Conventional signalling acts and conversation

Andrew J.I. Jones[1] and Xavier Parent[2]

[1] Department of Computer Science, King's College London, The Strand, London WC2R 2LS, UK,
ajijones@dcs.kcl.ac.uk,
[2] Department of Computer Science, King's College London, The Strand, London WC2R 2LS, UK,
xavier@dcs.kcl.ac.uk

Abstract. This article aims to provide foundations for a new approach to Agent Communication Languages (ACLs). First, we outline 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 outline one way 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 a conversational protocol.

1 Introduction

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

- those that are intention-based or mentalistic. Inspired by Grice [14], these approaches focus on the effects communicative acts have on participants' mental states (see e.g. [30, 20]);
- those that are commitment-based, in that they assign a key role to the notion of commitment (see e.g. [36, 29, 9]).

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 [14, 22, 26, 3]. 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 Language (ACL) 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 [29] and Colombetti [9]. However, there are also

some strong reasons to believe that that account too is 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 [36], and by Brandom [4, 5]. However, in line with Levi [21], 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.[3] Does that mean that j is not making an assertion? If not, 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 regulation, 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 paper is structured as follows. Section 2 outlines the basic assumptions and intuitions which motivate the theory of conventional signalling acts. Section 3 outlines one way 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 a conversational protocol.

2 Conventional signalling acts

The account of signalling acts outlined in this section bases the characterisation of communicative action neither on the intentions of communicators, nor on their commitments, but rather on the publically accessible conventions the use of which makes possible the performance of meaningful signalling acts. Consideration, first, of the communicative act of asserting will serve as a means of presenting the basic assumptions and intuitions which guide this approach.

2.1 Indicative signalling systems

The term 'indicative signalling system' is here used to refer to a signalling system in which acts of asserting can be performed. Such systems are 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

[3] For instance, Levi gives the example of a teacher explaining a thesis to a group of students.

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 [26], 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 [31].) 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 relativised 'counts as' conditional (see, for a detailed formal account, [19]), 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'.[4] 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

[4] The logic of the relativised action operator is given in [19] and [17]. The best available introduction to this kind of approach to the logic of agency is to be found in [11].

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

$$\text{(sc-assert)} \qquad E_jC \Rightarrow_s I_s^\star A \tag{1}$$

where $I_s^\star$ is a relativised optimality, or ideality, operator (a normative operator of the evaluative kind[5]), $I_s^\star A$ 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.

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. Secondly, 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. Thirdly, we assume that the members of U are all mutually aware of the (sc-assert) conditionals associated with s.[6]

2.2 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:

$$((E_jC \Rightarrow_s I_s^\star A) \wedge B_kE_jC) \rightarrow B_kI_s^\star A \tag{2}$$

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,

[5] On the distinction between *evaluative* and *directive* normative modalities, see [17]. For the logic of the $I_s^\star$ operator we adopt a (relativised) classical modal system of type EMCN. As is shown in [8], a classical system of this type is identical to the smallest normal system K. For details, see [17].

[6] See [17] for some remarks on the analysis of mutual belief.

but *if* he does so he will then also go on to form the belief that A. In summary, assuming (sc-assert) and (2), and supposing that

$$B_k E_j C \tag{3}$$

it now follows that

$$B_k I_s^\star A \tag{4}$$

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

$$B_k A \tag{5}$$

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:

$$B_k B_j A \tag{6}$$

Note the various possibilities here: k might trust neither the reliability nor the sincerity of j's assertion, in which case neither (5) nor (6) holds. Alternatively, k might trust j's sincerity without trusting the reliability of his assertion ((6), but not (5)), or k might trust the reliability of j's assertion without trusting j's sincerity ((5) but not (6)). 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.[7]

It can readily be seen that, in contrast to the approach advocated in the FIPA COMMUNICATIVE ACT LIBRARY SPECIFICATION [XC00037G, 29.01.2001][8], 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 $E_j C$, intends thereby to produce in k one or both of the beliefs (5) and (6). Indeed the only background assumption about the communicator's intention that is implicit in this account is that k, when forming the belief

[7] 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., [18] and [28] 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 [16].

