A model of delegation for multi-agent systems

Timothy J. Norman[†] Chris Reed[‡]

[†] Department of Computing Science, University of Aberdeen,
Aberdeen, AB24 3UE, Scotland, U.K.
tnorman@csd.abdn.ac.uk

[‡] Department of Applied Computing, University of Dundee,
Dundee, DD1 4HN, Scotland, U.K.
chris@computing.dundee.ac.uk

Abstract. The act of delegating a task from one agent to another can be carried out through the performance of one or more imperative communication acts. In this paper, the semantics of such imperatives are specified using a language of actions and states. The logical system that is developed then supports a notion of responsibility. An agent may not only be issued an imperative to directly carry out an event, or achieve some state, but also to be responsible for an event being carried out or state achieved - and these latter commitments might then be serviced through a subsequent act of delegation. The model thus clearly distinguishes between different classes of responsibility and different forms of delegation, and it is shown how this sound theoretical foundation can then be applied in specifying the semantics of imperatives in agent communication languages.

1 Introduction

To delegate is to entrust a representative to act on your behalf. This is an important issue for agents that may be forced to rely on others. Although autonomous agents have a high degree of self-determination, they may be required to achieve a goal that is made easier, satisfied more completely or only possible with the aid of other, similarly autonomous, agents. For delegation to be successful, there must be a relationship between the agent delegating the goal or task and the agent to whom it is delegated. Furthermore, after successful delegation, responsibility for the task concerned is now shared. For example, the manager of a business unit, in delegating a task, will no longer be solely responsible for that task. The manager must, however, ensure that the employee to whom the task has been delegated acts appropriately (e.g. by completing the task, asking for help or further delegating the task).

This is, according to Castelfranchi and Falcone [5], *strong delegation*; i.e. where there is mutual awareness and social commitment. In particular, Castelfranchi and Falcone [5, 4], amongst others [24], address the question: What is the nature of the relationship on which the ability to delegate is predicated? Here, however, we focus on the specification of communicative acts that can be used to delegate tasks, and on a number of the dimensions that characterise what is being delegated and what are the conditions under which it can be said that delegation was successful. We focus on two key dimensions:

1. the distinction between the delegation of *actions to be performed* and *states of affairs to be achieved*; and
2. the restrictions on the further delegation of the activity concerned (i.e. whether further delegation is *permitted*, *forbidden* or *required*.

There are, of course, many more issues that must be considered in the development of a complete theory of delegation; for example, the discussion is restricted to individuals rather than groups of agents [19, 23]. However, considering these two issues alone does mean that a number of distinct situations must be considered:

(a) "I don't care who achieves (state of affairs) A, but it must be achieved."
(b) "I don't care who does (action) α, but it must be done."
(c) "You must, through your own, direct, intervention achieve A."
(d) "You must, through your own, direct, intervention do α."
(e) "You must ensure that A is achieved by someone other than yourself."
(f) "You must ensure that α is done by someone other than yourself."

Possibly the most common mechanism employed in the delegation of activity (especially in multi-agent systems) is through direct communication. Knowledge level [22] communication between agents through speech act based [2, 28] agent communication languages (ACLs) is both an active area of research [31, 26] and of standardisation [11, 10]. These languages typically include *indicatives* (or assertions) such as 'tell' (KQML) and 'inform' (FIPA); for example, "It's raining". Queries or questions are also common (i.e. *interrogatives*) such as 'ask-if' (KQML) and 'query-if' (FIPA); for example, "Is it raining?". In addition to these, *imperatives* are used to issue commands, give advice or request action; for example, "bring the umbrella". Examples of imperative message types in ACLs are 'achieve' (KQML) and 'request' (FIPA). An intuitive explanation of these message types is that the sender is attempting to influence the recipient to act in some way. In fact, an attempt to influence the mental state of the hearer (or recipient of a message) is common to all knowledge level communication. For example, an agent utters an indicative such as "It's raining" with the intention of inducing a belief by means of the recognition of this intention [12]. In other words, the speaker is attempting to influence the hearer to adopt a belief about the weather.

