and containing the dialectical obligations currently incurred by $P_i$. All participants can view this store, but only $P_i$ may write to it (by means of the appropriate utterances). We denote the contents of $DOS(P_i)$ by triples, $(P_i, X, Y)$, where $P_i$ is a participant, where $X \in \mathcal{C}$ is a well-formed formula in the content language (here a proposition about the world), or $X \in \mathcal{A}$ is a well-formed formula in the argument language, and $Y \in \{+, -\}$. The triple $(P_i, \phi, +) \in DOS(P_i)$ denotes that participant $P_i$ has a dialectical obligation to provide justification or argument in support for the proposition $\phi$, while the triple $(P_i, \phi, -) \in DOS(P_i)$ denotes that participant $P_i$ has a dialectical obligation to provide justification or argument against the proposition $\phi$. Likewise, for triples $(P_i, \Phi, .\ ) \in DOS(P_i)$, where the second element of the triple is an argument $\Phi$, the participant $P_i$ has a dialectical obligation to provide justification for or against $\Phi$.

Our semantics is specified in terms of two classes of modal operators, $\{B_i, D_i\}$, where $i$ is an agent identifier. Other symbols have the same definitions as in Section 4.1. These classes have the following intended interpretations:[6]

$B_i\phi$ : "Agent $i$ believes that $\phi$ is true."

$D_i\phi$ : "Agent $i$ desires that $\phi$ be true."

We also use simplified elements of FIPA's action language [6, Annex A]; in particular, we let

*Done [*illocution$(P_i, \phi)$*, pre-con]*

indicate that *illocution*$(P_i, \phi)$ has been uttered by participant $P_i$ with content $\phi$, and with pre-conditions *pre-con* true just before the utterance. These operators may be embedded to any depth. Accordingly, we can now define the five locutions of *Fatio* in terms of these modal operators and the *done* operator, presenting pre- and post-conditions for each instantiated locution. Where there are two similar locutions, for reasons of space we only present the full details for one.

- *assert*$(P_i, \phi)$

  *Pre-conditions:* A speaker $P_i$ desires that each participant $P_j (j \neq i)$, believes that $P_i$ believes the proposition $\phi \in \mathcal{C}$.

  $((P_i, \phi, +) \notin DOS(P_i)) \wedge (\forall j \neq i)(D_i B_j B_i \phi)$.

  *Post-conditions:* Each participant $P_k (k \neq i)$, believes that participant $P_i$ desires that each participant $P_j (j \neq i)$, believe that $P_i$ believes $\phi$.

  $(P_i, \phi, +) \in DOS(P_i) \wedge$
  $(\forall k \neq i)(\forall j \neq i)(B_k D_i B_j B_i \phi)$.

  *Dialectical Obligations:* $(P_i, \phi, +)$ is added to $DOS(P_i)$, the Dialectical Obligations Store of speaker $P_i$.

- *question*$(P_j, P_i, \phi)$

  *Pre-conditions:* Some participant $P_i (i \neq j)$ has a dialectical obligation to support $\phi$ and participant $P_j$ desires that each other participant $P_k (k \neq j)$, believe that $P_j$ desires that $P_i$ utter a *justify*$(P_i, \phi, .)$ locution.

  $\exists i (i \neq j)(((P_i, \phi, +) \in DOS(P_i)) \wedge$
  $(\forall k \neq j) D_j B_k D_j (\exists \Delta \in \mathcal{A})$
  *Done [justify*$(P_i, \Delta \vdash^+ \phi)$*)*$, ((P_i, \phi, +) \in DOS(P_i))]))$.

  *Post-conditions:* Participant $P_i$ must utter a *justify* locution.

  $(\exists \Delta \in \mathcal{A})$ *Done (justify*$(P_i, \Delta \vdash^+ \phi)$*)*,
  *Done [question*$(P_j, P_i, \phi)$*)*$, ((P_i, \phi, +) \in DOS(P_i))])$.

  *Dialectical Obligations:* No effect.

- *justify*$(P_i, \Phi \vdash^+ \phi)$

  *Pre-conditions:* A speaker $P_i$ has a dialectical obligation to support $\phi \in \mathcal{C}$, another speaker $P_j (j \neq i)$ has uttered a *question*$(P_j, P_i, \phi)$ or a *challenge*$(P_j, P_i, \phi)$ locution, and $P_i$ desires that each participant $P_k (k \neq i)$ believes that $P_i$ believes that $\Phi \in \mathcal{A}$ is an argument for $\phi$.