[8] See http://www.fipa.org/

represented by (4), 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 (2) 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 publically 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.[9]

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*.

2.3 Commitment

Some recent approaches to ACLs have assigned a key role to the notion of *commitment* (e.g., Singh in [29] and Colombetti in [9]), 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 $I_s^\star$ 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.

[9] 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. This is not the place to enter into that discussion. Suffice it to say that 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.

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 [13] 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.

2.4 Some other types of communicative acts

Asserting is of course just one type of communicative act among many. This section provides just a sketch of some other types, and certainly does not pretend to give anything like an exhaustive characterisation of communicative act-types. But it does illustrate the flexibility and expressive power of the logical framework here employed. We consider four types:

- Commands
- Commissives (placing oneself under an obligation, e.g., promising)
- Requests
- Declaratives (in the sense of Searle & Vanderveken, [27])

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 be suggested in section 2.5, there is nevertheless one very important difference between pure assertives and these sub-species.

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

is:

$$\text{(sc-command)} \qquad E_jC \Rightarrow_s I_s^{\star}OE_kA \tag{7}$$

where the 'O' operator is a directive normative modality representing *obligation*. 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.

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

$$\text{(sc-commit)} \qquad E_jC \Rightarrow_s I_s^{\star}OE_jA \tag{8}$$

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.[10]

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

$$\text{(sc-request)} \qquad E_jC \Rightarrow_s I_s^{\star}H_jE_kA \tag{9}$$

where the relativised 'H' operator represents the modality 'attempts to see to it that...'. 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.[11]

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 [19], 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.

[10] On the logic of the directive normative modality, see [17].
[11] On the logic of the 'attempts to see to it that...' modality, see [17].

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:

$$\text{(sc-declare)} \qquad E_j C \Rightarrow_s I_s^\star E_j A \tag{10}$$

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 sees 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.

2.5 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.[12] 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.[13]

We have now outlined 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.

---

[12] 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.

[13] 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 [15] for references.

3 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 [32–34] and colleagues [35, 24].

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.[14]

The reason why we do not use dynamic logic in its traditional form (see Pratt [25]), 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, 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. Exploration of this second possibility is the main focus of our current investigations. 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.

3.1 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" (or "circle");
- 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 turning the static account into a dynamic 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] As far

[14] Of course, the account outlined in this paper can only be suggestive of how future work should proceed. For instance, the account says nothing about the specific criterion the agent should apply in choosing which utterance will constitute its next contribution. For a discussion of this issue, see P.E. Dunne and P. McBurney's contribution to this volume ([10]).

[15] In this approach, arrows are not required to have some particular internal structure (to be "ordered pairs", for instance).

as the 'dynamification' of the static framework is concerned, it then suffices to keep the package of truth-clauses already employed in the static framework, and to introduce those usually used for the three new building blocks.

The full account of the framework will be the focus of attention in a longer report on this work. Here we will characterise the arrow formalism only in terms of its proof-theory, and in terms of the graphics which help to give an intuitive account of the three new building blocks. Semantically, the introduction of these modalities is straightforward, by adding relations between arrows.

For instance, the evaluation rule for $\circ$ ("circle") states 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:

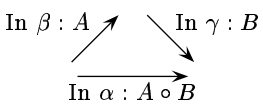

Fig. 1. Composition

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 follows:

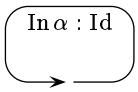

Fig. 2. Identity

Finally, the truth-clause for $\breve{}$ ("reverse") says that, for $A\breve{}$ to be true at α, there must be an arrow β that is the reversal of α and at which A holds:

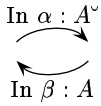

Fig. 3. Reverse

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 [33], these are designed for reasoning about both states and transitions. It seems very natural to try to refine the formalism in such a way that what obtains within states is also taken into account. We shall explore this issue in future research.