Similarly, the imperative "bring the umbrella" is an attempt to influence the hearer's future actions by means of the hearer recognising that this is the intention of the speaker. Following Searle's [28] description of the types of illocutionary (or communicative) act, Cohen and Levesque [8] provide a model in which such acts are construed as attempts by the speaker to change the mental state of the hearer. For example, a request for the hearer to do some action, α, is an attempt to change the hearer's mental state in such a way that it becomes an intention of the hearer to do α. The hearer, being an autonomous agent, may refuse. This is, of course, different from misunderstanding the speaker.

With this in mind, Cohen and Levesque [8] distinguish between a speaker's goal of performing an act, and the speaker's intention behind the act. The goal, in the case of an imperative, is that the hearer believes that the speaker intends the hearer to act and the hearer acts accordingly. The intention, by contrast, is only that the hearer believes that the speaker had that communicative goal. If the intention of the speaker is not understood by the hearer, then the communicative act is unsuccessful. A communicative act

is then an attempt to achieve the goal, but at least to satisfy the intention that the hearer believes that this is what the speaker wants. Through this definition of 'attempts', Cohen *et al.* [8, 31] provide a concrete characterisation of the communicative acts that are common in ACLs, and go on to specify conversations. For example, one agent offers a service to another, which may be responded to with acceptance, rejection or silence (cf. Barbuceanu and Fox [3]). This extension of an agent communication language to capture typical conversations between agents is the approach taken by the FIPA specification [11], and it is the specification of imperatives within FIPA that is returned to in section 6, where we return to the dimensions introduced above to discuss to what extent existing standards of agent communication are able to express them.

The grounding of agent communication languages in such formal models is essential to ensure that the meaning of communicative acts are clear to those designing agents for practical applications. Without such a grounding, agent communication languages can suffer from inherent ambiguity which, when implemented, can lead to unexpected, undesirable and counter-intuitive results. Although we do require that specifications are unambiguous, it is essential that ACLs support rich dialogues between agents, and so analyses of the possible dimensions of the semantics of communication are important [27]. The work presented in this paper focuses on imperatives, aims to present an account of delegation, and show how this may be better understood by considering both existing models of imperatives [13, 33] and normative positions [20, 29]. Before presenting a formal model of agentative action (sections 3 and 4) upon which the model of delegation proposed in this paper is built, it is important to discuss imperatives in more detail.

2 Imperatives

Numerous proposals have been laid out in both philosophical and computational literature for classification of utterance types, or, more specifically, of illocutionary acts. Austin [2, p. 150] and Searle [28, pp. 66–67] are perhaps the two most prominent.

Though there are a range of similarities and dissimilarities, these schemes have at least one thing in common: not all utterances are indicative. This is not in itself remarkable, until it is considered that the logics employed to handle and manipulate utterances are almost always exclusively based upon the predominant formal tradition of treating only the indicative. The interrogative and imperative utterances (which figure amongst Austin's Exercitives and Expostives, and include Searle's Request, Question, Advise and Warn) rarely benefit from the luxury of a logic designed to handle them.

Interrogative logics for handling questions have been proposed by Åqvist [1] and Hintikka *et al.* [15] among others, and these form an interesting avenue for future exploration. The focus of the current work, however, is on imperative logic. Hamblin's [13] book *Imperatives* represents the first thorough treatment of the subject, providing a systematic analysis not only of linguistic examples, but also of grammatical structure, semantics and the role imperatives play in dialogue.

His classification goes into some detail, but one key distinction is drawn between imperatives which are *wilful*, and those which are not. The former class are characterised by advantage to the utterer, the latter by advantage to the hearer. Thus com-

mands, requests, and demands are all classed as wilful; advice, instructions, suggestions, recipes and warnings are all classed as non-wilful.