  $((P_i, \phi, +) \in DOS(P_i)) \wedge$
  (*Done [question*$(P_j, P_i, \phi)$*)*$, ((P_i, \phi, +) \in DOS(P_i))] \vee$
  *Done [challenge*$(P_j, P_i, \phi)$*)*$, ((P_i, \phi, +) \in DOS(P_i))]) \wedge$

6 Note that beliefs and desires are time-dependent. Because we are ignoring time in the locution syntax, we also ignore it here.

$(\exists \Phi \in \mathcal{A})(\forall k \neq i)(D_i B_k B_i(\Phi \vdash^+ \phi))$.

*Post-conditions:* Each participant $P_k (k \neq i)$ believes that $P_i$ desires that each participant $P_j (j \neq i)$ believes that $P_i$ believes that $\Phi \in \mathcal{A}$ is an argument for $\phi$.

$((P_i, \phi, +) \in DOS(P_i)) \wedge ((P_i, \Phi, +) \in DOS(P_i)) \wedge$
$(\forall k \neq i)(\forall j \neq i)(B_k D_i B_j B_i(\Phi \vdash^+ \phi))$.

*Dialectical Obligations:* $(P_i, \Phi, +)$ is added to $DOS(P_i)$, the Dialectical Obligations Store of speaker $P_i$.

- *challenge*($P_j, P_i, \phi$)

  *Pre-conditions:* Some participant $P_i (i \neq j)$ has a dialectical obligation to support $\phi$ and participant $P_j$ desires that each other participant $P_k (k \neq j)$, believe both that $P_j$ desires that $P_i$ utter a *justify*($P_i, \Delta \vdash^+ \phi$) locution for some $\Delta \in \mathcal{A}$ and that $P_j$ does not believe $\phi$.

  $\exists i (i \neq j)(((P_i, \phi, +) \in DOS(P_i)) \wedge$
  $(\forall k \neq j)(D_j B_k \neg B_j \phi)$
  $\wedge (\forall k \neq j) D_j B_k D_j (\exists \Delta \in \mathcal{A})$
  *Done* [*justify*($P_i, \Delta \vdash^+ \phi$), $((P_i, \phi, +) \in DOS(P_i))$]])).

  *Post-conditions:* Participant $P_i$ must utter a *justify* locution and speaker $P_j$ becomes dialectically obligated to provide an argument against $\phi$ if questioned or challenged.

  $((P_j, \phi, -) \in DOS(P_j) \wedge$
  $(\exists \Delta \in \mathcal{A})$ *Done* [*justify*($P_i, \Delta \vdash^+ \phi$),
  *Done* [*question*($P_j, P_i, \phi$), $((P_i, \phi, +) \in DOS(P_i))$]]].

  *Dialectical Obligations:* $(P_j, \phi, -)$ is added to $DOS(P_j)$, the Dialectical Obligations Store of speaker $P_j$.

- *retract*($P_i, \phi$)

  *Pre-conditions:* For proposition $\phi \in \mathcal{C}$, with $(P_i, \phi, +) \in DOS(P_i)$, $P_i$ desires that each participant $P_j (j \neq i)$ believes that $P_i$ no longer believes $\phi$. For proposition $\phi \in \mathcal{C}$, with $(P_i, \phi, -) \in DOS(P_i)$, $P_i$ desires that each participant $P_j (j \neq i)$ believes that $P_i$ no longer does not believe $\phi$.

  $((P_i, \phi, +) \in DOS(P_i) \wedge (\forall j \neq i)(D_i B_j \neg B_i \phi))$
  $\vee$
  $(((P_i, \phi, -) \in DOS(P_i)) \wedge$
  $(\forall j \neq i)(D_i B_j \neg\neg B_i \phi))$.

  *Post-conditions:* Depending on the two cases in the pre-conditions, either each participant $P_k (k \neq i)$, believes that participant $P_i$ desires that each participant $P_j (j \neq i)$, believe that $P_i$ no longer believes $\phi$, or each participant $P_k (k \neq i)$, believes that participant $P_i$ desires that each participant $P_j (j \neq i)$, believe that $P_i$ no longer does not believe $\phi$.

  $((P_i, \phi, +) \notin DOS(P_i) \wedge$
  $(\forall k \neq i)(\forall j \neq i)(B_k D_i B_j \neg B_i \phi))$
  $\vee$
  $((P_i, \phi, -) \notin DOS(P_i) \wedge$
  $(\forall k \neq i)(\forall j \neq i)(B_k D_i B_j \neg\neg B_i \phi))$.

