Group Decision and Negotiation (2007) 16:101–141
DOI 10.1007/s10726-006-9059-1

A Convention-based Approach to Agent Communication Languages*

ANDREW J. I. JONES
Department of Computer Science, King's College London, The Strand, London, WC2R 2LS UK
(E-mail: andrewji.jones@kcl.ac.uk)

XAVIER PARENT
54 avenue de l'Elisa, 83100, Toulon, France
(E-mail: xp@up.univ-mrs.fr)

Abstract

This article aims to provide foundations for a new approach to Agent Communication Languages (ACLs). First, we present the theory of signalling acts. In contrast to current approaches to communication, this account is neither intention-based nor commitment-based, but convention-based. Next, we explore ways of embedding that theory within an account of conversation. We move here from an account of the basic types of communicative act (the statics of communication) to an account of their role in sequences of exchanges in communicative interaction (the dynamics of communication). Finally, we apply the framework to the analysis of conversational protocols such as the English auction protocol. We propose to give a compact expression of conversation protocols by means of a formula of the object-language. We also use this kind of representation to provide the basis for a procedure for keeping a record of the conventional effects achieved in a conversation. A corresponding axiomatic presentation is given, and shown to be sound and complete with respect to our proposed semantics.

Key words agent communication languages, speech acts, convention, conversational protocol, dynamic logic, arrow logic

1. Introduction

Current approaches to conversation can be divided into two basic categories:

- Those that are intention-based or mentalistic. Inspired by Grice (1957), these approaches focus on the effects communicative acts have on participants' mental states (see e.g. Labrou and Finin 1997; Smith et al. 1998);

* This paper is a revised and considerably extended version of Jones and Parent (2004). Some of the material included herein appeared in the document 'A Logical Framework', written by Andrew J. I. Jones, an unpublished chapter that formed part of the deliverables from the project ALFEBIITE (to which reference is made in the concluding acknowledgements).

- Those that are commitment-based, in that they assign a key role to the notion of commitment – see e.g. Colombetti (2000), Singh (1998) and Walton and Krabbe (1995).

What the relative merits are of intention-based and convention-based approaches to communication is a question that has been much debated within the Philosophy of Language (Bennett 1976; Grice 1957; Lewis 1969; Searle 1969). We cannot here enter into the details of this debate. Suffice it to say that it has become increasingly clear that the role played by the Gricean recognition-of-intention mechanism is not as important as one might think. Indeed, as far as literal speech acts are concerned, it is necessary to assume such a mechanism only for those cases where communicative acts are performed in the absence of established conventional rules. On the other hand, as some researchers working on Agent Communication Languages (ACLs) have also observed, the intention-based account takes for granted a rather controversial assumption, according to which agents' mental states are verifiable. This last observation is in fact one of the starting points of the commitment-based account as proposed by Singh (1998) and Colombetti (2000). However, there are also some strong reasons to believe that such models are fundamentally problematic. The most obvious reason has to do with the fact that it is not entirely clear what it means for speaker j to commit himself to an assertion of p. Should not the propositional content of a commitment be a future act of the speaker? If so, to what action is j preparing to commit himself, when asserting p? A natural reaction is to say that, in asserting p, speaker j in fact commits himself to defend p if p is challenged by k. This is the view defended by Walton and Krabbe (1995) and Brandom (1983, 1994). However, in line with Levi (1996), we believe that this defence does not stand up to close scrutiny. What counts as an assertion in a language–game may correlate very poorly with j's beliefs. For instance, j can say that p without being able to defend p.[1] Does that mean that j is not making an assertion? If so, what *is* he doing? As we shall see, to focus exclusively on agents' commitments amounts, ultimately, to confusing two kinds of norms, which have been called "preservative" and "constitutive". The first are the kind that control antecedently existing activities, e.g. traffic regulations, while the second are the kind that create or constitute the activity itself, e.g. the rules of the game.

Objections of these kinds, we believe, indicate the need for an account of signalling acts based not on *intentions*, or *commitments*, but on *public conventions*.

The plan of this paper is as follows. Section 2 presents the syntax and semantics of the multi-modal language upon which the theory of conventional signalling acts is based. Section 3 presents a detailed account of the latter theory. Section 4 outlines ways of embedding that theory within an account of conversation. We move here from an account of the basic types of communicative act (the statics of communication) to an account of their role in sequences of exchanges in communicative interaction (the dynamics of communication). The proposed framework is applied to the analysis of conversational protocols such as the English auction protocol. We propose to give a compact expression of conversation protocols by means of a formula of the object-language. We also use this kind of representation to provide the basis for a procedure for keeping a record of the conventional effects achieved in a conversation. In Section 5, finally, we show that the extended framework is sound and complete with respect to our proposed semantics.

2. A Multi-modal Framework

2.1. The basic components: axiomatics

In this section we first specify the syntax of a multi-modal logical language, and then proceed to axiomatic characterisations of the logics of several modal operators. These operators will serve as 'basic building blocks' in terms of which the statics of communication will be described in Section 3.

2.1.1. Syntax

The basic modal language Φ is founded upon a set of *atomic sentences*, usually denoted P_1, P_2, P_3 and so on. These are the simplest sentences. *Complex expressions* are formed from these by means of 15 syntactic operations

$$\top, \bot, \neg, \wedge, \vee, \rightarrow, \leftrightarrow, \Box, \Diamond, E_i, H_i, B_i, I_i, O, \text{ and } \Rightarrow_s .$$

The set of sentences of Φ is defined formally as follows:

1. P_n is a sentence, for $n = 0, 1, 2,...$
2. $\top$ is a sentence.
3. $\bot$ is a sentence.
4. $\neg A$ is a sentence, if A is a sentence.
5. $A \wedge B$, $A \vee B$, $A \rightarrow B$, $A \leftrightarrow B$, and $A \Rightarrow_s B$ are sentences if A and B are sentences.
6. $\Box A$, $\Diamond A$, E_iA, H_iA, B_iA, I_iA, and OA are sentences if A is a sentence.

The sentence $\Box A$ is the *necessitation* of A, and $\Diamond A$ is the *possibilitation* of A.[2] There is no standard terminology for sentences of the form E_iA, H_iA, B_iA, I_iA, OA. We will refer to them as e.g. *agentive sentences*, *doxastic sentences* and *normative sentences* depending on the interpretation of the relevant modal operator. A sentence of the form $A \Rightarrow_s B$ will be called a *counts as-conditional* of which A is the *condition* and B is the *consequence*.

2.1.2. Necessity

We read expressions of the form $\Box A$ as 'it is necessary that A'. The logic of the $\Box$-modality is that of the system traditionally known as S5, and is characterised by extending PL (classical propositional logic) with the following rule and axiom schemas:

$$(\Box RN) \quad \frac{A}{\Box A} \text{(i.e. if } \vdash \text{ A then } \vdash \Box A) \tag{1}$$

$$\Box K. \quad \Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B) \tag{2}$$

$$\Box T. \quad \Box A \rightarrow A \tag{3}$$

$$\Box 4. \quad \Box A \rightarrow \Box\Box A \tag{4}$$

$$\Box 5. \quad \neg\Box A \rightarrow \Box\neg\Box A \tag{5}$$

The dual of $\Box$ is defined as usual: $\Diamond A \stackrel{\text{def}}{=} \neg\Box\neg A$. We read expressions of the form $\Diamond A$ as 'it is possible that A'.

2.1.3. Action and attempted action

We first outline the logic for the concept of action based on the theory presented in (Elgesem 1997), which was in turn greatly influenced by the line of approach introduced in (Kanger 1972; Pörn 1970, 1977, 1989).

Let expressions of the type E_iA be read as 'agent i brings it about that A', or 'agent i sees to it that A'. In contrast to those logics of action which may be more familiar to computer scientists, this logic focuses exclusively on who the agent is, and on the state of affairs which the exercise of his agency produces, ignoring both temporal aspects and considerations of the means by which a particular state is brought about.

The logic for this modality is *classical* (cf. Chellas 1980) in the sense that it is closed under logical equivalence:

$$(E_iRE) \qquad \frac{A \leftrightarrow B}{E_iA \leftrightarrow E_iB} \tag{6}$$

Other characteristic principles are these:

$$E_iT. \qquad E_iA \rightarrow A \tag{7}$$

$$E_iC. \qquad (E_iA \wedge E_iB) \rightarrow E_i(A \wedge B) \tag{8}$$

$$E_i\neg N. \qquad \neg E_i\top. \tag{9}$$

Elgesem (1997) also shows how further praxeological concepts may be defined, including *opportunity*, *unavoidability* and *independence*, and he maps out the implication relation between the members of this family (Elgesem 1997, Sections 4.3–4.8). One notion which is not covered, however, is that of *attempting to see to it*. So we supplement Elgesem's framework by incorporating one of the modalities defined in Santos, Jones and Carmo (1997), reading expressions of the kind H_iA as 'the agent i attempts to see to it that A'. The logic of the modality H_i is also closed under the rule (RE):

$$(H_iRE) \qquad \frac{A \leftrightarrow B}{H_iA \leftrightarrow H_iB} \tag{10}$$

and contains the following axiom schemas:

$$H_iC. \qquad (H_iA \wedge H_iB) \rightarrow H_i(A \wedge B) \tag{11}$$

$$H_i\neg N. \qquad \neg H_i\top \tag{12}$$

$$E_iH_jT \rightarrow . \qquad E_iA \rightarrow H_iA \tag{13}$$

$$E_iE_jH_j \rightarrow . \qquad E_iE_jA \rightarrow E_iH_jA \tag{14}$$

$$H_iE_jH_j \rightarrow . \qquad H_iE_jA \rightarrow H_iH_jA. \tag{15}$$

Note that we do not adopt the schema $E_iE_j \rightarrow .$:

$$E_iE_j \rightarrow . \qquad E_iE_jA \rightarrow E_iA$$

since it is commonly the case that we would not want to say that *i himself* brings it about that A, even though i does bring it about that j brings it about that A [for further discussion of the distinction between direct and indirect action, see Santos et al. (1997)].

2.1.4. Belief

We read expressions of the form B_iA as 'agent i believes that A'. The belief modality will be classical in the sense that it is closed under logical equivalence

$$(B_iRE) \qquad \frac{A \leftrightarrow B}{B_iA \leftrightarrow B_iB} \tag{16}$$

We also adopt the following axiom schemas:

$$B_iC. \qquad (B_iA \wedge B_iB) \rightarrow B_i(A \wedge B) \tag{17}$$

$$B_i\neg F. \qquad \neg B_i\bot \tag{18}$$

from which we may in turn derive

$$B_iD. \qquad \neg(B_iA \wedge B_i\neg A) \tag{19}$$

which represents a consistency constraint on agents' beliefs. We also adopt

$$B_iM. \qquad B_i(A \wedge B) \rightarrow (B_iA \wedge B_iB). \tag{20}$$

As is readily shown, $B_iM.$ together with (B_iRE) yields the result that i believes all logical consequences of that which he believes, and thus, also, that if i has any beliefs at all then i believes all tautologies.

From (B_iRE), $B_iC.$ and $B_iM.$ the schema $B_iK.$ follows:

$$B_iK. \qquad (B_iA \wedge B_i(A \rightarrow B)) \rightarrow B_iB. \tag{21}$$

In certain contexts we may wish to adopt one or both of the positive and negative introspection axiom schemas:

$$B_i4. \qquad B_iA \rightarrow B_iB_iA \tag{22}$$

$$B_i5. \qquad \neg B_iA \rightarrow B_i\neg B_iA \tag{23}$$

but we do not build these principles into the basic belief logic.

2.1.5. Normative modalities
The characterization of communicative acts will call for the introduction of two distinct normative modalities.[3] The *directive* modality, designated by O, will be used to specify agents' obligations: what they are required to do and not to do. The *evaluative* modality, designated by I_i, is a relativised modality specifying that which, for a given agent i, is ideal, or optimal.[4]

The problems of standard deontic logic (SDL) (Hilpinen and Follesdal 1971), which is a normal modal system of type KD, are well documented. [For a new overview, see Carmo and Jones (2002).]. It seems clear that some, but by no means all, of these problems derive from the strong closure rule (RK) which is a defining characteristic of normal modal systems. For this reason in particular, we opt for classical, non-normal logics for the two normative modalities. Nevertheless there is also a very good reason to believe that merely switching to non-normal modal logic will not itself provide a solution to the notorious problems associated with deontic conditional sentences – specifically, problems about Contrary-to-Duty conditionals (CTDs), and problems about the defeasibility of conditional obligation. However, in regard to CTDs, an elaborate formal treatment is available (Carmo and Jones 2002), and the dyadic obligation operator (and its associated necessity operators) defined there can replace the present O-operator if we are confronted with scenarios in which CTDs play a critical role. And as regards the issue of defeasible conditional obligations, on our view this (unlike the CTD problem) is not a problem specific to *deontic* logic, but rather an instance of the more general task of devising:

1. An appropriate logical analysis of default (exception-allowing) conditionals, and
2. An associated default reasoning mechanism.[5]

Since they are classical modalities, both O and I_i are closed under logical equivalence:

$$(ORE) \qquad \frac{A \leftrightarrow B}{OA \leftrightarrow OB} \tag{24}$$

$$(I_iRE) \qquad \frac{A \leftrightarrow B}{I_iA \leftrightarrow I_iB}. \tag{25}$$

In addition, we adopt the following schemas: (assume that $\Diamond$ is the dual of $\Box$ and that $\Box$ is a KT5 (i.e. S5) modality)

$$OC\Diamond \qquad (OA \wedge OB \wedge \Diamond(A \wedge B)) \rightarrow O(A \wedge B) \tag{26}$$

$$I_iC\Diamond \qquad (I_iA \wedge I_iB \wedge \Diamond(A \wedge B)) \rightarrow I_i(A \wedge B) \tag{27}$$

$$O\neg N. \qquad \neg O\top \tag{28}$$

$$O\neg F \qquad \neg O\bot \tag{29}$$

$$I_i\neg F \qquad \neg I_i\bot \tag{30}$$

Following Kanger (1972) and Pörn (1977) we also adopt the following bridging principle linking the two modalities

$$OI_i \rightarrow \qquad I_iOA \rightarrow I_iA \tag{31}$$

i.e., if it is optimal for i that A is obligatory, then it is optimal for i that A.