The distinction is useful because it highlights the importance of the contextual environment of the utterance: commands would fail to have an effect if the utterer was not in a position of authority over the hearer; advice would fail if the hearer did not trust the utterer, and so on. Any logic of imperatives must both be able to cope with this wide range of locutionary acts, but also be insensitive to any of the extralinguistic (and thereby extra-logical) factors affecting the subsequent effect of issuing imperatives.

Hamblin states [13, p. 137] that to handle imperatives there are several features, "usually regarded as specialised", which are indispensable for a formal model: (1) a time-scale; (2) a distinction between actions and states; (3) physical and mental causation; (4) agency and action-reduction; and (5) intensionality. Following the second feature listed above, both events and states of affairs are explicitly represented in the Action State Semantics: a world is a series of states connected by events. The states can be seen as collections of propositions. Events are of two types: deeds, which are performed by specific agents, and happenings, which are world effects. This distinction gives the model an unusual richness: most other formal systems have explicit representation of one or the other, defining either states in terms of the sequences of events (true of most action and temporal logics), or else events in terms of a succession of states such as in classical AI planning.

The situation calculus [21] allows both states and events to be represented, but the commonly adopted "axioms of arboreality" [30] restrict the flexibility so that a given sequence of events is associated with a single, unique situation. Even if all the fluents in two situations have identical values, under the axioms of arboreality, those two situations are only the same if the events leading to them have also been the same. In Hamblin's work, however, there can be several different histories up to a given state and the histories are not themselves a part of those states.

This rich underlying model is important in several respects. First, it allows, at a syntactic level, the expression of demands both that agents bring about states of affairs, and that they perform actions. Secondly, it avoids both ontological and practical problems of having to interrelate states and events — practical problems often become manifest in having to keep track of 'Done events' in every state [9]. Finally, this construction of a world as a chain of states connected by deeds and happenings makes it possible to distinguish those worlds in which a given imperative i is satisfied (in some set of states). Thus the imperative "Shut the door" is satisfied in those worlds in which the door is shut (given appropriate deixis). This 'extensional' satisfaction, however, is contrasted with a stronger notion, of 'whole-hearted' satisfaction, which characterises an agent's involvement and responsibility in fulfilling an imperative. Whole-hearted satisfaction is based upon the notion of a strategy. A strategy for a particular agent is the assignment of a deed to each time point. A partial i-strategy is then a set of incompletely specified strategies, all of which involve worlds in which i is extensionally satisfied. The whole-hearted satisfaction of an imperative i by an agent x is then defined as being x's adoption of a partial strategy and the execution of a deed from that strategy at every time point after the imperative is issued.

A Hamlinian world $w \in W$ is defined such that for every time point in T there is:

1. a state from the set of states S,
2. a member of the set H of 'big happenings' (each of which collect together all happenings from one state to the next), and
3. a deed (in D) for every agent (in X), i.e. an element from D^X.

The set W of worlds is, therefore, defined as $(S \times H \times D^X)^T$. The states, happenings and deed-agent assignments of a given world w are given by $S(w)$, $H(w)$ and $D(w)$.

Let j_t be a history of a world up to time t, including all states, deeds and happenings of the world up to t. Thus j_t is equivalent to a set of worlds which have a common history up to (at least) time t. J_t is then the set of all possible histories up to t; i.e. all the ways by which the world could have got to where it is. A strategy q_t is then an allocation of a deed to each $j_{t'} \in J_{t'}$ for every $t' \geq t$.[1]

Let the possible worlds in which the deeds of agent x are those specified by strategy q_t be $W_{strat}(x, q_t)$, and the worlds in which an imperative, i, is extensionally satisfied be W_i. A strategy for the satisfaction of an imperative i (i.e. an i-strategy) can, therefore, be defined as follows: A strategy $q_t \in Q_t$ is an i-strategy for agent x if and only if the worlds in which x does the deeds specified by q_t are also worlds in which i is extensionally satisfied: $W_{strat}(x, q_t) \subseteq W_i$.

In practice, however, it is not feasible for an agent to select a particular strategy in Q_t at time t that specifies every deed for every time t' after t. For this reason, an agent will adopt a *partial* i-strategy. A partial i-strategy is a disjunction of i-strategies, $Q'_t \subseteq Q_t$, and the world set for x adopting this partial i-strategy is $W_{strat}(x, Q'_t)$.