  *Dialectical Obligations:* Either $(P_i, \phi, +)$ or $(P_i, \phi, -)$ is removed from $DOS(P_i)$, the Dialectical Obligations Store of speaker $P_i$.

The illocutions *justify*($P_i, \Phi \vdash^- \phi$) and *retract*($P_i, \Phi$) have similar semantics to that for *justify*($P_i, \Phi \vdash^+ \phi$) and *retract*($P_i, \phi$), respectively. Because $\mathcal{C}$ is a proper subset of $\mathcal{A}$, this semantics permits a speaker to use proposition $\phi$ as a justification for itself. Whether or not this is acceptable to the other participants in a dialogue depends on their attitudes at the time [15].

### 4.2. Operational Semantics

We also present an operational semantics for the *Fatio* protocol. An operational semantics indicates how the states of a system change as a result of execution of the commands in a programming language [22]. In this case, the commands in question are the locutions in an argumentation dialogue conducted according the rules of the protocol.[7]

Our definition of the protocol in Section 3 was deliberately exclusively syntactical: we made no assumptions regarding the decision-making architectures or the mental states of the participants before, during or after the dialogue in which they engage. Consequently, any agent willing to submit to the defined rules of the argumentation dialogue may participate in it, regardless of the meaning(s) the agent may place on the utterances made. We believe this property ensures wide applicability. However, the rules for *Fatio* are not sufficient to ensure the automatic generation of agent dialogues. To achieve this, the individual participants need to be vested with mechanisms which will invoke particular utterances at particular points in the dialogue, responding to past and anticipated future utterances. We call these mechanisms *agent decision mechanisms*, although they still may be simulated by the participants, and thus bear little or no relationship to the true decision-making processes or associated “mental states” of the participants.

**4.2.1. Agent Decision Mechanisms** We present a portfolio of internal agent decision mechanisms. Although defined here at a high level, each mechanism is readily implementable using argumentation reasoning methods (**D1–D4**) or meta reasoning methods (**D5**).

**D1($\phi$): Claim or Not:** A procedure, for each statement $\phi$, to enable an agent $P_i$ to decide to utter an *assert*($P_i, \phi$)

7 Other agent communications protocols for which operational semantics have been defined include [5, 10].

locution. If the agent is vested with a reasoning process using argumentation, as in [1, 15], then this procedure may operate by assessing the arguments for and against $\phi$, and then deciding to speak or not on the basis of the agent's *argument assertion attitudes* [15]. The two outputs of this mechanism are: *listen* and *utter-assert($\phi$)*.

**D2: React or Not:** A procedure to enable an agent $P_j$ to decide to utter a *question($P_j, P_i, \phi$)* or a *challenge($P_j, P_i, \phi$)* locution, following an *assert($P_i, \phi$)* utterance. As with mechanism **D1**, an agent using argumentation-based reasoning may decide to speak on the basis of the agent's *argument acceptance attitudes* [15]. The three outputs of this mechanism are: *listen*, *utter-question($P_i, \phi$)* and *utter-challenge($P_i, \phi$)*.

**D3($\phi$): Defend or Not:** A procedure to enable an agent $P_i$ with a dialectical obligation to provide justification for some statement or argument to utter a *justify(.)* locution to meet this obligation. This procedure could include, as a sub-procedure, the identification of the best justification for the statement at this time in the dialogue. The two outputs of this mechanism are: *listen* and *utter-justify(.)*.

**D4($\phi$): Fold or Not:** A procedure to enable an agent $P_i$ with a dialectical obligation to provide justification for some statement or argument to utter a *retract(.)* locution. The two outputs of this mechanism are: *listen* and *utter-retract(.)*.

**D5: Listen or Do:** A procedure to await a new utterance from other participants, and, upon its receipt, to decide which of the four classes of mechanisms **D1–D4** to execute. The five outcomes are: *listen* and *do-mech(Di)*, for $i = 1, 2, 3, 4$.

Note that mechanism **D5** is a meta-level decision mechanism, and may include procedures for intention-reconsideration, as in [18].