2.1.6. 'Counts as'

Jones and Sergot (1996) give a formal representation of situations in which an agent has an *institutionalised power* to bring about a certain state of affairs. They discuss a particular kind of conditional relation called *counts as*, which is used to represent the idea that, for instance, within a given institution (e.g., a given church), the performance of a particular kind of act by a designated agent (e.g., a priest conducting a marriage ceremony), perhaps in a particular context (witnesses must be present), counts as a way of establishing a particular institutional fact (that two agents are married). To capture this notion of *counts as* formally, they introduce a relativised conditional connective $\Rightarrow_s$. Sentences of the form $A \Rightarrow_s B$ are intended to express the idea that, according to institution s, when the state of affairs described by A obtains, that counts as a means of establishing that the state of affairs described by B also obtains. Very often, although not always, the antecedent and consequent will be act-descriptions, as when one says that, according to a legal system s, if the registrar of marriages utters a particular form of words in a particular context, this counts as a means whereby s sees to it that two agents attain the status of 'married couple'. The conditional $A \Rightarrow_s B$ describes an institutionalised means-end relationship: the registrar, acting as an empowered agent for, or on behalf of, s, fulfils the antecedent condition which is sufficient to guarantee that the institution s then classifies the given couple as 'married'. Clearly, the conditional connective needs to be relativised to the institution concerned, for what counts for one institution as a means of creating a state of affairs may well not so count for some other institution.

Among other uses, we shall apply the Jones and Sergot analysis of institutionalised power to the description of signalling systems, since the rules or conventions which define a signalling system may also be seen as specifying institutionalised means-end relationships. They say, essentially, that the performance of a particular kind of act counts (as far as that signalling system is concerned) as a means of indicating that a particular state of affairs obtains. First we give an outline of the axiomatic characterisation of the new connective.

The logic of $\Rightarrow_s$ is that of a classical conditional logic [in the sense of (Chellas 1980, chapter 10)]. That is, it is closed under the following two rules:

$$\Rightarrow_s RCEC \qquad \frac{A \leftrightarrow A'}{(B \Rightarrow_s A) \leftrightarrow (B \Rightarrow_s A')} \tag{32}$$

$$\Rightarrow_s RCEA \qquad \frac{A \leftrightarrow A'}{(A \Rightarrow_s B) \leftrightarrow (A' \Rightarrow_s B)}. \tag{33}$$

In addition, the following axiom schemas are adopted:

$$\Rightarrow_s CC. \qquad ((A \Rightarrow_s B) \wedge (A \Rightarrow_s C)) \rightarrow (A \Rightarrow_s (B \wedge C)) \tag{34}$$

$$\Rightarrow_s CA. \qquad ((A \Rightarrow_s B) \wedge (C \Rightarrow_s B)) \rightarrow ((A \vee C) \Rightarrow_s B) \tag{35}$$

$$\Rightarrow_s S. \qquad (A \Rightarrow_s B) \rightarrow ((B \Rightarrow_s C) \rightarrow (A \Rightarrow_s C)). \tag{36}$$

For more detailed discussion of this system see Jones and Sergot (1996). But note at least the following two features: first, this is not a normal conditional logic, so the consequent of a $\Rightarrow_s$ conditional is not closed under classical logical consequence. This is as it should be, for the kinds of application we are proposing here: suppose for instance that the act of raising one yellow flag on a ship indicates, according to signalling system s, that the ship is carrying explosives. Would it not then seem odd to be able to draw the further conclusion that the same signal indicates either that the ship is carrying explosives or that the ship is carrying injured passengers? Second, the conditional does not have the strengthening of the antecedent property:

$$\Rightarrow_s SA. \qquad (A \Rightarrow_s B) \rightarrow ((A \wedge C) \Rightarrow_s B). \tag{37}$$

The failure of $\Rightarrow_s SA.$ reflects the fact that *counts as* conditionals are defeasible: circumstances may arise in which, even though the antecedent A is fulfilled, this nevertheless does not count as establishing the truth of the consequent. In the case of exercise of institutionalised power, examples of defeasibility abound (the registrar carries out the ceremony, but does so under duress). Likewise, in the case of signalling one may find circumstances in which the performance of the signalling act does not count.

2.2. The basic components: semantics

We here describe a semantical framework for the multi-modal logical language Φ presented in subsection 2.1, and in terms of it we give truth-conditions for the various kinds of modal sentences that are expressible in that language.

2.2.1. Model structure

Consider a minimal model structure

$$M = \langle W, f_n, f_{c_i}, f_{h_i}, f_o, f_{i_i}, f_{b_i}, f_{\Rightarrow_s}, P \rangle$$

W is a set of possible worlds, and P is a valuation function which assigns to each atomic sentence in the language Φ a set of possible worlds. Intuitively, P is thought of as returning, for each atomic sentence P_n, the set of worlds in M at which P_n is true.

Where A is any sentence of Φ, the truth-set of A in a model M is the set of worlds in M at which A is true, and is denoted by $\| A \|^M$. So, we adopt the following definition, where α is any world in W in M:

$$\| A \|^M =_{df} \{\alpha : M, \alpha \models A\}. \tag{38}$$

Thus, where P_n is any atomic sentence:

$$M, \alpha \models P_n \text{ iff } \alpha \in P(P_n). \tag{39}$$

And the truth-sets for the sentences $\top$ and $\bot$ are specified by:

$$\| \top \|^M = W \tag{40}$$

$$\| \bot \|^M = \emptyset \tag{41}$$

Where A and B are any sentences in Φ,

$$M, \alpha \models \neg A \text{ iff } \alpha \notin \| A \|^M \tag{42}$$

$$M, \alpha \models A \wedge B \text{ iff } \alpha \in (\| A \|^M \cap \| B \|^M) \tag{43}$$

$$M, \alpha \models A \vee B \text{ iff } \alpha \in (\| A \|^M \cup \| B \|^M) \tag{44}$$

$$\begin{gathered} M, \alpha \models A \rightarrow B \text{ iff } \alpha \in (- \| A \|^M \cup \| B \|^M) \\ \text{where } - \| A \|^M =_{df} W - \| A \|^M \end{gathered} \tag{45}$$

$$\begin{gathered} M, \alpha \models A \leftrightarrow B \text{ iff } \alpha \in (- \| A \|^M \cup \| B \|^M) \text{ and} \\ \alpha \in (- \| B \|^M \cup \| A \|^M). \end{gathered} \tag{46}$$

A proposition is a set of possible worlds. Accordingly, the proposition expressed by a sentence A is the set of possible worlds at which A is true (the truth-set of A). The members f_n, f_{h_i}, f_o, f_{i_i} and f_{b_i} of M are unary functions each of which assigns, to each world in W, a set of propositions, i.e., a set of sets of possible worlds. The binary functions $f_{\Rightarrow_s}$ and f_{c_i} each assign a set of propositions to each proposition at each world. The subscripts to the functions f_{c_i}, f_{h_i}, f_{i_i} and f_{b_i} indicate their relativisation to individual agents, and – as was explained in Section 2.1.6 – the function $f_{\Rightarrow_s}$ is relativised to institutions. The functions f_n, f_{c_i}, f_{h_i}, f_o, f_{i_i} and f_{b_i} will be employed in the specifications of the truth conditions for (respectively) modal sentences of type necessity, action, attempted action, directive norm, evaluative norm and belief.

But before proceeding to those specifications, it will be useful to list and name a number of constraints to which the unary functions might be subjected. Let $\underline{f}$ be any one of the unary functions in M, let X and Y be any subsets of W, and let α and β be any members of W:

$(m\underline{f})$ if $X \cap Y \in \underline{f}(\alpha)$ then $X \in \underline{f}(\alpha)$ and $Y \in \underline{f}(\alpha)$

$(c\underline{f})$ if $X \in \underline{f}(\alpha)$ and $Y \in \underline{f}(\alpha)$ then $X \cap Y \in \underline{f}(\alpha)$

$(cp\underline{f})$ if $X \in \underline{f}(\alpha)$ and $Y \in \underline{f}(\alpha)$ and $X \cap Y \neq \emptyset$ then $X \cap Y \in \underline{f}(\alpha)$

$(n\underline{f})$ $W \in \underline{f}(\alpha)$

$(nn\underline{f})$ $W \notin \underline{f}(\alpha)$

$(n\emptyset\underline{f})$ $\emptyset \notin \underline{f}(\alpha)$

$(d\underline{f})$ if $X \in \underline{f}(\alpha)$ then $-X \notin \underline{f}(\alpha)$

$(t\underline{f})$ if $X \in \underline{f}(\alpha)$ then $\alpha \in X$

$(4\underline{f})$ if $X \in \underline{f}(\alpha)$ then $\{\beta \in M : X \in \underline{f}(\beta)\} \in \underline{f}(\alpha)$

$(5\underline{f})$ if $X \notin \underline{f}(\alpha)$ then $\{\beta \in M : X \notin \underline{f}(\beta)\} \in \underline{f}(\alpha)$.

We may illustrate the way in which this list will be used, as follows: suppose that the function f_n is subject to the condition/constraint $(t\underline{f})$. Then we shall say that (tf_n) holds, meaning by that, where X is any subset of W, and α any member of W, if $X \in f_n(\alpha)$ then $\alpha \in X$.

Similarly, suppose that the function f_{b_i} is subject to the condition/constraint $(c\underline{f})$. Then we shall say that (cf_{b_i}) holds, meaning by this that, where i is any agent, X and Y any subsets of W, and α any member of W, if $X \in f_{b_i}(\alpha)$ and $Y \in f_{b_i}(\alpha)$ then $X \cap Y \in f_{b_i}(\alpha)$.

2.2.2. The necessity operator

From the intuitive point of view, the function f_n is thought of as selecting, for each world α, the set of propositions that are necessary relative to α. Truth of a sentence of form $\Box A$ at a world α in a model M is specified as follows:

$$M, \alpha \models \Box A \text{ iff } \| A \|^M \in f_n(\alpha). \tag{47}$$

Sentences of the form $\Box A$ are read 'it is necessary that A'. Sentences of the form $\Diamond A$ are read 'it is possible that A', where $\Diamond A =_{df} \neg\Box\neg A$.

For the function f_n we adopt the following constraints: (mf_n), (cf_n), (nf_n), (tf_n), ($4f_n$) and ($5f_n$). In short, the $\Box$-operator expresses necessity of type S5.

2.2.3. The praxeological operators

Where i is any agent, α is any world in W, and A any sentence, the value of $f_{c_i}(\alpha, \| A \|^M)$ is understood intuitively to be the set of worlds where the agent i realises the ability he has in α to bring about the state of affairs described by A. [This approach derives from the work of Elgesem (1997).] Accordingly, we fix truth conditions for sentences of the form E_iA ('i sees to it that/brings it about that A') as follows:

$$M, \alpha \models E_iA \text{ iff } \alpha \in f_{c_i}(\alpha, \| A \|^M). \tag{48}$$

The validity of the closure rule (6) immediately follows. The validity of the 'success' condition (7) is secured by adopting the constraint:

$$f_{c_i}(\alpha, X) \subseteq X \tag{49}$$

(where X, as before, is any subset of W, α is any member of W, and i is any agent).

The validity of (9) is secured by adopting the constraint:

$$f_{c_i}(\alpha, W) = \emptyset. \tag{50}$$

Finally, the validity of the C-schema for the E_i-operator (the schema (8)) is secured by adopting the constraint:

$$\beta \in f_{c_i}(\alpha, X) \text{ and } \beta \in f_{c_i}(\alpha, Y) \text{ then } \beta \in f_{c_i}(\alpha, X \cap Y) \tag{51}$$

Where i is any agent, the function f_{h_i} selects, for each world α, the set of propositions corresponding to the states of affairs i attempts to bring about at α. Accordingly, we fix truth conditions for sentences of the form H_iA ('agent i attempts to see to it that A') as follows:

$$M, \alpha \models H_iA \text{ iff } \| A \|^M \in f_{h_i}(\alpha). \tag{52}$$

The validity of the closure rule (10) is now secured. To guarantee the validity of (11) we adopt the constraint (cf_{h_i}) – see Section 2.2.1, above. And the validity of (12) is secured by adopting the constraint (nnf_{h_i}).