We now turn to the axiomatic characterization of the framework. When no particular constraints are imposed on the semantical counterparts of the dynamic operators, the proof theory of the integrated framework can in fact be obtained by adding the following rules of inference and axiom schemata to the basic logic:

Rules of inference

$$\frac{\vdash B \rightarrow C}{\vdash (A \circ B) \rightarrow (A \circ C)}\,(\text{r1}) \qquad \frac{\vdash A \rightarrow C}{\vdash (A \circ B) \rightarrow (C \circ B)}\,(\text{r2}) \qquad \frac{\vdash A \rightarrow B}{\vdash (A)^{\smile} \rightarrow (B)^{\smile}}\,(\text{r3})$$

$$\frac{\vdash A}{\vdash \neg(\neg A \circ B)}\,(\text{r4}) \qquad \frac{\vdash A}{\vdash \neg(B \circ \neg A)}\,(\text{r5}) \qquad \frac{\vdash A}{\vdash \neg((\neg A)^{\smile})}\,(\text{r6})$$

Axiom schemata

$$\vdash (A \vee B) \circ C \rightarrow (A \circ C) \vee (B \circ C) \qquad (\text{a1})$$
$$\vdash A \circ (B \vee C) \rightarrow (A \circ B) \vee (A \circ C) \qquad (\text{a2})$$
$$\vdash (A \vee B)^{\smile} \rightarrow A^{\smile} \vee B^{\smile}. \qquad (\text{a3})$$

The first three rules express a principle of closure under logical consequence. The next three are the arrow counterparts of the necessitation rule. Axioms (a1)-(a3) say that $\circ$ and $^{\smile}$ distribute over $\vee$. The converse of implication (a1) can easily be derived by using (r2), and similarly for (a2) and (a3).

A proof of soundness and (strong) completeness for the extended framework will be included in a longer report on our work, in preparation. The proof is based on the standard technique of canonical model construction.

3.2 The English Auction Protocol

In this section, we illustrate the expressive capacity of the logic, by showing how it can be applied to the analysis of a conversational protocol. We focus on what are called English auctions, at least as a starting point. We here give the basic idea of the treatment.

Figure 4 below 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 conversation. The circles with a double-line represent the final states of the conversation.

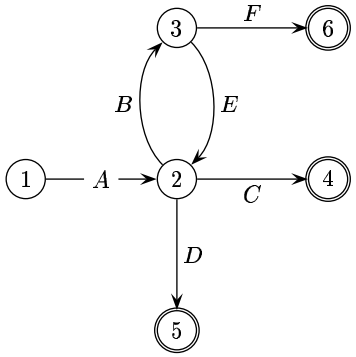

Fig. 4. English Auction Protocol

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 [6]). [16] C, E and F each correspond to the antecedent of a signalling convention taking the form of (sc-assert).

[16] The formalization of contrary-to-duty (CTD) scenarios raises a problem that is usually considered as a hard one by deontic logicians. We note in passing that the concept of norm violation has an obvious counterpart in commitment-based approaches

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

$$(A \circ D) \vee (A \circ C) \vee (A \circ (B \circ F)) \vee (A \circ (B \circ (E \circ C))) \tag{11}$$

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 $\vee$, (11) can be simplified into

$$A \circ \big(D \vee C \vee (B \circ (F \vee (E \circ C)))\big) \tag{12}$$

(11) considers in isolation the sequences of acts that are allowed by the protocol. The first disjunct in (11), $A \circ D$, 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 paths 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 (12) puts the sequences of speech acts together, and indicates the points when interactants have the opportunity to choose between two or more speech acts. (12) 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 (11) 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 the fact that users of signalling system s are empowered agents, or the fact that one agent j trusts some other agent k. 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 a construction proposed by Fitting [12]

$$S \models U \rightarrow X \tag{13}$$

which exploits the idea that the local and the global consequence relations used in modal logic can be subsumed under one more general relation. The formal definition of this notion can easily be adapted to the present framework. Intuitively, S expresses global assumptions, holding at all arrows. In contrast, U enumerates

---

to conversation. In particular, Mallya et al.'s contribution to this volume (see [23]) puts some emphasis on the notion of breach of a commitment in a conversation. It would be interesting to investigate what such frameworks have to say about CTD scenarios.