**4.2.2. Transition System** We now present the transition rules of the operational semantics for *Fatio*. We assume the participating agents are imbued with the decision mechanisms above, enabling them to initiate utterances and to respond to utterances in the dialogue, and so the states we will take to be the inputs and outputs of these decision mechanisms. The locutions uttered in the dialogue effect transitions between states of the decision mechanisms, as utterances serve as inputs to one or more of the mechanisms of the participating agents, and then these mechanisms in turn produce outputs causing further utterances in the dialogue. Thus, our operational semantics provides a formal linkage between the dialogue utterances and agent decision mechanisms.

To define these links, we let the triple $\langle P_i, \mathbf{K}, s\rangle$ denote the mechanism with number **K** and with an output $s$ of participant $P_i$. For ease of presentation, where a transition is invoked by or invokes a particular output of a mechanism **K** this is denoted by the specific output $s$ in the third place of the triple; where no specific output is invoked, we denote this by a period in the third place, $\langle P_i, \mathbf{K}, .\rangle$. Some transitions occur between mechanisms of different agents by means of dialogue locutions; these are denoted by arrows, labelled by the relevant locution number (**F1–F5**) from Section 3. Other transitions occur between the mechanisms of a single agent; these are denoted by unlabelled arrows. Where different agent subscripts appear in the same transition rule, they refer to distinct agents. The rules are defined as follows, for any statement $\phi$, and for any agents $P_i$, $P_j$ and $P_k$:

**TR1:** $\langle P_i, \mathbf{D1}(\phi), \textit{listen}\rangle \rightarrow \langle P_i, \mathbf{D5}, .\ \rangle$

**TR2:** $\langle P_i, \mathbf{D1}, \textit{utter-assert}(\phi)\rangle \overset{\mathbf{F1}}{\rightarrow} \langle P_i, \mathbf{D5}, \textit{listen}\ \rangle$

**TR3:** $\langle P_i, \mathbf{D1}, \textit{utter-assert}(\phi)\rangle \overset{\mathbf{F1}}{\rightarrow}$
$\langle P_j, \mathbf{D5}, \textit{do-mech(D2)}\ \rangle$

**TR4:** $\langle P_j, \mathbf{D2}, \textit{listen}\rangle \rightarrow \langle P_j, \mathbf{D5}, .\ \rangle$

**TR5:** $\langle P_j, \mathbf{D2}, \textit{utter-question}(P_i, \phi)\rangle \overset{\mathbf{F2}}{\rightarrow} \langle P_j, \mathbf{D5}, \textit{listen}\ \rangle$

**TR6:** $\langle P_j, \mathbf{D2}, \textit{utter-question}(P_i, \phi)\rangle \overset{\mathbf{F2}}{\rightarrow}$
$\langle P_i, \mathbf{D5}, \textit{do-mech(D3(}\phi\textit{))}\ \rangle$

**TR7:** $\langle P_j, \mathbf{D2}, \textit{utter-question}(P_i, \phi)\rangle \overset{\mathbf{F2}}{\rightarrow} \langle P_k, \mathbf{D5}, \textit{listen}\ \rangle$

**TR8:** $\langle P_j, \mathbf{D2}, \textit{utter-challenge}(P_i, \phi)\rangle \overset{\mathbf{F3}}{\rightarrow} \langle P_j, \mathbf{D5}, \textit{listen}\ \rangle$

**TR9:** $\langle P_j, \mathbf{D2}, \textit{utter-challenge}(P_i, \phi)\rangle \overset{\mathbf{F3}}{\rightarrow}$
$\langle P_i, \mathbf{D5}, \textit{do-mech(D3(}\phi\textit{))}\ \rangle$

**TR10:** $\langle P_j, \mathbf{D2}, \textit{utter-challenge}(P_i, \phi)\rangle \overset{\mathbf{F3}}{\rightarrow}$
$\langle P_k, \mathbf{D5}, \textit{listen}\ \rangle$

**TR11:** $\langle P_i, \mathbf{D3}(\phi), \textit{listen}\rangle \rightarrow \langle P_i, \mathbf{D5}, .\ \rangle$

**TR12:** $\langle P_i, \mathbf{D3}(\phi), \textit{utter-justify}(\phi)\rangle \overset{\mathbf{F4}}{\rightarrow} \langle P_i, \mathbf{D5}, \textit{listen}\ \rangle$

**TR13:** $\langle P_i, \mathbf{D3}(\phi), \textit{utter-justify}(\phi)\rangle \overset{\mathbf{F4}}{\rightarrow} \langle P_k, \mathbf{D5}, .\ \rangle$