With this grounding, the whole-hearted satisfaction of an imperative, i, can now be defined. An agent x may be said to whole-heartedly satisfy an imperative i issued at t if and only if for every $t' \geq t$:

1. x has a partial i-strategy, $Q'_{t'}$; and
2. x does a deed from the set of deeds specified by that $Q'_{t'}$.

In addition to Hamblin's monograph [13], more detail on the role of such a model in the wider context of dialogue and a more complete set-theoretic précis is given by Walton and Krabbe [33]. Here, motivated by the work outlined in this section, we present a theory of agentative activity that is appropriate for modelling imperatives. First, an axiomatisation of the two action modalities S and T is presented. Secondly, we develop a possible worlds semantics for these operators as an approximation to the Action State Semantics. Finally, we go on to show how this theory can be used to model imperatives in agent communication.

3 Axiomatisation of S and T

With the intuitive grounding in Hamblin's Action State Semantics provided by the previous section, we present a syntax that may be used to explicitly refer to agents performing actions and achieving goals. We refer to actions (or deeds) by the symbols

[1] This notion of a strategy has an intensional component, since it prescribes over a set of possible w, rather than picking out, at this stage, the actual world.

RES	$\dfrac{A \leftrightarrow B}{\mathsf{S}_x A \leftrightarrow \mathsf{S}_x B}$
TS	$\mathsf{S}_x A \rightarrow A$
CS	$(\mathsf{S}_x A \wedge \mathsf{S}_x B) \rightarrow \mathsf{S}_x (A \wedge B)$
MS	$\mathsf{S}_x (A \wedge B) \rightarrow (\mathsf{S}_x A \wedge \mathsf{S}_x B)$
RS	$\mathsf{S}_x (A \wedge B) \leftrightarrow \mathsf{S}_x A \wedge \mathsf{S}_x B$
KS	$\mathsf{S}_x (A \rightarrow B) \rightarrow (\mathsf{S}_x A \rightarrow \mathsf{S}_x B)$
DS	$\mathsf{S}_x A \rightarrow \neg \mathsf{S}_x \neg A$

Fig. 1. Rules of inference & axiom schemas of S.

$\alpha, \beta, \ldots \in D$, states by $A, B, \ldots \in S$ and agents by $x, y, \ldots \in \mathcal{X}$. World effects, or happenings, are not considered explicitly here; we assume that there is a special agent that models world effects.

In the following discussion, a number of rules of inference and axiom schemas are considered. Those that are included in the logic of the modality S are summarised in figure 1 (these axioms are analogous for T, but do not represent a minimal set — they are listed exhaustively in the interests of clarity). A few others are given in figure 2 for the purposes of discussion, but are rejected for modality S (similarly, they are rejected for T).

The logic of the operators S and T is a regular modal logic [6]. As with other classical modal logics, both are closed under equivalence by the rules RE (see figure 1 for RES). Furthermore, following Jones and Sergot's exposition of their modality E_x, both S_x and T_x use the axiom schema T. The adoption of schema T can be justified on intuitive grounds by reading it as follows for modality S: if an agent sees to it that a state of affairs holds, then that state of affairs does, in fact, hold. Following Jones and Sergot, then, the current work develops a logic of *successful* action.[2] A similar gloss can be constructed for T_x — if an agent sees to it that an action is performed then that action is performed — but this implicitly requires stretching a possible-worlds interpretation as far as, and perhaps further, than is reasonable, as explained below.