The validity of schema (13), which expresses the assumption that seeing to it that A implies attempting to see to it that A, is captured by imposing the constraint:

$$\text{if } \alpha \in f_{c_i}(\alpha, \| A \|^M) \text{ then } \| A \|^M \in f_{h_i}(\alpha). \tag{53}$$

Finally, adoption of the following two constraints secures the validity of (respectively) the two bridge principles (14) and (15), where i and j are any agents:

$$\begin{aligned}&\text{if } \alpha \in f_{c_i}(\alpha, \{\beta : \beta \in f_{c_j}(\beta, \| A \|^M)\})\\ &\qquad \text{then } \alpha \in f_{c_i}(\alpha, \{\beta : \| A \|^M \in f_{h_j}(\beta)\})\end{aligned} \tag{54}$$

$$\begin{aligned}&\text{if } \{\beta : \beta \in f_{c_j}(\beta, \| A \|^M)\} \in f_{h_i}(\alpha)\\ &\qquad \text{then } \{\beta : \| A \|^M \in f_{h_j}(\beta)\} \in f_{h_i}(\alpha).\end{aligned} \tag{55}$$

2.2.4. The doxastic operator

Where i is any agent, the function f_{b_i} picks out, for each world α, the set of propositions believed by i at α. So we adopt the following truth condition for sentences of the form B_iA:

$$M, \alpha \models B_iA \text{ iff } \| A \|^M \in f_{b_i}(\alpha).$$

The validity of the closure rule (16) is thereby secured.

We also adopt the following constraints on the function f_{b_i} : $(mf_{b_i}), (cf_{b_i}), (n\emptyset f_{b_i})$, thereby validating (respectively) schemas (20), (17) and (18).

The positive and negative introspection schemas (22) and (23) would be validated by the adoption of (respectively) constraints $(4f_{b_i})$ and $(5f_{b_i})$. However, we do not build these constraints into the semantical characterisation of the belief modality.

2.2.5. The normative operators

The function f_o picks out, for each world α, the set of propositions that correspond to that which is obligatory at α. (Typically, these propositions will be those expressed by some set of act descriptions, relativised to agents, since that which is obligatory – typically – is the performance of some particular acts.)

We adopt the following truth condition:

$$M, \alpha \models OA \text{ iff } \| A \|^M \in f_o(\alpha) \tag{56}$$

reading sentences of the form OA as 'it is obligatory/normatively required that A'. The validity of the closure rule (24) is now secured. We adopt the constraints (cpf_o), (nnf_o) and ($n\emptyset f_o$) to grant validity to (26), (28) and (29), respectively.

Where i is any agent, the function f_{i_i} picks out, for each world α, the set of propositions corresponding to that which is ideal, or optimal, for i at α. So we adopt the truth condition:

$$M, \alpha \models I_iA \text{ iff } \| A \|^M \in f_{i_i}(\alpha). \tag{57}$$

This guarantees the validity of the closure rule (25). The adoption of constraints (cpf_{i_i}) and $(n\emptyset f_{i_i})$ now also secures the validity of (27) and (30), respectively.

The bridging principle between the two normative modalities, expressed by (31), is validated by adopting the following constraint:

$$\text{if } \{\beta : || A ||^M \in f_o(\beta)\} \in f_{i_i}(\alpha) \text{ then } || A ||^M \in f_{i_i}(\alpha). \quad (58)$$

2.2.6. The 'counts-as' operator

The reader is referred to Jones and Sergot's 1996 paper on the formal characterisation of institutionalised power (Jones and Sergot 1996) for a detailed account motivating the introduction of the 'count as' operator. In this section, we simply describe the relevant truth condition and associated constraints on the function $f_{\Rightarrow_s}$.

Where s is an institution, α is a world in a model M, and $|| A ||^M$ is the proposition expressed by sentence A in M, the function $f_{\Rightarrow_s}$ picks out, for the pair $(\alpha, || A ||^M)$, a set of propositions: intuitively, the set of propositions that, from the perspective of institution s, the truth of A at α counts as a means of establishing as true – as when, for instance, the fact that, at α, John is under the age of 14 years counts, for s, as a means of establishing the truth of the proposition that John is a child. Viewed in this way, the 'count as' notion is expressible as a conditional, whose truth conditions we specify as follows:

$$M, \alpha \models A \Rightarrow_s B \text{ iff } || B ||^M \in f_{\Rightarrow_s}(\alpha, || A ||^M). \quad (59)$$

This truth condition suffices to validate the two closure rules (32) and (33). Furthermore, we adopt the following three constraints, in order to validate the schemas (34), (35) and (36), respectively:

$$\text{if } Y \in f_{\Rightarrow_s}(\alpha, X) \text{ and } Z \in f_{\Rightarrow_s}(\alpha, X) \text{ then } Y \cap Z \in f_{\Rightarrow_s}(\alpha, X) \quad (60)$$

$$\text{if } X \in f_{\Rightarrow_s}(\alpha, Y) \text{ and } X \in f_{\Rightarrow_s}(\alpha, Z) \text{ then } X \in f_{\Rightarrow_s}(\alpha, Y \cup Z) \quad (61)$$

$$\text{if } Y \in f_{\Rightarrow_s}(\alpha, X) \text{ and } Z \in f_{\Rightarrow_s}(\alpha, Y) \text{ then } Z \in f_{\Rightarrow_s}(\alpha, X). \quad (62)$$

3. Conventional Signalling Acts

This section is principally concerned with the characterisation of some fundamental types of signalling acts, and aims to provide foundations for a new approach to Agent Communication Languages (ACLs). A principal feature of the approach taken here, in contrast to those which dominate much current work on ACLs, is that communicators' intentions, and in particular the effects they intend to achieve in their audience, will not be assigned a central role. Rather, the focus will be on the public conventions whose existence makes possible the performance of intentional acts of communication. A close look, first, at the communicative act of asserting will serve as a means of presenting the basic assumptions and intuitions which guide this approach.

3.1. Indicative signalling systems

An indicative signalling system is a signalling system in which acts of asserting can be performed. It is constituted by conventions which grant that the performance, in particular circumstances, of instances of a given class of act-types *count as* assertions, and which also specify what the assertions mean. For example, the utterance with a particular intonation pattern of a token of the sentence "The ship is carrying explosives" will count, in an ordinary communication situation, as an assertion that the ship is carrying explosives. The raising, on board the ship, of a specific sequence of flags, will also count as an assertion that the ship is carrying explosives. In the first case the signal takes the form of a linguistic utterance, and in the second it takes the form of an act of showing flags. These are just two of a number of different types of media employed in signalling systems. For present purposes, it is irrelevant which medium of communication is employed. But for both of these signalling systems there are conventions determining that particular acts count as assertions with particular meanings.

According to Searle (1969), if the performance by agent j of a given communicative act counts as an assertion of the truth of A, then j's performance *counts as an undertaking to the effect that A is true*. What lies behind that claim, surely, is that when j asserts that A what he says *ought* to be true, in some sense or other of 'ought'. The problem is to specify what sense of 'ought' this is. [cf. Stenius (1967).] The view adopted here is that the relevant sense of 'ought' pertains to the specification of the conditions under which an indicative signalling system is in an optimal state: given that the prime function of an indicative signalling system is to facilitate the transmission of reliable information, the system is in a less than optimal state, relative to that function, when a false signal is transmitted. The relevant sense of 'ought' is like that employed in "The meat ought to be ready by now, since it has been in the oven for 90 minutes". The system, in this case the oven with meat in it, is in a sub-optimal state if the meat is not ready – things are not then as they ought to be, something has gone wrong. The fact that the principles on which the functioning of the oven depends are physical laws, whereas the principles on which the signalling system depends are man-made conventions, is beside the point: in both cases the optimal functioning of the system will be defined relative to the main purpose the system is meant to achieve, and thus in both cases failure to satisfy the main purpose will represent a less-than-optimal situation.

Suppose that agents j and k are users of an indicative signalling system s, and that they are mutually aware that, according to the signalling conventions governing s, the performance by one of them of the act of seeing to it that C is meant to indicate that the state of affairs described by A obtains. The question of just what kind of act 'seeing to it that C' is will be left quite open. All that matters is that, by convention (in s), seeing to it that C counts as a means of indicating that A obtains. The content of the convention which specifies the meaning, in s, of j's seeing to it that C will be expressed using a 'counts as' conditional relativised to s, with the sentence E_jC as its antecedent, where E_jC is read 'j sees to it that C' or 'j brings it about that C'. How, then, is the form of the consequent to be represented ? The communicative act is an act of asserting that A, and thus counts as an undertaking to the effect that the state of affairs described by A obtains. As proposed in the previous paragraph, this is interpreted as meaning that, when the communicative act E_jC is performed, s's being in an optimal state would require that the sentence A be true. So the form of the signalling convention according to which, in s, j's seeing to it that C counts as an undertaking to the effect that A, is given by

$$\text{(sc-assert)} \qquad E_jC \Rightarrow_s I_s^\star A \tag{63}$$

where $I_s^\star$ is a relativised optimality, or ideality, operator.[6] $I_s^\star A$ expresses the proposition that, were s to be in an optimal state relative to the function s is meant to fulfil, A would have to be true, and $\Rightarrow_s$ is the relativised 'counts as' conditional. Simplifying, we can say that (63) expresses the following: by the conventions constituting signalling system s, if j brings it about that C, then A ought to be true.

We state informally some assumptions we associate with (sc-assert). First, signalling system s is likely to contain a number of other conventions of the same form, according to which j's seeing to it that C' counts as an undertaking to the effect that A', j's seeing to it that C'' counts as an undertaking to the effect that A'', ... and so on. So the conventions expressed by conditionals of form (sc-assert) may be said to contain the *code* associated with indicative signalling system s – the code that shows what particular kinds of assertive signalling acts in s are meant to indicate. We might then also say that s itself is *constituted* by this code. Second, we assume that the (sc-assert) conditionals constituting s hold true for *any* agent j in the group U of agents who use s; that is, each agent in U may play the role of communicator. Third, we assume that the (sc-assert) conditionals associated with s are all mutually believed by the members of U. We do not here state in full what 'mutual belief' amounts to, except to say that we take it to include at least the following: where X is a conditional of form (sc-assert) associated with s, and j and k are any pair of members of the group U that use s, then

$$\begin{array}{l} B_jX \\ B_kX \\ B_jB_kX \\ B_kB_jX \\ B_kB_jB_kX \\ B_jB_kB_jX \\ \vdots \end{array}$$

and so on, to some suitably high level of iteration.[7]

3.2. *The logic of the modality $I_s^\star$*

For the logic of this particular optimality/ideality operator, we adopt a (relativised) classical modal system of type EMCN. [As is shown in Chellas (1980) – see chapter 8 – a classical system of this type is identical to the smallest normal system K.] So, the logic contains the rule of closure under logical equivalence ($I_s^\star RE$) and, in addition, the schemas $I_s^\star M.$ and $I_s^\star C.$, and the sentence $I_s^\star N.$:

$$(I_s^\star RE) \quad \frac{A \leftrightarrow B}{I_s^\star A \leftrightarrow I_s^\star B} \tag{64}$$

$$I_s^\star M. \qquad I_s^\star(A \wedge B) \rightarrow (I_s^\star A \wedge I_s^\star B) \tag{65}$$

$$I_s^\star C. \qquad (I_s^\star A \wedge I_s^\star B) \rightarrow I_s^\star(A \wedge B) \tag{66}$$

$$I_s^\star N. \qquad I_s^\star \top. \tag{67}$$

The following rule now also holds:

$$(I_s^\star RM) \quad \frac{A \rightarrow B}{I_s^\star A \rightarrow I_s^\star B} \tag{68}$$

This is reasonable for the intended interpretation of $I_s^\star$, given the assumption that the prime function of an indicative signalling system is to facilitate the transmission of reliable information. For if the optimal state of such a system were to require the truth of A, then it would surely also require the truth of any logical consequence of A. Furthermore, the adoption of $I_s^\star N.$ amounts to the assumption that the optimal state of an indicative signalling system requires the truth of all tautologies.

Note that the D. schema is *not* contained in the logic of $I_s^\star$:

$$I_s^\star D. \qquad \neg(I_s^\star A \wedge I_s^\star \neg A) \tag{69}$$

Thereby we leave open the (not infrequently realised) possibility that one or more agents might perform indicative signalling acts the meaning-contents of which are mutually inconsistent. As Chellas (1980), chapter 6, observes, no distinction can be drawn in a normal modal system between the D. schema and the sentence P. – these are logically equivalent in normal systems. For the modality $I_s^\star$, the sentence P. is

$$I_s^\star P. \qquad \neg I_s^\star \bot \tag{70}$$

So, since $I_s^\star P.$ is not contained in our logic, we also leave open the possibility that an agent might make a single assertion the content of which is an explicit contradiction. However, the logic of belief we adopt secures the result that no agent could ever accept that such an assertion was true.

From the model-theoretic point of view, the logic of the modality $I_s^\star$ may be characterised in terms of minimal models, where the basic truth condition for sentences of the form $I_s^\star A$ is given by

$$M, \alpha \models I_s^\star A \text{ iff } \| A \|^M \in f_s^\star(\alpha)$$

The function $f_s^\star$ will also be subject to the following constraints (see subsection 2.2.1): $(mf_s^\star)$, $(cf_s^\star)$, $(nf_s^\star)$.