**TR14:** $\langle P_i, \mathbf{D4}(\phi), \textit{listen}\rangle \rightarrow \langle P_i, \mathbf{D5}, .\ \rangle$

**TR15:** $\langle P_i, \mathbf{D4}(\phi), \textit{utter-retract}(\phi)\rangle \overset{\mathbf{F5}}{\rightarrow} \langle P_i, \mathbf{D5}, \textit{listen}\ \rangle$

**TR16:** $\langle P_i, \mathbf{D4}(\phi), \textit{utter-retract}(\phi)\rangle \overset{\mathbf{F5}}{\rightarrow} \langle P_k, \mathbf{D5}, .\ \rangle$

**TR17:** $\langle P_i, \mathbf{D5}, \textit{listen}\ \rangle \rightarrow \langle P_i, \mathbf{D5}, .\ \rangle$

**TR18:** $\langle P_i, \mathbf{D5}, \textit{do-mech(D1(}\phi\textit{))}\ \rangle \rightarrow \langle P_i, \mathbf{D1}(\phi), .\ \rangle$

**TR19:** $\langle P_i, \mathbf{D5}, \textit{do-mech(D2)}\ \rangle \rightarrow \langle P_i, \mathbf{D2}, .\ \rangle$

**TR20:** $\langle P_i, \mathbf{D5}, \textit{do-mech(D3(}\phi\textit{))}\ \rangle \rightarrow \langle P_i, \mathbf{D3}(\phi), .\ \rangle$

**TR21:** $\langle P_i, \mathbf{D5}, \textit{do-mech(D4(}\phi\textit{))}\ \rangle \rightarrow \langle P_i, \mathbf{D4}(\phi), .\ \rangle$

To briefly illustrate the meaning of these rules, consider rule **TR6**, which indicates that when agent $P_j$ utters locution **F2** — *question*($P_j, P_i, \phi$) — to agent $P_i$, then $P_i$ initiates mechanism **D5** with output mechanism **D3**($\phi$)**: Defend or Not**. Transition Rules **TR11–TR13** show the possible transitions on from agent $P_i$'s execution of mechanism **D3**. And so on.

## 5. Example

We give a brief example of a dialogue conducted under the *Fatio* protocol, between participants labelled *A, B* and *C*. The argument uses four propositions relating to a fictional restaurant, the Brigade Brigade:

**R:** The Brigade Restaurant is excellent.

**P:** I had a great meal at the Brigade.

**Q:** I am vegetarian.

**S:** Vegetarian food at the Brigade is awful.

In the dialogue, utterances are numbered in sequence.

1. *assert(A, R)*
   Agent A asserts *R*.
2. *challenge(B, A, R)*
   Agent B issues a challenge to A for a justification for *R*.
3. *justify*(*A*, $P \vdash^{+} R$)
   Agent A provides a justification for *R*, namely *P*.
4. *question(C, B, R)*
   Agent C questions B over her challenge to A, seeking a justification for the challenge.
5. *justify*(*B*, $Q \,\&\, S \vdash^{-} R$)
   Agent B provides a justification for her challenge to A, namely that *Q* and *S*.
6. *retract(A, R)*
   Hearing this, Agent A retracts his prior assertion of *R*.
7. *question(C, B, Q)*
   Agent C asks B to justify her argument of *Q*.

Although this is only a very simple example, it illustrates the use of the *Fatio* protocol for argument-based interaction, with agents requesting, providing and receiving reasons for statements made in a dialogue.

## 6. Conclusions

The primary contribution of this paper has been to define a new protocol for argument-based dialogue between autonomous, intelligent agents. The protocol, *Fatio*, allows participants to make assertions, request justifications for assertions, make challenges to assertions, provide justifications (or arguments) for assertions, and retract prior assertions. The content of dialogues conducted under the protocol may range over any domain: objectively-verifiable beliefs about the world beyond the dialogue; internal mental states of the participants, such as their preferences and intentions; statements about the social relationships between the participants; commitments of the participants to actions in the world beyond the dialogue; even dialectical obligations of the participants to one another.