One of the most fundamental disagreements between theories of agency concerns the rule of necessitation (RN for modality S is given in figure 2). This arises from a deep intuitive dilemma. The argument for adopting the reverse R$\neg$N proposed by Jones and Sergot is simply stated: "Whatever else we may have in mind ... on no account could we accept that an agent brings about what is logically true" [17, p. 435]. Thus Jones and Sergot, like Belnap and Perloff (whose *negative condition* entails R$\neg$N) are trying to capture some notion of responsibility, such that no agent can be said to be 'responsible' for a tautology. Chellas' intuitions, by contrast run rather differently. He

[2] This notion of "successful action" may be better viewed as "successful interaction with the world" considering our distinction between S and T. This alternative reading more clearly indicates that the formula to which the modality is applied is not in any way equivalent or *logically* related to the actions that an agent may carry out.

$$\mathrm{R}\neg\mathsf{NS} \quad \frac{A}{\neg \mathsf{S}_x A}$$

$$\mathrm{RNS} \quad \frac{A}{\mathsf{S}_x A}$$

$$\mathrm{RMS} \quad \frac{A \to B}{\mathsf{S}_x A \to \mathsf{S}_x B}$$

$$5\mathsf{S} \quad \neg \mathsf{S}_x A \to \mathsf{S}_x \neg \mathsf{S}_x A$$

$$4\mathsf{S} \quad \mathsf{S}_x A \to \mathsf{S}_x \mathsf{S}_x A$$

Fig. 2. Further candidate rules of inference and axiom schemas discussed.

is happy to accept RN, a much more conventional rule of a normal modal logic, and his argument too is tabled very briefly: "Can it ever be the case that someone sees to it that something logically true is so? I believe the answer is yes. When one sees to something, one sees to anything that logically follows, including the easiest such things, such as those represented by $\top$. One should think of seeing to it that, for example, $0 = 0$ as a sort of trivial pursuit, attendant upon seeing to anything at all." [7, p. 508]. Chellas' decision, in particular, is motivated by the logical consequences of the rule, and in particular on the availability of schemata C and M.

The *outward* distributivity of an action modality is adopted in the axiom schema C. Schema C is adopted by Chellas, Jones and Sergot, Belnap and Perloff, and, similarly, in the work presented here (see figure 1 for CS); it is difficult to argue from an intuitive basis how C might fail.

The *inward* distributivity axiom schema, M, however, is more troublesome. M, like C, seems intuitively appealing, but, for Jones and Sergot (and other systems adopting R¬N), it is pathological, since, with RE, it yields the rule RM (RMS is shown in figure 2). Taking the tautology $A \to \top$, RM gives $\mathsf{S}_x A \to \mathsf{S}_x \top$. Since R¬N gives $\neg \mathsf{S}_x \top$, any $\mathsf{S}_x A$ is thus a contradiction. Jones and Sergot, therefore, reject M because they are committed to the notion of responsibility captured by R¬N; Chellas on the other hand, accepts RN and, thereby, the loss of agentative responsibility, but does, as a result, maintain M.

The solution proposed for the modalities S and T represents a half-way house, eschewing both the restrictive nature of a (smallest) classical modal logic, and the counterintuitive results of a normal modal logic, in favour of a (smallest) regular modal logic. We also defer the issue of necessitation (versus "anti-necessitation") to the semantics. Both modalities thus include the rule RE and the axiom schema R (and, consequently, M, C and K), but they require neither the rule of necessitation (RN), nor the rule of anti-necessitation (R¬N).

The preceding discussion has already mentioned the intuitive appeal of M and C; it is also worth digressing to offer an intuitive gloss on the schema K to demonstrate its role, particularly as Jones and Sergot implicitly reject K. An imperative with the form of an implication is, linguistically, quite straightforward: "Make sure that if you go out then you lock the door". If an agent brings it about that the implication holds

then K states that if the agent brings about the antecedent then it is logically responsible also for bringing about the consequent. This does not impinge upon the autonomy of an agent to decide not to fulfil some imperative; rather, it states only that if the agent brought about the antecedent, then it can only also be said to have brought about the implication if it is responsible for the consequent.