3.3. Communicator and audience

Suppose that j and k are both users of signalling system s, and that (sc-assert) is any of the signalling conventions in s. Then we adopt the following schema:

$$((E_j C \Rightarrow_s I_s^\star A) \wedge B_k E_j C) \rightarrow B_k I_s^\star A. \tag{71}$$

The import of the schema is essentially this: if k (the audience) believes that j performs the communicative act specified in the antecedent of (sc-assert), then k will accept that the consequent of (sc-assert) holds. He believes, then, that were signalling system s to be in an optimal state, A would be true. Another way of expressing the main point here is as follows: since k is familiar with the signalling conventions governing s, he is aware of what j's doing C is meant to indicate, and so, when k believes that j has performed this act, k is also aware of what would then have to be the case if the reliability of j's assertion could be *trusted*. This is not of course to say that k will necessarily trust j's reliability, but *if* he does so he will then also go on to form the belief that A. In summary, assuming (sc-assert) and (71), and supposing that

$$B_k E_j C \tag{72}$$

it now follows that

$$B_k I_s^\star A. \tag{73}$$

If k now also *trusts the reliability of* j's *assertion*, k goes on to form the belief

$$B_k A. \tag{74}$$

This type of trust is to be distinguished from 'trust-in-sincerity'. For we may say that, in this same communication situation, *if* k also *trusts the sincerity of* j's *assertion*, k goes on to form the belief:

$$B_k B_j A. \tag{75}$$

Note the various possibilities here: k might trust neither the reliability nor the sincerity of j's assertion, in which case neither (74) nor (75) holds. Alternatively, k might trust j's sincerity without trusting the reliability of his assertion [(75), but not (74)], or k might trust the reliability of j's assertion without trusting j's sincerity [(74) but not (75)]. The latter case may arise if, for instance, k believes that the source of information supplying j is indeed reliable, even though he (k) also believes that j does not think the source is reliable. Finally, of course, k might trust both the reliability and the sincerity of j's assertion.

Note, furthermore, that the set of four *trust positions* we have just indicated may be expanded into a larger set of positions, depending on whether or not j is *in fact* reliable and *in fact* sincere.[8]

It can readily be seen that, in contrast to the approach advocated in the FIPA COMMUNICATIVE ACT LIBRARY SPECIFICATION [SC00037J, 2002-12-06][9], the present account of asserting makes no assumptions about the sincerity of the communicator. Furthermore, there is no assumption to the effect that j, when performing the act E_jC, intends thereby to produce in k one or both of the beliefs (74) and (75). Indeed the only background assumption about the communicator's intention that is implicit in this account is that k, when forming the belief represented by (73), supposes that j's communicative act is to be taken as a serious, *literal* implementation of the governing convention (sc-assert); i.e., k does not think that j is play-acting, communicating ironically, talking in his sleep, etc. In such *non*-literal communication situations there are good reasons (which will not be developed here) for supposing that (71) does not hold for a rational audience k. One distinctive feature of the present approach is that this background assumption about the communicator's intention can *remain* implicit, since the mechanism by means of which assertoric signalling is effected turns essentially on the governing signalling conventions – the publicly accessible rules which show what particular types of communicative acts are taken to indicate – rather than on the intentions of agents who employ those conventions.[10]

It might also be observed that it is very natural indeed to adopt this background assumption in the contexts for which the theory of ACLs is currently being developed. For the primary interest there is certainly not in *non*-literal communication, or in 'communicating one thing but meaning another', but in the *literal* (albeit quite possibly *deceitful*) usage of signals with *public*, *conventional meanings*.

3.4. Commitment

Some recent approaches to ACLs have assigned a key role to the notion of *commitment* [e.g., Singh (1998) and Colombetti (2000)], and it might be suggested that when an agent j asserts that A, his act counts as an *undertaking* to the effect that A is true in the sense that *j commits* himself to the truth of A. So it might be supposed that there is here an alternative way of understanding the essential rule governing asserting to that offered above in terms of the I_s^* operator.

However, this suggestion raises a number of difficulties. First, just what is meant by saying that an agent commits himself to the truth of some sentence A? Does it mean that j is under some kind of obligation to accept that A is true? If so, in relation to which other agents is this obligation held, i.e., who is it that requires of j that j shall accept the truth of A? Everyone to whom he addresses his assertion? Surely not, for there may well be members of the audience who do not care whether j is being sincere, and there may also be others who require j to be insincere: perhaps j is their designated 'spokesman' whom they have instructed to engage in deception when that strategy appears to meet their interests. Furthermore, since the current concern with

ACLs is related to the design of *electronic* agents, it has to be said that there is very little agreement on what it might mean for an electronic agent to enter into a commitment.

The view taken here is that the move towards agent *commitment* (as the basis for understanding the *undertaking* involved in an act of asserting) is the result of a confusion – a confusion which was already indicated by Føllesdal (1967) in his discussion of Stenius. The point is this: the reason why it is very commonly required of communicators that they shall tell the truth, or at least attempt to tell the truth as they see it, is that conformity to that requirement (that norm) will help to *preserve* the practice of asserting *qua* practice whose prime function is to facilitate the transmission of reliable information. But norms designed to *preserve* the practice should not be confused with the rules or conventions which themselves *constitute* the practice – the conventions whose very existence makes possible the game of asserting, and which determine that the performance of an instance of a given act-type counts as a means of saying that such-and-such a state of affairs obtains. An attempt to use the notion of communicator's commitment to characterise the nature of asserting confuses preservative norms with constitutive conventions. To be sure, those conventions will eventually become de-valued, relative to the function they were designed to meet, if there is continual violation of the preservative norms. But this should not be allowed to obscure the fact that it is the conventions, and not the preservative norms, that create the very possibility of playing the asserting game, in an honest way, or deceitfully.

In a reply to our criticism of the commitment-based approach, Verdicchio and Colombetti (2004) claim that they do not intend to interpret the term 'commitment' as equivalent to 'obligation'. "Committing to the truth of a sentence, *s*," they say, "simply means that the debtor of the commitment will be in a state of fulfillment if *s* is settled true, in a state of violation if *s* is settled false, and in a pending state if the truth value of *s* is still undetermined." (Verdicchio and Colombetti 2004, p. 141) At first this appears difficult to comprehend, since the language of 'debtor' and 'creditor' that they use is strongly suggestive of obligation, and since it remains unclear what it is – if not an obligation – that the debtor is fulfilling/violating when *s* is settled true/settled false.

However, Verdicchio and Colombetti later go on to say (Verdicchio and Colombetti 2004, p. 142) that they too accept as fundamental the distinction between constitutive conventions and preservative norms, and maintain further that "All the rules connecting the messages to commitments are constitutive conventions: they say that messages of certain forms *count as* certain operations on commitments." (*loc. cit.*) Comparing their approach with ours, they claim that we "both consider agent communication as fully conventional, and regard a false assertion as some kind of violation." (*loc. cit.*)

So, if the 'kind of violation' that occurs when an agent makes a false assertion is not the violation of an obligation, of what sort is it? Our view, as indicated above, is indeed that we are *not* here concerned with violation of some *directive* norm – for directive norms do not constitute communication, although they may serve to preserve it. Rather, the making of a false assertion should be seen as 'violation' of an optimality or ideality condition, of the type captured by the $I_s^\star$ operator, which expresses a normative modality of the *evaluative*, rather than directive, type. The term 'violation' fits uncomfortably here, and is best replaced by the notion of *deviation or departure from the ideal.*

Our conjecture, then, is that our account, expressed essentially in terms of the counts-as and optimality operators, provides an explicit formal theory of the intuitions about the nature of communicative conventions that Verdicchio and Colombetti's approach leaves unarticulated.

3.5. Some other types of communicative acts

We have so far considered the communicative act-type of asserting, but there are of course other types as well, and we here indicate how the approach advocated above can be extended to incorporate them. The account provides no more than a sketch, and makes no claims to being complete in the sense of giving an exhaustive characterisation of communicative act-types. Nevertheless, the sketch should indicate the flexibility and expressive power of the logical framework employed.

We consider five types:

- Commands
- Permissives (granting permission)
- Commissives (placing oneself under an obligation, e.g., promising)
- Requests
- Declaratives [in the sense of (Searle and Vanderveken 1985)].

In each case, the governing signalling convention will take the form of (sc-assert) with, crucially, some further elaboration of the scope-formula A in the consequent. This means that each of these signalling act-types is a sub-species of the act of asserting – a consequence which is harmless, and which simply reflects the fact that all communicative acts are acts of transmitting information – information which may, or may not, be true. However, as will emerge in due course, there is one very important difference between pure assertives and these sub-species, and this difference may also be thought to provide an answer to one of the key questions from which Austin started in *How to Do Things with Words* (Austin 1962) – the question of how to distinguish *constatives* from *performatives.*[11]

3.5.1. Commands

Let j be the agent issuing the command, and let k be the agent who is being commanded to see to it that A. Then the form of the governing signalling convention is:

$$\text{(sc-command)} \quad E_jC \Rightarrow_s I_s^\star OE_kA \qquad (76)$$

where the 'O' operator is a directive normative modality. So, according to (sc-command), if j sees to it that C, s would then be in an optimal state, relative to its function of facilitating the transmission of reliable information, if there were an obligation on k to see to it that A.

3.5.2. Permissives

Let j be the agent issuing the permission, and let k be the agent who is to be permitted to see to it that A. Then the form of the governing signalling convention is:

$$\text{(sc-permit)} \quad E_jC \Rightarrow_s I_s^{\star}PE_kA \tag{77}$$

where the 'P' operator is the dual of the directive normative modality, and we are thus adopting the simplification that 'permitted to do A' is interpreted as 'not obliged not to do A'. So, according to (sc-permit), if j sees to it that C, s would then be in an optimal state, relative to its function of facilitating the transmission of reliable information, if k were permitted to see to it that A.

3.5.3. Commissives

Let j be the agent issuing the commissive. Then the form of the governing signalling convention is:

$$\text{(sc-commit)} \quad E_jC \Rightarrow_s I_s^{\star}OE_jA. \tag{78}$$

So, according to (sc-commit), if j sees to it that C, s would then be in an optimal state, relative to its function of facilitating the transmission of reliable information, if j were himself under an obligation to see to it that A.

3.5.4. Requests

Let j be the agent making the request, and let the aim of the request be to get agent k is to see to it that A. Then the form of the governing signalling convention is:

$$\text{(sc-request)} \quad E_jC \Rightarrow_s I_s^{\star}H_jE_kA \tag{79}$$

where the relativised 'H' operator represents the modality 'attempts to see to it that...' (see above, Section 2.1.3). So, according to (sc-request), if j sees to it that C, s would then be in an optimal state, relative to its function of facilitating the transmission of reliable information, if j were attempting to see to it that k sees to it that A.

3.5.5. Declaratives

These are the kinds of signalling acts that are performed by, for instance, the utterance of such sentences as:

- 'I pronounce you man and wife'.
- 'I name this ship *Generalissimo Stalin*'.
- 'I pronounce this meeting open'.

The point of declaratives is to create a new state of affairs, which will itself often carry particular normative consequences concerning rights and obligations, as when two persons become married, or a meeting is declared open. In the spirit of the approach developed in Jones and Sergot (1996), we may say that declaratives are used by designated agents within institutions as a means of generating institutional facts: facts which, when recognised by the institution as established, are deemed to have particular kinds of normative consequences.

Let j be the agent issuing the declarative, and let A describe the state of affairs to be created by the declarative. Then the form of the governing signalling convention is:

$$\text{(sc-declare)} \quad E_j C \Rightarrow_s I_s^\star E_j A. \tag{80}$$

So, according to (sc-declare), if j sees to it that C, s would then be in an optimal state, relative to its function of facilitating the transmission of reliable information, if j had seen to it that A. For instance, j utters the words ‘I pronounce you man and wife’, and then s’s being in an optimal state would require that j has indeed seen to it that the couple are married.[12]

3.6. Being empowered

For each of the four types just considered, if j is an empowered/authorised agent, then the *mere performance* by j of the act of seeing to it that C will be sufficient in itself to *guarantee* the truth of the respective formula to the right of the $I_s^\star$ operator.[13] For instance, if j is empowered/authorised to command k, then his seeing to it that C will indeed create an obligation on k to do A. Likewise, if j is empowered/authorised to commit himself, then performing the appropriate communicative act will be enough to place himself under an obligation. And if j is empowered/authorised to make a request to k, then his communicative act will constitute an attempt to get k to do the requested act. And so on.

Here lies the key to the crucial difference, alluded to above, between pure assertions and the other types of communicative act. For pure assertions, there is no notion of empowerment or authorisation which will license the inference of A from the truth of $I_s^\star A$. The closest one could get to such a notion would be the case where j is deemed to be an authority on the subject about which he is making an assertion: but even then, his *saying* that A does not *make it the case* that A.[14]

We have now presented a new formal approach to the theory of ACLs, in which a class of signalling conventions, governing some distinct types of communicative acts, can be represented. Other types of communicative act remain to be characterised. But we now turn to the task of embedding this ‘static’ account of communication within a theory of *conversation*, in which sequences of inter-related signalling acts are transmitted.