In this paper, we also presented a novel classification of illocutions in agent interactions, drawing on speech act theory and the philosophy of communicative action. The definition of the *Fatio* protocol presented illocutions in the category of Argumentation Statements, namely utterances questioning, challenging and defending the contents of other speech acts. As our work shows this is a category ignored by the *FIPA ACL* agent communications language. We also presented both an axiomatic semantics and an operational semantics for the *Fatio* protocol. The axiomatic semantics was defined for the case where the content of utterances concerns the beliefs of the participants and in terms of modal operators representing the beliefs and desires of the participants. The basis of this semantics is similar to the semantic language defined for the *FIPA ACL*. Thus, the five *Fatio* locutions could readily be added to the 22 standard *FIPA ACL* locutions should an agent designer so wish. Such addition would help to ameliorate the impoverished argumentation theory of *FIPA ACL*. The *Fatio* protocol could also be used on a stand-alone basis, or as an addition to other agent interaction protocols.

It is well-known that the axiomatic semantics of the *FIPA ACL* is not verifiable in general [24]. To seek to ameliorate this, in other work we have defined the notion of a *contestability semantics* [13], in which claims made by agents in a dialogue are contestable by other participants, who can question or challenge the claims, and seek justifications for them. In this manner, agent claims may be assessed, for example, for correspondence to the truth; for consistency; for sincerity; and so on. Such assessments take place at run-time, in the dialogue itself, by the other participants as and when required, rather than being undertaken by the design teams in some conformance testing process before the dialogue commences. The *Fatio* protocol provides the means to undertake such run-time assessments, and therefore is an operationalization of our notion of contestability semantics.

**Acknowledgments** We are grateful to Iyad Rahwan and Mike Wooldridge for their comments.

## References


[1] L. Amgoud, S. Parsons, and N. Maudet. Arguments, dialogue, and negotiation. In W. Horn, editor, *Proc. ECAI 2000*, pages 338–342, Berlin, 2000. IOS Press.

[2] J. L. Austin. *How To Do Things with Words*. Oxford University Press, Oxford, 1962.

[3] M. Colombetti and M. Verdicchio. An analysis of agent speech acts as institutional actions. In *Proc. AAMAS 2002*, pages 1157–1164, New York, 2002. ACM Press.

[4] M. Dummett. What is a Theory of Meaning? (II). In *The Seas of Language*, pages 34–93. Clarendon, Oxford, 1993.

[5] R. van Eijk, F. S. de Boer, W. van der Hoek, and J-J. Ch. Meyer. A verification framework for agent communications. *J. Auton. Agents and Multi-Agent Systems*, 6:185–219, 2003.

[6] FIPA. Communicative Act Library Specification. Standard SC00037J, Foundation for Intelligent Physical Agents, 2002.

[7] J. Habermas. *The Theory of Communicative Action: Volume 1: Reason and the Rationalization of Society*. Heinemann, London, 1984. (Published in German 1981).

[8] J. Habermas. *On the Pragmatics of Communication*. MIT Press, Cambridge, MA, 1998. Edited by M. Cooke.

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

[10] P. McBurney, R. M. van Eijk, S. Parsons, and L. Amgoud. A dialogue-game protocol for agent purchase negotiations. *J. Auton. Agents & Multi-Agent Systems*, 7(3):235–273, 2003.

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

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

[13] P. McBurney and S. Parsons. Engineering democracy in open agent systems. In A. Omicini *et al.*, editor, *Engineering Societies in the Agents World*, LNAI, Berlin, 2004. Springer.

[14] P. McBurney, S. Parsons, and M. Wooldridge. Desiderata for agent argumentation protocols. In *Proc. AAMAS 2002*, pages 402–409, New York, 2002. ACM Press.

[15] S. Parsons, M. Wooldridge, and L. Amgoud. An analysis of formal interagent dialogues. In *Proc. AAMAS 2002*, pages 394–401, New York, 2002. ACM Press.

[16] S. Paurobally, J. Cunningham, and N. R. Jennings. Ensuring consistency in the joint beliefs of interacting agents. In *Proc. AAMAS 2003*, pages 662–669, New York, 2003. ACM Press.

[17] J. Pitt and A. Mamdani. Some remarks on the semantics of FIPA's Agent Communications Language. *J. Auton. Agents and Multi-Agent Systems*, 2:333–356, 1999.

[18] M. Schut and M. Wooldridge. The control of reasoning in resource-bounded agents. *Knowledge Engineering Review*, 16(3):215–240, 2001.

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

[20] A. Sen. *Rationality and Freedom*. Harvard University Press, Cambridge, MA, 2002.

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

[22] R. D. Tennent. *Semantics of Programming Languages*. Prentice-Hall, Hemel Hempstead, UK, 1991.

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

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

[25] C. Wright. Strict finitism. In *Realism, Meaning and Truth*, pages 107–175. Blackwell, Oxford, 1993.