The axioms 4 and 5 are commonly employed in mentalistic modalities, and, less frequently, in agentative modalities. First, consider schema 5S (figure 2). This is explicitly rejected for several reasons, not least of which is that with T, it would yield RN, which we wish to avoid. We return to the problems that 5 would throw up in the context of forbearance, section 5.1. Schema 5 is also rejected across the board by Jones and Sergot, Belnap and Perloff, and Chellas. Axiom schema 4, however, is accepted by Belnap and Perloff. Consider schema 4S (figure 2). With TS, this yields the following equivalence, which we reject: $\mathsf{S}_x A \leftrightarrow \mathsf{S}_x \mathsf{S}_x A$. The importance of avoiding this equivalence and the problems that 5 would present with respect to forbearance are discussed in section 5.1.

Finally, the adoption of T in the models of Jones and Sergot, of Belnap and Perloff, and of Chellas entails the inclusion of axiom schema D (see figure 1 for DS).

To summarise then, the logics of S_x and T_x are relativised classical regular modal logics of type RT [6, p. 237].

4 Semantic Model

As the axiomatisation indicates, the proposed logic is considerably smaller than a normal modal logic, and as a result, a standard model is inappropriate. To provide a possible-worlds semantics, we therefore use a minimal model [6].

The simplest approach is to define S_x (and T_x analogously) in the same style as a conventional modal logic. Thus with a model $\mathcal{M} = \langle \mathcal{W}, \mathcal{N}, \mathcal{P} \rangle$ with worlds $\mathcal{W}$, "necessitation function" $\mathcal{N}$, and interpretation functions abbreviated by $\mathcal{P}$, we can define the truth conditions of the unrelativised modality S. To characterise the *relativised* modality S_x, we introduce multiple necessitation functions, one for each agent $x \in \mathcal{X}$, thus $\mathcal{M} = \langle \mathcal{W}, \mathcal{N}^x, \mathcal{P} \rangle$. $\mathcal{N}^x$ maps from a given world ω, to a collection of sets of worlds (i.e. $\mathcal{N}^x \subseteq \wp(W)$), picking out those propositions which are brought about (by x) at ω. The standard truth conditions for propositional logic are captured in 1–8, and for the modality S_x in 9. (Note that, $\mathcal{P}$ abbreviates an infinite sequence, $\mathcal{P}_0, \mathcal{P}_1, \mathcal{P}_2, \ldots$, of subsets of $\mathcal{W}$, where, for each n, $\mathcal{P}_n$ represents those possible worlds in which the corresponding atomic sentence P_n holds — this is condition 1.)

$$\models^{\mathcal{M}}_{\omega} P_n \quad \text{iff} \quad \omega \in \mathcal{P}_n \text{ for } n = 1, 2, 3, \ldots \tag{1}$$

$$\models^{\mathcal{M}}_{\omega} \top \tag{2}$$

$$\not\models^{\mathcal{M}}_{\omega} \bot \tag{3}$$

$$\models^{\mathcal{M}}_{\omega} \neg A \quad \text{iff} \quad \not\models^{\mathcal{M}}_{\omega} A \tag{4}$$

$$\models^{\mathcal{M}}_{\omega} A \wedge B \quad \text{iff} \quad \models^{\mathcal{M}}_{\omega} A \text{ and } \models^{\mathcal{M}}_{\omega} B \tag{5}$$

$$\models^{\mathcal{M}}_{\omega} A \vee B \quad \text{iff} \quad \models^{\mathcal{M}}_{\omega} A \text{ or } \models^{\mathcal{M}}_{\omega} B \text{ or both} \tag{6}$$

$$\models^{\mathcal{M}}_{\omega} A \rightarrow B \quad \text{iff} \quad \text{if } \models^{\mathcal{M}}_{\omega} A \text{ then } \models^{\mathcal{M}}_{\omega} B \tag{7}$$

$$\models^{\mathcal{M}}_{\omega} A \leftrightarrow B \quad \text{iff} \quad \models^{\mathcal{M}}_{\omega} A \text{ if and only if } \models^{\mathcal{M}}_{\omega} B \tag{8}$$

$$\models^{\mathcal{M}}_{\omega} \mathsf{S}_x A \quad \text{iff} \quad \|A\|^{\mathcal{M}} \in \mathcal{N}^x_{\omega} \tag{9}$$

Unfortunately, quite apart from practical difficulties in using such a model as the basis for implementation of a multi-agent system [34] the approach fails to provide a good foundation upon which to develop an account of not just static states of affairs but of dynamic states, and of not just individual actions but of series of actions. These extensions are vital to any account of real agentative action, which has motivated works such as those of Chellas [7], Horty and Belnap [16] and others to adopt a much richer "metaphysical backdrop", substantially extending the Leibnizian model.