4. Modelling Conversations

Conversations are essentially dynamic in nature. In this section, we outline one possible way of adding a dynamic dimension to the theory of signalling acts, by combining it with the arrow logic of van Benthem (1991, 1994, 1996) and colleagues (Marx 1996; de Venema 1994).

Our proposal is twofold. First, we suggest giving a compact expression to conversation protocols, by means of a formula of the object-language. Second, we suggest using this kind of representation to provide the beginning of a procedure for keeping a record of the conventional effects achieved in a conversation.

The reason why we do not use dynamic logic in its traditional form – see (Pratt 1976) – is that it presupposes a kind of approach to the logic of agency that is very different from the treatment provided in the theory of signalling acts. As indicated in Section (2.1.3), the present framework treats agency as a modal operator, with some reading such as "agent *j* sees to it that". Dynamic logic has explicit labels for action terms. These are not propositions but (to put it in Castañeda's terms) practitions.

It might well be the case that temporal logic provides a better account than arrow logic. The reason why we have chosen to concentrate first on arrow logic is that, when moving to the dynamics, we do not have to redefine the main ingredients of the semantics used for the static account. Indeed all we need to do is to interpret the points in a model as transitions. The completeness problem for the integrated framework is, then, relatively easy.

4.1. Embedding the static account within arrow logic

The syntax of arrow logic has in general the following three building blocks:

- A binary connective denoted by $\circ$ referred to as "composition";
- A unary connective denoted by $\breve{}$ referred to as "reverse" (or "cap");
- A propositional constant denoted by Id referred to as "identity".

The sentences that replace A, B, ..., that the first two connectives take as arguments, are supposed to describe an event, an action, etc. More expressive modal operators can be added into the vocabulary of the logic. For present purposes, we need not introduce them. Suffice it to observe that this way of adding dynamics to our static account is very natural, because a frame in arrow logic is no more than an ordinary Kripke frame. The only difference is that the universe W is viewed as consisting of arrows. These are not links between possible worlds. In fact they are treated themselves as the possible worlds.[15]

Once $\circ$, $\breve{}$ and Id have been introduced as new building blocks, it seems natural to proceed as follows.

Definition 1 Let F be a minimal frame as defined within the static framework. F is simply a minimal model structure M (see above, Section 2.2.1), stripped of its last ingredient, the valuation function P. By an enriched minimal frame, let us mean a quadruplet

$$\mathcal{F} = (F, C, R, I)$$

where F is a minimal frame, and

- $C \subseteq W \times W \times W$ is the semantical counterpart of $\circ$. Expressions of the type $C\alpha\beta\gamma$ are read as: α is a "composition" of β and γ. One might also heuristically read $C\alpha\beta\gamma$ as: transition α consists of transitions β and γ.
- $R \subseteq W \times W$ is the semantical counterpart of $\breve{}$. Expressions of the type $R\alpha\beta$ are read as: β is a "reversal" of α.
- $I \subseteq W$ is the semantical counterpart of Id. Expressions of the type $I\alpha$ are read as: α is an "identity" arrow.

The semantics will be easier to handle if R and C are each assumed to be functional:

$$\forall\alpha\beta\gamma\ ((R\alpha\beta\ \&\ R\alpha\gamma) \rightarrow \beta = \gamma) \tag{81}$$

$$\forall\alpha\beta\gamma\delta\epsilon\ ((C\alpha\beta\gamma\ \&\ C\alpha\delta\epsilon) \rightarrow (\beta = \delta\ \&\ \gamma = \epsilon)). \tag{82}$$

Now let us define a model as a pair

$$\mathcal{M} = (\mathcal{F}, V)$$

where $\mathcal{F}$ is an enriched minimal frame, and V is an assignment that associates a set of arrows with each propositional letter P_n. Informally, we think of $V(P_n)$ as the set of arrows in our model where P_n is true.

Armed with these notions, we can easily redefine what it means for any sentence A to be true at an arrow α in $\mathcal{M}$, in symbols $\mathcal{M}, \alpha \models A$. It suffices to keep the package of truth-clauses already used in the static framework, and to introduce those usually employed for $\circ$, $\breve{}$ and Id. We state this formally.

Definition 2 Let $\mathcal{M} = (\mathcal{F}, P)$ be a model. For non-dynamic sentences – those of the forms described in the statics, the truth conditions remain unchanged. The dynamic sentences are evaluated as follows:

$$\begin{aligned}\mathcal{M}, \alpha \models A \circ B \text{ iff } \exists\beta, \gamma \text{ in } \mathcal{M} \text{ such that } C\alpha\beta\gamma \text{ and}\\ \mathcal{M}, \beta \models A \text{ and } \mathcal{M}, \gamma \models B\end{aligned} \tag{83}$$

$$\mathcal{M}, \alpha \models \text{Id iff } I\alpha \tag{84}$$

$$\mathcal{M}, \alpha \models A^{\breve{}} \text{ iff } \exists\beta \text{ in } \mathcal{M} \text{ such that } R\alpha\beta \text{ and } \mathcal{M}, \beta \models A. \tag{85}$$

The first truth-clause says that $A \circ B$ is true at an arrow α iff it can be decomposed into two arrows at which A and B hold, respectively. This can be pictured as in Figure 1. The intended meaning of this connective is relatively transparent. A sentence of the form $A \circ B$ can be read as meaning that the event described by A is followed by the event described by B. The two arrows at which A and B are evaluated can be seen as two intervals (periods of time).

Next, the evaluation rule for Id ("identity") says that, for Id to be true at α, α must be a transition that does not lead to a different state. This can be pictured as in Figure 2.

Finally, the truth-clause for $^{\smile}$ ("reverse") says that, for $A^{\smile}$ to be true at α, there must be an arrow β that is the reversal of α and at which A holds. This is shown in Figure 3. It is natural to say that such an operator has the meaning of "undo-ing" an action. In Figure 3, arrow β, at which A is true, leads from one state to another. Intuitively, the endpoint of β contains the effects of the performance of A in β. Arrow α, at which $A^{\smile}$ is true, goes in the opposite direction, so that the effects of the performance of A in transition β are cancelled. Of course, we give this model for heuristic purposes only, since the formalism is not expressive enough to allow us to reason about states as well. However, it is possible (at least in principle) to remove this limitation, by switching to so-called two-sorted arrow logics. Introduced in van Benthem (1994), these are designed for reasoning about both states and transitions. Accounts of agency in terms of a "brings it about" operator focus on the agent concerned and the state of affairs that results from their action. But in our account "bringing it about that A" is evaluated at transitions. A tempting reading of the evaluation rule employed in the statics, i.e.

$$\mathcal{M}, \alpha \models E_j A \text{ iff } \alpha \in f_{c_j}(\alpha, ||A||^{\mathcal{M}}), \tag{86}$$

is

> $E_j A$ is true at transition α iff α belongs to the set of transitions where j realizes the ability he has in α to bring about the transition described by A.

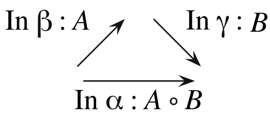

Figure 1. Composition.

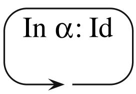

Figure 2. Identity.

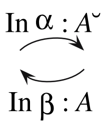

Figure 3. Reverse.

It seems very natural to try to refine the formalism in such a way that what obtains within states is also taken into account.[16] We shall explore this issue in future research. Some work along these lines has already been conducted by Segerberg (1989) in the context of dynamic logic. It may also be valuable to explore the relationship between that work and the approach outlined in the present paper.

We now turn to the axiomatic characterization of the framework. When no particular constraints are imposed on the way C, R and I interact, the proof theory of the integrated framework can in fact be obtained by adding the following rules of inference and axiom schemata to the multi-modal system presented in Section 2:

Rules of inference

$$\text{From } \vdash B \rightarrow C \text{ infer } \vdash (A \circ B) \rightarrow (A \circ C) \tag{87}$$

$$\text{From } \vdash A \rightarrow C \text{ infer } \vdash (A \circ B) \rightarrow (C \circ B) \tag{88}$$

$$\text{From } \vdash A \rightarrow B \text{ infer } \vdash A^{\smile} \rightarrow B^{\smile} \tag{89}$$

$$\text{From } \vdash A \text{ infer } \vdash \neg(\neg A \circ B) \tag{90}$$

$$\text{From } \vdash A \text{ infer } \vdash \neg(B \circ \neg A) \tag{91}$$

$$\text{From } \vdash A \text{ infer } \vdash \neg((\neg A)^{\smile}) \tag{92}$$

Axiom schemata

$$\vdash (A \vee B) \circ C \rightarrow (A \circ C) \vee (B \circ C) \tag{93}$$

$$\vdash A \circ (B \vee C) \rightarrow (A \circ B) \vee (A \circ C) \tag{94}$$

$$\vdash (A \vee B)^{\smile} \rightarrow A^{\smile} \vee B^{\smile} \tag{95}$$

$$\vdash (A \circ C) \wedge (B \circ C) \rightarrow (A \wedge B) \circ C \tag{96}$$

$$\vdash (A \circ B) \wedge (A \circ C) \rightarrow A \circ (B \wedge C) \tag{97}$$

$$\vdash A^{\smile} \wedge B^{\smile} \rightarrow (A \wedge B)^{\smile}. \tag{98}$$

Rules (87)–(89) express a principle of closure under consequence. Rules (90)–(92) are the arrow counterparts of the necessitation rule. Axioms (93)–(95) say that $\circ$ and $^{\smile}$ distribute over disjunction. Axioms (96)–(98) say that $\circ$ and $^{\smile}$ factorize over conjunction. (96)–(98) are the axiomatic counterparts of constraints (81) and (82) described above. Note that, in the general case where no constraints are placed on the way I, R and C interact, it is not necessary to introduce any specific axiom involving the constant Id. In this general case, no new validities involving this constant appear except those that are an instantiation of one of the above laws.

A proof of soundness and completeness for the extended framework is given in Section 5. The proof is based on the standard technique of canonical model construction (see, e.g., Blackburn, de Rijke and de Venema 2001).

4.2. Conversation protocols

In this section, we illustrate the expressive capacity of the logic, by showing how it can be applied to the analysis of conversation protocols. Examples abound, but a good starting point is the conversation protocols defined as part of the FIPA standardisation process. Formalisms that have been proposed to model conversation protocols include: finite-state diagrams, Petri Nets (Cost et al. 2000; Ferber 1995; Lin et al. 2000), Dooley graphs (Parunak 1996) dialogue-games (Mc Burney and Parsons 2002; Reed 1998) and denotational semantics (Pitt 1999). The style of analysis we will outline below is reminiscent of the first two, but differs from them in two significant ways. First, we suggest representing the protocol as a formula giving compact expression to the set of permissible sequences of speech acts. Second, this style of analysis paves the way for an approach to the study of conversation that (to the best of our knowledge) has never been considered in the literature on ACLs. It is in terms of a procedure for keeping a record of the conventional effects achieved during a conversation. A close look at the so-called English Auction Protocol will serve as a means of presenting the basic idea of the treatment. Although we need to subject this point to further investigation, we believe that a similar formal treatment might be devised for the other conversational protocols usually discussed in the literature on ACLs.

Figure 4 depicts the English Auction Protocol used between an auctioneer agent a and each agent buyers b. The nodes (circles) represent states of the conversation, and the arcs (lines) represent speech acts that cause transition from state to state in the

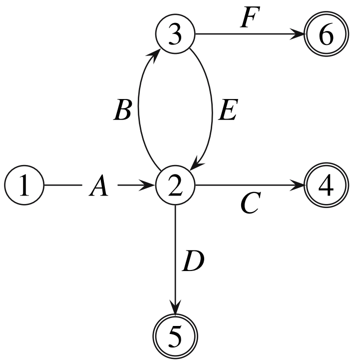

Figure 4. English Auction Protocol.

conversation. The circles with a double-line represent the final states of the conversation.

The propositional letters attached to the arcs are notational shorthand for the following speech acts:

- A: a puts item c up for auction;
- B: b makes a bid;
- C: a informs b that the item is sold to another buyer;
- D: a declares that the auction is at an end;
- E: a informs b that another buyer overbids;
- F: a informs b that his bid wins.

We use propositional letters for clarity's sake only. In fact, A corresponds to the antecedent of a conventional signalling rule of type (sc-declare), and likewise for D. B is to be replaced by the antecedent of a signalling convention taking the form of (sc-commit). The scope formula in the consequent uses a conditional obligation, $O(E_b A_2/A_1)$, according to which b is under the obligation to pay if his offer is accepted. We leave aside discussion of the problem of how to analyse the conditional obligation operator $O(/)$ [an elaborate formal treatment is available in Carmo and Jones (2002)]. C, E and F each correspond to the antecedent of a signalling convention taking the form of (sc-assert).