local assumptions, holding at particular arrows. In line with our previous analysis – see section 2.1 – 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:

$$E_a A_1 \Rightarrow_s I_s^\star E_a A_4 \tag{14}$$

$$E_b A_2 \Rightarrow_s I_s^\star O(E_b A_6 / A_5) \tag{15}$$

$$E_a A_3 \Rightarrow_s I_s^\star A_7 \tag{16}$$

Now let us adopt the point of view of an external observer x. This means that we can specify U in (13) as

$$B_x E_a A_1 \circ (B_x E_b A_2 \circ B_x E_a A_3) \tag{17}$$

As can easily be verified, the doxastic form of modus ponens (2) used in the 'static' framework allows us to specify X in (13) as

$$B_x I_s^\star E_a A_4 \circ (B_x I_s^\star O(E_b A_6 / A_5) \circ B_x I_s^\star A_7), \tag{18}$$

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:

$$B_x E_a A_4 \circ (B_x O(E_b A_6 / A_5) \circ B_x A_7). \tag{19}$$

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. (In this case, all b sees is the new announcement.). 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 $\star$ and referred to as "iteration" (also "Kleene star"). We defer the full discussion of this issue to another occasion.

4 Concluding remarks

Although the 'dynamic' account outlined in this paper is preliminary, we believe it points the way to a comprehensive theory of conversation, providing 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, we are inclined to think that this kind of representation will be able to facilitate the systematic comparison of protocols.

In closing, let us add one further remark in connection with the second suggestion we have made. It is that the record process should 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 MIT's Kasbah prototype [7]. 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

$$(E_k C \Rightarrow_s I_s^\star E_k E_j A) \wedge (E_j A \Rightarrow_s I_s^\star E_j B) \tag{20}$$

If k is empowered to make an offer (if, for instance, a time-out has not taken place), then the truth of $E_k E_j A$ (and, hence, the truth of $E_j A$) 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 $E_j B$ also obtains. Here the idea is to classify the performance of a communicative act as valid or invalid according to whether or not the agent that performed the action was institutionally empowered. Some work along these lines has already been conducted in the context of the study of the Contract-Net-Protocol (see Artikis et al. [1, 2]). A thorough investigation of the relation between that account and the one outlined in this paper remains to be done.

References

1. A. Artikis, J. Pitt, and M. Sergot. Animated specifications of computational societies. In C. Castelfranchi and L. Johnson, editors, *Proceedings of Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 1053–1062, 2002.
2. A. Artikis, M. Sergot, and J. Pitt. Specifying electronic societies with the Causal Calculator. In F. Giunchiglia, J. Odell, and G. Weiss, editors, *Proceedings of Workshop on Agent-Oriented Software Engineering III (AOSE)*. Springer, 2003.
3. J. Bennett. *Linguistic Behavior*. Cambridge, 1976.
4. R. Brandom. Asserting. *Nous*, 17:637–650, 1983.
5. R. B. Brandom. *Making it Explicit - Reasoning, Representing, and Discursive Commitment*. Harvard University Press, London, England, 1994.
6. J. Carmo and A.J.I. Jones. Deontic logic and contrary-to-duties. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 8, pages 265–343. Kluwer Academic Publishers, Dordrecht, Holland, 2nd edition, 2002.