The main function of a protocol is to define the sequences of speech acts that are permissible during a conversation. The basic idea is to assume that such sequences can be expressed in a compact way, by means of a disjunction containing $\circ$, $\breve{}$ and/or Id. For instance, the English Auction Protocol is an instantiation of the formula

$$(A \circ D) \vee (A \circ C) \vee (A \circ (B \circ F)) \vee (A \circ (B \circ (E \circ C))) \tag{99}$$

where (as we have just indicated) A, B, C, D, E and F stand for the antecedents of the appropriate signalling conventions. Since $\circ$ distributes over $\vee$, (99) can be simplified into

$$A \circ \big(D \vee C \vee (B \circ (F \vee (E \circ C)))\big) \tag{100}$$

(99) considers in isolation the sequences of acts that are allowed by the protocol. The first disjunct in (99), $A \circ D$, translates the path 1-2-5. The second disjunct, $A \circ C$, translates the path 1-2-4. The third disjunct, $A \circ (B \circ F)$ translates the path 1-2-3-6. The fourth and last disjunct, $A \circ (B \circ (E \circ C))$, translates the path 1-2-3-2-4. Formula (100) puts the sequences of speech acts together, and indicates the points when interactants have the opportunity to choose between two or more speech acts. (100) can be read as follows. Once A has been done, then we can have either D, C or B. And once B has been done, we can have either F or E-followed-by-C. For simplicity's sake, we assume here that auctioneer a receives at most two bids. The fact that auctioneer a can receive more than two bids might be captured by an operator expressing iteration.

As the auction evolves, there is a shift in focus from the whole disjunction (99) to one specific disjunct. The latter records the acts (which are not necessarily verbal) performed in a conversation. It seems reasonable to expect a formal language for ACLs to also provide a way of keeping a record of the conventional effects achieved by these acts. As a further refinement, the recording might take into account which users of signalling system s are empowered agents, and which trust relationships exist between agents. Although we need to subject this issue to further investigation, we can already give some hint of how such a record can be achieved in the present framework. It consists in using the notion of logical consequence as defined by Fitting (1983):

Definition 3 Let S and U be sets of formulas, and X be a formula. By $S \models U \rightarrow X$, we mean: for every model $\mathcal{M}$ in which the members of S are valid, and for every arrow α in $\mathcal{M}$ at which the members of U are true, it is the case that $\mathcal{M}, \alpha \models X$.

Such a construction exploits the idea that the local and the global consequence relations used in modal logic can be subsumed under one more general relation. In the notation

$$S \models U \rightarrow X, \tag{101}$$

S expresses global assumptions, holding at all arrows. In contrast, U enumerates local assumptions, holding at particular arrows. In line with our previous analysis – see Section 3.1 – we assume that S contains the signalling conventions adopted by institution s. These are mutually believed by the agents who use s. Here, S plays the role of a black box that takes U (a sequence of communicative acts) as input and gives X (a list of conventional effects) as output. For instance, if the focus is on the sequence $A \circ (B \circ F)$, then S is the set having the following three elements:

$$E_a A_1 \Rightarrow_s I_s^\star E_a A_4 \tag{102}$$

$$E_b A_2 \Rightarrow_s I_s^\star O(E_b A_6 / A_5) \tag{103}$$

$$E_a A_3 \Rightarrow_s I_s^\star A_7. \tag{104}$$

Now let us adopt the point of view of an external observer x. This means that we can specify U in (101) as

$$B_x E_a A_1 \circ (B_x E_b A_2 \circ B_x E_a A_3). \tag{105}$$

As can easily be verified, the doxastic form of modus ponens (71) used in the 'static' framework allows us to specify X in (101) as

$$B_x I_s^\star E_a A_4 \circ (B_x I_s^\star O(E_b A_6 / A_5) \circ B_x I_s^\star A_7), \tag{106}$$

which represents a record of the conventional effects achieved in the conversation.

Depending on x's beliefs about the empowerment and trustworthiness of the communicators a and b, the record will include some further features. For instance, if x believes that a and b are empowered to declare and commit, respectively, and if x also believes that a's assertion of A_7 is trustworthy (reliable), then the record will also show:

$$B_x E_a A_4 \circ (B_x O(E_b A_6/A_5) \circ B_x A_7). \tag{107}$$

One last remark is to be made. So far we have used only the operator $\circ$, in order not to distract the reader from the main point we wish to make in this paper. It is possible to use the other two operators, Id and $\breve{}$, so as to capture further aspects of the protocol. The modal constant Id can be used to capture the obvious fact that, once a has suggested a starting-price for the goods, it may happen that another agent, call it b', opens the bid. Operator $\breve{}$ can be used to express the fact that, once E has been performed, the conversation returns to the prior state 2. Finally, it should be mentioned that the presence of a potential cycle might easily be captured by using the unary connective usually denoted by * and referred to as "iteration" (also "Kleene star"). We defer the full discussion of this issue to another occasion.

5. Soundness and completeness

This section reports a soundness and completeness result for the integrated logical framework we have presented. We shall refer to the axiomatic characterization of the entire framework as $\mathcal{L}$, and to the class of frames with which $\mathcal{L}$ is associated as $\wp$. To keep matters simple, we look at the $\star$-free fragment – that is, at formulas without occurrences of the iteration operator, referred to in the previous paragraph.

We stress that the result of the present section concerns only the so-called *local* semantic consequence relation, which demands that the maintenance of truth should be guaranteed point to point or *locally*. Let Γ and A be a set of formulas and a single formula. As usual, we say that A is a local semantic consequence of Γ (notation : $\Gamma \models A$) if and only if for all models $\mathcal{M}$ and all arrows α in W, $\mathcal{M}, \alpha \models A$ whenever $\mathcal{M}, \alpha \models B$ for all B in Γ. The associated proof theoretic consequence relation $\vdash$ is defined as follows. We say that A is a syntactic consequence of Γ (notation : $\Gamma \vdash A$) if and only if there is some finite part $\{A_1, ..., A_n\}$ of Γ such that $\vdash (A_1 \wedge \ldots \wedge A_n) \rightarrow A$. The completeness problem for the Fitting construction discussed in the previous section remains open. As we have seen, the distinctive feature of the latter construction is that it subsumes the local and the global consequence relations under a more general notion.

Before proceeding to the proof, we need to introduce some further terminology. A set Γ of wffs is inconsistent iff $\Gamma \vdash \bot$; otherwise Γ is consistent. A set Γ of wffs is maximal iff for every wff A, either $A \in \Gamma$ or $\neg A \in \Gamma$. Γ is maximal consistent iff it is both maximal and consistent. Finally, a set Γ of wffs is satisfiable iff and there exists a model $\mathcal{M}$ and an arrow α such that $\mathcal{M}, \alpha \models A$ for every A in Γ; otherwise, Γ is unsatisfiable.

The soundness theorem states that every formula that can be proved in the system from some set of assumptions can also be obtained as a semantic consequence from that set. Formally, for all formulae A, and all sets of formulae Γ,

$$\Gamma \vdash A \longrightarrow \Gamma \models A.$$

The verification of the soundness part is straightforward, by showing that every axiom is valid, and every rule of inference preserves validity. Thus, we concentrate our attention on the strong completeness theorem, which states that every formula that can be obtained as a semantic consequence from some set of formulae can also be proved from that set, i.e., for all formulae A, and all set of formulae Γ,

$$\Gamma \models A \longrightarrow \Gamma \vdash A.$$

The proof is based on the standard canonical model construction (see, e.g., Blackburn et al. 2001). The first step involves defining the so-called canonical model $\mathcal{M}^{\mathcal{L}}$ for $\mathcal{L}$. The usual definition carries over to the present framework with obvious modifications to take into account the new building blocks. Let W be the set of all maximal consistent sets of sentences (in short, MCSs). We use the symbolism $|A|_{\mathcal{L}}$ for the class of maximal consistent sets containing the sentence A. The ingredients employed in the static framework are given by:

- $f_{\Rightarrow_s}$ is a function from $W \times P(W)$ to $P(P(W))$ such that, for each arrow α in W, and each $|A|_{\mathcal{L}}$,

$$f_{\Rightarrow_s}(\alpha, |A|_{\mathcal{L}}) = \{|B|_{\mathcal{L}} \subseteq W : A \Rightarrow_s B \in \alpha\} \tag{108}$$

- Let $\underline{f} \in \{f_n, f_{h_i}, f_o, f_{i_i}, f_s^{\star}, f_{b_i}\}$; $\underline{f}$ is a mapping from W to $P(P(W))$ such that, for each arrow α in W,

$$\underline{f}(\alpha) = \{|A|_{\mathcal{L}} \subseteq W : \dagger A \in \alpha\}. \tag{109}$$

Here $\dagger$ is for the modal operator with which $\underline{f}$ is associated.

- f_{c_i} is a mapping from $W \times P(W)$ to $P(W)$ such that, for each arrow α in W, and each $|A|_{\mathcal{L}}$,

$$f_{c_i}(\alpha, |A|_{\mathcal{L}}) = \{\beta \in W : E_i A \in \beta\}. \tag{110}$$

The relations associated with the dynamic modalities are given by: [17]

$$C^{\mathcal{L}}\alpha\beta\gamma \text{ iff } \forall A \in \beta\ \forall B \in \gamma\ A \circ B \in \alpha \tag{111}$$

$$R^{\mathcal{L}}\alpha\beta \text{ iff } \forall A \in \beta\ A^{\smile} \in \alpha \tag{112}$$

$$I^{\mathcal{L}}\alpha \text{ iff } \mathrm{Id} \in \alpha. \tag{113}$$

The next (non-trivial) step is to prove the following:

Lemma 1 [Existence Lemma]. *For any MCS $\alpha \in W$, if $A^{\smile} \in \alpha$ then there is some MCS $\beta \in W$ such that $R^{\mathcal{L}}\alpha\beta$ and $A \in \beta$. Similarly, for any MCS $\alpha \in W$, if $A \circ B \in \alpha$ then there are $\beta,\gamma \in W$ such that $C^{\mathcal{L}}\alpha\beta\gamma$, $A \in \beta$ and $B \in \gamma$.*

Proof: We only prove the first part of the lemma (for the second part, the argument is similar). Let α be a maximal consistent set of formulae. Suppose $A^{\smile} \in \alpha$. We need to construct a MCS β such that $R^{\mathcal{L}}\alpha\beta$ and $A \in \beta$. By setting $\beta^{-} = \{B : B^{\smile} \in \alpha\}$, we get exactly what we need; all that remains to be checked is that β^{-} is maximal consistent. The proof will *inter alia* illustrate the usefulness of the rules (89) and (92), and of the axiom schemata (95) and (98). In the proof we will use some familiar properties of maximal consistent sets.

Suppose, per absurdum, that β^{-} is inconsistent. This means that there are $B_1,\ldots,$ B_n in β^{-} such that

$$\vdash (B_1 \wedge \ldots \wedge B_n) \to \bot \tag{114}$$

hence by PC

$$\vdash (B_1 \wedge \ldots \wedge B_n) \to \neg A. \tag{115}$$

So by rule (92)

$$\vdash \neg[\big(\neg((B_1 \wedge \ldots \wedge B_n) \to \neg A)\big)^{\smile}]. \tag{116}$$

So by rule (89)

$$\vdash \neg[(B_1 \wedge \ldots \wedge B_n \wedge A)^{\smile}] \tag{117}$$

from which we get (since a MCS always contains the theorems of the logic)

$$\neg((B_1 \wedge \ldots \wedge B_n \wedge A)^{\smile}) \in \alpha \tag{118}$$

and finally (given that α is consistent)

$$(B_1 \wedge \ldots \wedge B_n \wedge A)^{\smile} \notin \alpha. \tag{119}$$

But, on the other hand, B_1, ... and B_n each are in β^{-}. By construction of β^{-}, it follows that $B_1{}^{\smile} \in \alpha$, ... and $B_n{}^{\smile} \in \alpha$. We also have $A^{\smile} \in \alpha$. Since a MCS is closed under conjunction, $B_1{}^{\smile} \wedge \ldots \wedge B_n{}^{\smile} \wedge A^{\smile} \in \alpha$. By axiom (98) we then get $(B_1 \wedge \ldots \wedge B_n \wedge A)^{\smile} \in \alpha$ – contradicting equation (119) above. We thus conclude that β^{-} is consistent after all.

We now argue that β^{-} is maximal. By rule (89),

$$\vdash A^{\smile} \rightarrow (B \vee \neg B)^{\smile}. \tag{120}$$

Hence

$$A^{\smile} \rightarrow (B \vee \neg B)^{\smile} \in \alpha. \tag{121}$$

From this together with $A^{\smile} \in \alpha$, it follows that $(B \vee \neg B)^{\smile} \in \alpha$, since a MCS is closed under implication. Given that $^{\smile}$ distributes over disjunction – axiom (95), we then get $B^{\smile} \vee (\neg B)^{\smile} \in \alpha$. This means that either $B^{\smile} \in \alpha$ or $(\neg B)^{\smile} \in \alpha$. We may now apply the definition of β^{-} to conclude that either $B \in \beta^{-}$ or $\neg B \in \beta^{-}$, as required.

The proof of the second part of the Lemma follows the same pattern, and is therefore straightforward. □

With this established, the rest is easy. First, we establish the so-called Truth Lemma for $\mathcal{M}^{\mathcal{L}}$, to the effect that being a member of α in $\mathcal{M}^{\mathcal{L}}$ is equivalent to being true in α:

Lemma 2 [Truth Lemma]. *For every MCS* α in $\mathcal{M}^{\mathcal{L}}$:

$$\mathcal{M}^{\mathcal{L}}, \alpha \models A \textit{ if and only if } A \in \alpha$$

or, equivalently,

$$||A||^{\mathcal{M}^{\mathcal{L}}} = |A|_{\mathcal{L}}. \tag{122}$$

Proof: As usual, the proof is by induction on the form of A. The base case follows from the definition of V in the canonical model. For the inductive cases we adopt the hypothesis that the result holds for sentences shorter than A. The boolean cases are treated in the usual manner. It remains to deal with the modalities.