7. A. Chavez and P. Maes. 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)*, pages 76–90, London, 1996. UK.
8. B.F. Chellas. *Modal Logic - an introduction.* Cambridge University Press, Cambridge, 1980.
9. M. Colombetti. A commitment-based approach to agent speech acts and conversations. In *Proceedings of the Workshop on Agent Languages and Conversation Policies*, pages 21–29, Barcelona, 2000. (Held at the *'Agents 2000'* conference).
10. P.E. Dunne and P. McBurney. Concepts of optimal utterance in dialogue: Selection and complexity. In F. Dignum, editor, *Advances in Agent Communication*, LNAI, page in this volume. Springer Verlag.
11. D. Elgesem. *Action Theory and Modal Logic.* PhD thesis, Institutt for filosofi. Det historik-filosofiske fakultetet. Universitetet Oslo, 1993. See also Elgesem's paper 'The modal logic of agency'. In *The Nordic Journal of Philosophical Logic*, vol. 2, 1997, pp. 1-46.
12. M. Fitting. Basic modal logic. In Hogger C. Gabbay D.M. and Robinson J., editors, *Handbook of logic in Artificial Intelligence and Logic Programming.* Oxford University Press, Oxford, 1983. Volume 1.
13. D. Follesdal. Comments on Stenius's 'Mood and language game'. *Synthese*, 17:275–280, 1967.
14. H.P. Grice. Meaning. *The Philosophical Review*, 66:377–388, 1957.
15. A.J.I. Jones. *Communication and Meaning - an essay in applied modal logic.* D. Reidel, Dordrecht - Holland, 1983.
16. A.J.I. Jones. On the concept of trust. *Decision Support Systems*, 33:225–232, 2002.
17. A.J.I. Jones. A logical framework. In J. Pitt, editor, *The Open Agent Society*, chapter 3. John Wiley and Sons, Chichester, UK, 2003. To appear.
18. A.J.I. Jones and M. Sergot. Formal specification of security requirements using the theory of normative positions. In Y. Deswarte, G. Eizenberg, and J.-J Quisquater, editors, *Computer Security - ESORICS 92*, Lecture Notes in Computer Science 648, pages 103–121. Springer-Verlag, 1992.
19. A.J.I. Jones and M. Sergot. A formal characterisation of institutionalised power. *Journal of the Interest Group in Pure and Applied Logic*, 4:427–443, 1996. 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, 1997, pages 349-367.
20. Y. Labrou and T. Finin. Semantics and conversations for an agent communication language. In Martha E. Pollack, editor, *Proceedings of the Fifteenth International Joint Conference on Artifical Intelligence (IJCAI-97)*, pages 584–591, Nagoya, Japan, 1997. Morgan Kaufmann publishers.
21. I. Levi. Review of: Making it explicit. *Journal of Philosophy*, 93:145–158, 1996.
22. D.K. Lewis. *Convention: a Philosophical Study.* Harvard University Press, Cambridge, 1969.
23. A.U. Mallya, P. Yolum, and M.P. Singh. Resolving commitments among autonomous agents. In F. Dignum, editor, *Advances in Agent Communication*, LNAI, page in this volume. Springer Verlag.
24. M. Marx. Dynamic arrow logic. In M. Marx, M. Masuch, and P. Polos, editors, *Arrow Logic and Multi-Modal Logic*, pages 109–129. CSLI Publications, 1996.
25. V.R. Pratt. Semantical considerations on Floyd-Hoare logic. In *Proc. 17th IEEE Symp. on Foundations of Computer Science*, pages 109–121, 1976.

26. J.R. Searle. *Speech Acts - an essay in the philosophy of language.* Cambridge University Press, Cambridge, 1969.
27. J.R. Searle and D. Vanderveken. *Foundations of illocutionary logic.* Cambridge University Press, Cambridge, 1985.
28. M. Sergot. Normative positions. In P. McNamara and H. Prakken, editors, *Norms, Logic and Information Systems*, pages 289–308. IOS Press, Amsterdam, 1999.
29. M.P. Singh. Agent communication languages: rethinking the principles. *IEEE Computer*, 31(12):40–47, 1998.
30. I. A. Smith, P. R. Cohen, J. M. Bradshaw, M. Greaves, and H. Holmback. Designing conversation policies using joint intention theory. In *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS98)*, 1998.
31. E. Stenius. Mood and language game. *Synthese*, 17:254–274, 1967.
32. J. van Benthem. *Language in Action.* The MIT Press, Amsterdam, 1991.
33. J. van Benthem. A note on dynamic arrow logics. In J. van Eijck and A. Visser, editors, *Logic and Information Flow*, pages 15–29. MIT Press, Cambridge, 1994.
34. J. van Benthem. *Exploring Logical Dynamics.* CSLI Publications, Stanford, California, 1996.
35. Y. Venema. A crash course in arrow logic. In M. Marx, M. Masuch, and P. Polos, editors, *Arrow Logic and Multi-Modal Logic*, pages 3–34. CSLI Publications, 1994.
36. D.N. Walton and E.C.W. Krabbe. *Commitment in Dialogue.* State University of New York Press, 1995.

Index

Name Index