As for those employed in the static framework, we can restrict ourselves to the case where A is E_iB (for the other modalities, the argument is similar). One might here argue as follows:

$$\begin{aligned} \mathcal{M}^{\mathcal{L}}, \alpha \models E_iB \text{ iff } & \alpha \in f_{c_i}(\alpha, ||B||^{\mathcal{M}_{\mathcal{L}}}) \\ & \text{(truth-conditions for } E_i, \text{ i.e. (48))} \\ \text{iff } & \alpha \in f_{c_i}(\alpha, |B|_{\mathcal{L}}) \text{ (inductive hypothesis)} \\ \text{iff } & E_iB \in \alpha \text{ (definition of } f_{c_i} \text{ in the canonical model).} \end{aligned}$$

We now deal with the arrow modalities:

- If A is of the form Id, there is no difficulty. It suffices to invoke the truth-conditions for Id as well as the definition of $I^{\mathcal{L}}$ in the canonical model:

$\mathcal{M}^{\mathcal{L}}, \alpha \models \mathrm{Id}$ iff $I^{\mathcal{L}}\alpha$ (truth-conditions for Id)

iff $\mathrm{Id} \in \alpha$ (definition of $I^{\mathcal{L}}$).

- Suppose A is of the form $B^{\smile}$. The left to right direction is immediate too:

$\mathcal{M}^{\mathcal{L}}, \alpha \models B^{\smile}$ iff $\exists\beta$ ($R^{\mathcal{L}}\alpha\beta$ & $\mathcal{M}^{\mathcal{L}}, \beta \models B$) (truth-clause for $^{\smile}$)

iff $\exists\beta$ ($R^{\mathcal{L}}\alpha\beta$ & $B \in \beta$) (inductive hypothesis)

only if $B^{\smile} \in \alpha$ (definition of $R^{\mathcal{L}}$).

For the right to left direction, suppose $B^{\smile} \in \alpha$. By the equivalences above, it suffices to find a MCS β such that $R^{\mathcal{L}}\alpha\beta$ and $B \in \beta$ – and this is precisely what the Existence Lemma guarantees. If A is of the form $B \circ C$, the proof is similar, and is straightforward. □

Lemma 3 *The canonical model of $\mathcal{L}$ is based on a frame in $\wp$.*

Proof: First, we need to show that $R^{\mathcal{L}}$ and $C^{\mathcal{L}}$ satisfy conditions (81) and (82), respectively. This follows immediately from the definitions of $R^{\mathcal{L}}$ and $C^{\mathcal{L}}$ in the canonical model. Next, we need to show that the functions $f_n, f_{c_i}, f_{h_i}, f_o, f_{i_i}, f_{b_i}, f_s^{\star}$ and $f_{\Rightarrow_s}$ satisfy the constraints initially placed on them (see Section 2). As usual, this is shown by using the proof-theoretic properties of the system. The proof is straightforward. □

Lemma 4 *Every consistent set of sentences is satisfiable.*

Proof: The proof follows the usual pattern. It suffices to find, for any consistent set Γ, a model $\mathcal{M}$ and an arrow α in $\mathcal{M}$ such that each element of Γ is true at α. Simply take $\mathcal{M}$ to be $\mathcal{M}^{\mathcal{L}}$; $\mathcal{M}^{\mathcal{L}}$ has the right properties by the above Lemma 3. Given that Γ is consistent, we may now apply Lindenbaum's Lemma to conclude that there is some α in $\mathcal{M}^{\mathcal{L}}$ such that $\Gamma \subseteq \alpha$. By the above Truth Lemma, each element of Γ is true at α. □

With this last lemma in hand, we can prove the completeness theorem:

Theorem 1 [Completeness]. *Every formula that can be obtained as a semantic consequence from some set of formulae can also be proved from that set, i.e.,*

$$\Gamma \models A \longrightarrow \Gamma \vdash A.$$

Proof: The proof is the standard one. Assume $\Gamma \models A$. Then $\Gamma \cup \{\neg A\}$ is unsatisfiable and, hence, Lemma 4 allows us to conclude that $\Gamma \cup \{\neg A\}$ is inconsistent, that is to say $\Gamma \cup \{\neg A\} \vdash \bot$. The Deduction Theorem gives $\Gamma \vdash \neg A \rightarrow \bot$, from which we immediately get $\Gamma \vdash A$, as required. $\square$

6. Conclusion

In this paper we have presented a formal framework for agent communication that is sound and complete with respect to its proposed semantics. The formal framework is not, and is not intended to be seen as, a *computational* theory of ACLs. Rather, its purpose is to supply a 'middle-layer' between, on the one hand, an informal description of a communication protocol and, on the other hand, a computational model of inter-agent communication. In contrast to an informal description, the multi-modal language used in the formulation of the middle-layer captures in very precise terms the interpretation of the basic types of communicative act, and provides a means of formally articulating the sequence of acts that occur in a conversation. Since the multi-modal language is a *logical* language, it also affords the means for investigating the inferences that may be drawn from any fragment of a conversation, and for testing for consistency. Nevertheless, it is undeniably the case that the middle-layer may contain more detail than will be required for any given implementation: the requirements for the design of a particular inter-agent communication system may allow a number of simplifications to be made at the level of implementation; but the middle-layer will then serve to provide a clear picture (much clearer than that afforded by an informal description) of just *which* simplifications have been made. The middle-layer provides a rather comprehensive formal theory, against the backdrop of which the systems engineer may devise particular ACL implementations, meeting specific application requirements. The work reported in Jones and Kimbrough (2005) and Kimbrough et al. (2005) provides an indication of how this methodology may be applied in practice.

A conspicuous feature of our approach is that it is neither intention-based nor commitment-based, but convention-based. Although the 'dynamic' account outlined in this paper is preliminary, we hope it will lead to a comprehensive theory of conversation, and thus provide guidance to protocol designers. At the dynamic level, we have basically proposed a compact expression of conversation protocols, by using arrow logic. Although we need to subject this point to further investigation by applying the approach to a broad range of examples of protocols, we are inclined to think that this kind of representation can facilitate the systematic comparison of protocols. Such an issue has recently been addressed by McBurney, Parsons and Johnson (2002), but in the context of dialogue game protocols.[18] They focus on the question as to when two protocols may be considered to be the same, and they compare several reasonable definitions of equivalence of protocols. One such definition makes use of the notion of bisimulation, which arises in both modal logic and computer science. Roughly, two protocols are considered as equivalent if any state transition achievable in one is also achievable in the other.

To our knowledge, the idea of bringing to the fore the role of conventions in a conversation has not been previously considered in the Agent Communication Languages (ACLs) community. Clearly, there is much to be done before the account presented in this paper provides a complete theory. In particular we have to explore further the question of how the static component and the dynamic one interact. We have already indicated one way of moving from the static account to the dynamic one, but we still have to appreciate better the construction used to achieve this. In essence we have suggested that it is reasonable to expect a formal language for ACLs to provide a way of keeping a record of the conventional effects achieved in a conversation. We have also suggested that the record process should next take into account questions about whether users of signalling system s are empowered agents, or questions about whether one agent j trusts some other agent k. Considerations of the first type become particularly relevant when, for instance, we focus on those situations where agents buy and sell goods on *behalf of* some other agents. In recent years, we have seen the development of a number of systems that make it possible to advertise and search for goods and services electronically. Let us take the case of the Kasbah prototype (Chavez and Maes 1996). It is a Web-based system where users create autonomous agents to buy and sell goods on their behalf. Each of these agents is autonomous in that, once released into the marketplace, it negogiates and makes decisions on its own, without requiring user intervention. Suppose agent k makes a bid on behalf of user j. The background signalling convention (governing k's communicative act) takes the form

$$(E_kC \Rightarrow_s I_s^\star E_kE_jA) \wedge (E_jA \Rightarrow_s I_s^\star OE_jB). \tag{123}$$

The first conjunct expresses the acting-on-behalf-of aspect, the second expresses the commissive aspect. If k is empowered to make an offer (if, for instance, a time-out has not taken place), then the truth of E_kE_jA (and, hence, the truth of E_jA) is guaranteed. If user j is empowered as buyer (if j is not under age, or if j's credit is greater than or equal to the price of the good), then the truth of OE_jB also obtains. Here the idea is to classify the performance of communicative acts as valid or invalid according to whether or not the agent that performed that action had the institutional power to do it. Some work along these lines has already been conducted in the context of the study of the Contract-Net-Protocol [see Artikis, Pitt and Sergot (2002) and Artikis, Sergot and Pitt (2003)]. It may be valuable to further explore the relationship between this work and the approach outlined in the present paper.

Finally, it is worth mentioning that the static and the dynamic components also interact at a more general level than the one just described. Indeed one important question is whether it is possible to find in the static account general criteria of what counts as a well-formed conversation, that is to say some explanation of why some kinds of conversational turns are in general "preferred" to (more frequent than) others. This issue has been discussed at some length (Cohen and Levesque 1991; Grosz and Sidner 1990; Litman and Allen 1990). But most answers that have been suggested are based primarily on the analysis of the intentions of the speakers, and require a strong assumption of cooperativity. For instance, to account for the fact that *j*'s question is typically followed by *k*'s answer, such theories assume that *k* adopts *j*'s goal as his own. This is unrealistic in many contexts. So it is natural to attempt some conceptual clarification from other perspectives. According to another important research tradition, usually referred to as "conversation analysis", part of the answer lies in the particular considerations that arise in *face-to-face* interaction (Brown and Levinson 1987; Heritage 1984). On this account, the reason why *j*'s question is typically followed by *k*'s answer, is that a non-answer would imply a lack of consideration. Thus, the focus is still on individual mental states. However, this kind of explanation does not seem to be particularly well-suited for the modelling of conversation involving artificial agents instead of human agent participants.

Acknowledgements

This research was supported by the EU funded 5th Framework project ALFEBIITE (IST-1999-10298). The preparation of this paper continued during the EU funded Working Group iTrust (IST-2001-34910) and the EU funded 6th Framework Integrated Project TrustCoM (IST-2002-2.3.1.9). This support is gratefully acknowledged.

Notes

1. For instance, Levi gives the example of a teacher explaining a thesis to a group of students.
2. This terminology derives from Chellas (1980).
3. The distinction derives from Kanger (1972) and Pörn (1977). Pörn used 'Shall' where we use 'O' and 'Ought' where we use 'I'.
4. The modality I_i is not used in the characterisation of the types of communicative act analysed in Section 3, below. However, we believe that it would be needed in the formal account of some other types of communicative act (e.g., *advising*, *recommending*); accordingly, we include it here for the sake of completeness.
5. Cf. Delgrande (1988).
6. The '*' in the notation is introduced in the multi-modal language merely to distinguish this particular notion of ideality from the evaluative normative modality, 'I', that also figures in the same language (see Subsections 2.1.5 and 2.2.5). The logic of the I_s^* operator will be presented in the next section.
7. In an interesting comment, Kimbrough has pointed out to us that it is plausible that conventional signalling systems can arise naturally among agents that are not only unaware of the intentions of their communicants, but that have neither beliefs nor intentions. In this connection, he refers us to two books of Brian Skyrms (*Evolution of the Social Contract* and *The Stag Hunt and the Evolution of Social Structure*). Investigation of these issues will figure in our future work.
8. The use of the term 'position' here is quite deliberate, alluding to the theory of normative positions, and in particular to some well studied techniques for generating an exhaustive characterisation of the class of logically possible situations which may arise

for a given type of modality (or combination of modalities), for a given set of agents, vis-à-vis some state(s) of affairs. See, e.g., (Jones 2004; Jones and Sergot 1992; Sergot 1999) for illustrations of the development and application of the generation procedure. A more comprehensive account of the concept of trust, which incorporates the notion of 'trusting what someone says', is presented in Jones (2002).

9. See http://www.fipa.org/
10. Within the Philosophy of Language there has been a good deal of discussion of the relative merits of intention-based and convention-based approaches to the characterisation of communicative acts. FIPA's approach to ACLs seems to have been heavily, perhaps one-sidedly, influenced by theories deriving in large mesure from the Gricean, intention-based theory of meaning.
11. It was in part Austin's failure to find a satisfactory grammatical criterion for distinguishing between constatives and performatives that led to his development of the theory of illocutionary acts. For further discussion, see Jones and Kimbrough (2006).
12. In some cases it would be natural to say that it is not j, but rather the institution s (on whose behalf j acts), that brings it about that A. So a better formulation of the consequent in (80) would be: $I_s^{\star}(E_jA \vee E_sA)$.
13. We leave implicit here the obvious point that, in many cases, the communicative act has to be performed in a particular context – e.g., in the presence of witnesses – if it is to achieve its conventional effect.
14. This is an old idea in a new guise. A number of early contributors to the literature on performatives (Lemmon, Åqvist and Lewis, among them) suggested that the characteristic feature of performatives, in contrast to constatives, was 'verifiability by use', or the fact that 'saying makes it so'. See Jones (1983) for references.
15. In this approach, arrows are not required to have some particular internal structure (to be "ordered pairs", for instance).
16. In particular, as Elgesem observes, we want to be able to "express the important temporal dimension of actions like the opening of a window or the breaking of a vase; viz. that at the interval [in our terminology, the state] *immediately after* the action the window is open or the vase is broken. This last aspect, truth at an interval [a state] immediately after the action, is a characteristic temporal feature of 'bring-about' locutions of natural language" (Elgesem 1993, p. 114).
17. (111) and (112) are taken from van Benthem (1994, 1996). He claims that a completeness theorem is provable along standard lines by using these definitions, but he does not give the detailed proof of his claim. This section is an attempt to supply the missing details. In fact, the axiomatics $\mathcal{L}$ we have defined is different from the one used by van Benthem. He only works with the following three schemes of distribution, each of which is easily seen to be a thesis of $\mathcal{L}$:

$$(A \vee B) \circ C \leftrightarrow (A \circ C) \vee (B \circ C) \tag{124}$$

$$A \circ (B \vee C) \leftrightarrow (A \circ B) \vee (A \circ C) \tag{125}$$

$$(A \vee B)^{\smile} \leftrightarrow A^{\smile} \vee B^{\smile} \tag{126}$$

We have been unable to solve the completeness problem without modifying the axiomatics.

18. In general those modelling conversation protocols by using dialogue games start from a static account of communication that is commitment-based.

References

Artikis, A., J. Pitt and M. Sergot (2002). "Animated Specifications of Computational Societies," In: Castelfranchi C. and Johnson L. (eds.), *Proceedings of Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pp. 1053–1062.

Artikis, A., M. Sergot and J. Pitt (2003). "Specifying Electronic Societies with the Causal Calculator," In: F. Giunchiglia, J. Odell and G. Weiss (eds.), *Proceedings of Workshop on Agent-Oriented Software Engineering III (AOSE)*, Springer.

Austin, J. L. (1962). *How to do Things with Words*, Oxford: Oxford University Press.

Bennett, J. (1976). *Linguistic Behaviour*, Cambridge: Cambridge University Press.

Blackburn, P., M. de Rijke and Y. de Venema (2001). *Modal Logic*. Cambridge: Cambridge University Press.

Brandom, R. B. (1983). "Asserting," *Nous* 17, 637–650.

Brandom, R. B. (1994). *Making it Explicit – Reasoning, Representing, and Discursive Commitment*. London, England: Harvard University Press.

Brown, P. and S. C. Levinson (1987). *Politeness – Some Universals in Language Use*. Cambridge: Cambridge University Press.

Carmo, J. and A. J. I. Jones (2002). "Deontic Logic and Contrary-to-duties," In: D. Gabbay and F. Guenthner (eds.) *Handbook of Philosophical Logic*, Vol. 8, 2nd edn. Dordrecht, Holland: Kluwer Academic Publishers, pp. 265–343.

Chavez, A. and P. Maes (1996). "Kasbah: An Agent Marketplace for Buying and Selling Goods," In: *Proceedings of the 1st International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology (PAAM-96)*, London, UK, pp. 76–90.

Chellas, B. F. (1980). *Modal Logic – An Introduction*. Cambridge: Cambridge University Press.

Cohen, P. R. and H. J. Levesque (1991). "Confirmations and joint action," In: *Proceedings AJCAI-91*, pp. 951–957.

Colombetti, M. (2000). "A Commitment-based Approach to Agent Speech acts and Conversations," In: *Proceedings of the Workshop on Agent Languages and Conversation Policies*, Barcelona, pp. 21–29.

Cost, R. S., Y. Chen, T. Finin, Y. Labrou and Y. Peng (2000). "Issues in Agent Communication – An Introduction," In: F. Dignum and M. Greaves (eds) *Issues in Agent Communication* (pp. 178–192). Holland: Springer-Verlag.

Delgrande, J. (1988). "An Approach to Default Reasoning Based on a First-order Conditional Logic," *Artificial Intelligence*, 36, 63–90.

Elgesem, D. (1993). *Action Theory and Modal Logic*. PhD thesis, Institutt for filosofi. Det historik-filosofiske fakultetet. Universitetet i Oslo. See also Elgesem's paper 'The modal logic of agency' cited below.

Elgesem, D. (1997). "The Modal Logic of Agency," *Nordic Journal of Philosophical Logic* 2(2), 1–47.

Ferber, J. (1995). *Les Systèmes Multi-Agents – Vers une Intelligence Collective*. InterEditions.

Fitting, M. (1983). "Basic Modal Logic," In: C. Hogger, D. M. Gabbay and J. Robinson (eds) *Handbook of Logic in Artificial Intelligence and Logic Programming* 1 Oxford: Oxford University Press.

Føllesdal, D. (1967). "Comments on Stenius's 'Mood and Language Game'," *Synthese* 17, 275–280.

Grice, H. P. (1957). "Meaning," *The Philosophical Review* 66, 377–388.

Grosz, B. J. and C. L. Sidner (1990). "Plans for Discourse," In: P. R. Cohen, J. Morgan and M. E. Pollack (eds) *Intentions in Communication* (pp. 417–444). Cambridge: MIT Press.

Heritage, J. (1984). *Garfinkel and Ethnomethodology*. Cambridge: Polity Press.

Hilpinen, R. and D. Føllesdal (1971). "Introduction to Deontic Logic," In: R. Hilpinen (ed.) *Deontic Logic,* Dordrecht: D. Reidel Publishing Company.

Jones, A. J. I. (1983). *Communication and Meaning – An Essay in Applied Modal Logic*. Dordrecht, Holland: D. Reidel.

Jones, A. J. I. (2002). "On the Concept of Trust," *Decision Support Systems* 33, 225–232.

Jones, A. J. I. (2004). "On Normative- Informational Positions," In: A. Lomuscio and D. Nute (eds.) *Deontic Logic in Computer Science: Proceedings of the 7th International Workshop (DEON 2004)*, Vol. 3065 of *Lecture Notes in Artificial Intelligence*. Berlin: Springer, pp. 182–190.

Jones, A. J. I. and S. O. Kimbrough (2005). "A Note on Modelling Speech acts as Signalling Conventions," In: S. O. Kimbrough and D. J. Wu (eds) *Formal Modelling in Electronic Commerce, International Handbooks on Information Systems* (pp. 325–342). Berlin, Germany: Springer Verlag.

Jones, A. J. I. and S. O. Kimbrough (2006). "On the Normative Aspect of Signalling Conventions," In: L. Goble and J. -J. Ch. Meyer (eds.), *Proceedings of the Eighth International Workshop on Deontic Logic in Computer Science (DEON06)*, Lecture Notes in Computer Science, Berlin, July 2006, Springer (Forthcoming).

Jones, A. J. I. and X. Parent (2004). "Conventional Signalling Acts and Conversation," In: F. Dignum (ed.) *Advances in Agent Communication*, Vol. 2922 of *Lecture Notes in Artificial Intelligence*. Berlin, Heidelberg, New York: Springer-Verlag, pp. 1–17.

Jones, A. J. I. and M. Sergot (1992). "Formal Specification of Security Requirements Using the Theory of Normative Positions," In: Y. Deswarte, G. Eizenberg and J.-J. Quisquater (eds.) *Computer Security – ESORICS 92*, Vol. 648 of *Lecture Notes in Computer Science*. Springer-Verlag, pp. 103–121.

Jones, A. J. I. and M. Sergot (1996). "A Formal Characterisation of Institutionalised Power," *Journal of the Interest Group in Pure and Applied Logic*, 4, 427–443. Reprinted in E. Garzón Valdés et al. (eds.) *Normative Systems in Legal and Moral Theory – Festschrift for Carlos E. Alchourrón and Eugenio Bulygin*. Duncker & Humblot, Berlin, Germany, pp. 349–367.

Kanger, S. (1972). "Law and Logic," *Theoria* 38, 105–132.

Kimbrough, S. O., A. S. Abrahams, A. J. I. Jones, D. Eyers and J. Bacon (2005). "Introducing the Fair and Logical Trade Project," In: C. Bartolini, G. Governatori and Z. Milosevic (eds.) *Proceedings of the Workshop on Contract Architectures and Languages (CoALa 2005)*. Enschede, Holland: IEEE Press.

Labrou, Y. and T. Finin (1997). "Semantics and Conversations for an Agent Communication Language," In: M. E. Pollack (ed.) *Proceedings of the Fifteenth International Joint Conference on Artifical Intelligence (IJCAI-97)*. Nagoya, Japan: Morgan Kaufmann Publishers, pp. 584–591.

Levi, I. (1996). "Review of: Making it Explicit," *Journal of Philosophy* 93, 145–158.

Lewis, D. K. (1969). *Convention: A Philosophical Study*. Cambridge: Harvard University Press.

Lin, F., D. H. Norrie, W. Shen and K. Kremer (2000). "A Schema-based Approach to Specifying Conversation Policies," In: F. Dignum and M. Greaves (eds) *Issues in Agent Communication* (pp. 193–204). Holland: Springer-Verlag.

Litman, D. J. and J. F. Allen (1990). "Discourse Processing and Commonsense Plans," In: P. R. Cohen, J. Morgan and M. E. Pollack (eds) *Intentions in Communication* (pp. 365–388). Cambridge: MIT Press.

Marx, M. (1996). "Dynamic Arrow Logic," In: M. Marx, M. Masuch and P. Polos (eds.), *Arrow Logic and Multi-Modal Logic*. CSLI Publications, pp. 109–129.

McBurney, P. and S. Parsons (2002). "Games that Agents Play: A Formal Framework for Dialogue Between Autonomous Agents," *Journal of Logic, Language, and Information – Special issue on logic and games* 11, 315–334.

McBurney, P., S. Parsons and M. Johnson (2002). "When are Two Protocols the Same?", In: *AAMAS Workshop on Agent Communication Languages and Conversational Policies (ACL-02)*, Bologna, Italy, pp. 71–78.

Parunak, H. V. (1996). "Visualizing Agent Conversations: Using Enhanced Dooley Graphs for Agent Design and Analysis," In: *Proceedings of the Second International Conference on Multi-agent Systems (ICMAS96)*, AAII Press, pp. 275–282.

Pörn, I. (1970). *The Logic of Power*. Dordrecht: D. Reidel Publishing Company.

Pörn, I. (1977). *Action Theory and Social Science*. Dordrecht: D. Reidel Publishing Company.

Pörn, I. (1989). "On the Nature of a Social Order," In: J. F. Fenstad, et al (eds.) *Logic, Methodology and Philosophy of Science VIII* (pp. 553–567). New York: Elsevier Science Publishers B.V.

Pitt, J. (1999). "An Operational Semantics for Intentional Specifications of Agent Behaviour in Communication Protocols," In: M. Greaves (ed.), *Proceedings AA2000 Workshop on Agent Languages and Conversation Policies*. Barcelona: AAII Press, pp. 31–40.

Pratt, V. R. (1976). "Semantical Considerations on Floyd-Hoare logic," In: *Proc. 17th IEEE Symp. on Foundations of Computer Science*, pp. 109–121.

Reed, C. (1998). "Dialogue Frames in Agent Communication," In: *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS-98)*. New York: IEEE Press, pp. 246–253.

Santos, F., A. J. I. Jones and J. Carmo (1997). "Responsibility for Action in Organisations: A Formal Model," In: G. Holmström-Hintikka and R. Tuomela (eds.), *Contemporary Action Theory*, Vol. II of *Synthese Library 267*. Dordrecht, Holland: Kluwer Academic Press, pp. 333–350.

Searle, J. R. (1969). *Speech Acts – An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Searle, J. R. and D. Vanderveken (1985). *Foundations of Illocutionary Logic*. Cambridge: Cambridge University Press.

Segerberg, K. (1989). "Bringing it About," *Journal of Philosophical Logic* 18, 327–347.

Sergot, M. (1999). "Normative Positions," In: P. McNamara and H. Prakken (eds) *Norms, Logic and Information Systems* (pp. 289–308). Amsterdam: IOS Press.

Singh, M. P. (1998). "Agent Communication Languages: Rethinking the Principles," *IEEE Computer* 31(12), 40–47.

Smith, I. A., P. R. Cohen, J. M. Bradshaw, M. Greaves and H. Holmback (1998). "Designing Conversation Policies using Joint Intention Theory," In: *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS-98)*. IEEE Press: New York.

Stenius, E. (1967). "Mood and Language Game," *Synthese* 17, 254–274.

Benthem, J.van (1991). *Language in Action*. Amsterdam: The MIT Press.

van Benthem, J. (1994). "A Note on Dynamic Arrow Logics," In: J. van Eijck and A. Visser (eds) *Logic and Information Flow* (pp. 15–29). Cambridge: MIT Press.

van Benthem, J. (1996). *Exploring Logical Dynamics*. Stanford, California: CSLI Publications.

de Venema, Y. (1994). "A Crash Course in Arrow Logic," In: M. Marx, M. Masuch and P. Polos (eds.) *Arrow Logic and Multi-modal Logic*. CSLI Publications, pp. 3–34.

Verdicchio, M. and M. Colombetti (2004). "A Logical Model of Social Commitment for Agent Communication," In: F. Dignum (ed.). *Advances in Agent Communication, Vol. 2922 of Lecture Notes in Computer Science*. Springer, pp. 128–145.

Walton, D. N. and E. C. W. Krabbe (1995). *Commitment in Dialogue*. Albany: State University of New York Press.