The development of a full semantics based on Action State Semantics is the subject of current research and is beyond the scope of this paper. A compromise between familiarity and accuracy can be achieved though enriching the possible-worlds approach by building in structure to each world that approximates the Action State Semantics (an analogous approach is adopted by many works founded on branching time logics [34]). Such a compromise serves as a sufficient foundation upon which to explore a rich characterisation of delegation. Thus, we can say that $j \angle_t \upsilon$ can be read as a history, j, of the Hamblinian world υ up to t; j is an initial segment of υ and υ is a completion of j (following Walton and Krabbe [33, p. 191]). This is defined recursively as follows:

$$j \angle_0 \upsilon = \langle \emptyset, s_0 \in S(\upsilon), \delta^x_0 \in D(\upsilon) \rangle$$
$$j \angle_t \upsilon = \langle j \angle_{t-1} \upsilon, s_t \in S(\upsilon), \delta^x_t \in D(\upsilon) \rangle$$

where the functions S and D map from a Hamblinian world, υ, to a set of propositions corresponding to the state of the world, $S(\upsilon)$, and to a set of deed assignments (agent-action pairs), $D(\upsilon)$.

In simplifying the semantics, it is possible to provide an interpretation of the S_x and T_x modalities that is irrespective of time (this simplification constitutes one of the major restrictions by comparison to the full Action State Semantics model under development). This timelessness is achieved through building an entire Kripke structure for a single time point, t. Thus each possible world in the Kripke structure can be seen as containing one particular $j \angle_t \upsilon$ for each Hamblinian world υ. So a model $\mathcal{M}$, is defined as $\langle \mathcal{W}, \mathcal{X}, \mathcal{I}, \mathcal{S}^x, \mathcal{T}^x \rangle$ for a set of possible worlds $\mathcal{W}$, a set of agents $\mathcal{X}$, an interpretation function $\mathcal{I}$, and sets of functions $\mathcal{S}^x$ and $\mathcal{T}^x$ for each $x \in \mathcal{X}$. Following Chellas [6], $\mathcal{S}^x_{\omega}$ is the relativised necessitation function $\mathcal{S}^x$ at world ω, that gives a subset of the power set of worlds (i.e. $\mathcal{S}^x : \mathcal{W} \rightarrow \wp(\wp(\mathcal{W}))$).

Given that a Kripkean possible world encapsulates a Hamblinian history of the form $\langle j \angle \upsilon, s, \delta^x \rangle$, we need two components to the interpretation function to return either the current state of Hamblinian history (namely, the set s), or the deeds which are about to be (or are being, instantaneously) carried out by agent x (namely, the set δ^x). Let us use the functions $\mathcal{I}_S$ to map from a possible world ω and a specified state of affairs A to an element of the set $\{\top, \bot\}$ according to whether or not A is in the set s of ω. Similarly, $\mathcal{I}_D$ maps from a possible world ω and a deed-assignment α^x to an element of the set $\{\top, \bot\}$ according to whether or not α^x is in the set δ^x of ω. The interpretation function is thus constituted from $\mathcal{I}_S$ and $\mathcal{I}_D$, to refer to the appropriate parts of the Hamblinian history.

QS	$\mathsf{S}_x\mathsf{S}_y A \rightarrow \mathsf{S}_x A$
QT	$\mathsf{S}_x\mathsf{T}_y\alpha \rightarrow \mathsf{T}_x\alpha$

Fig. 3. Axioms of delegation.

We are now in a position to be able to describe the semantics of S_x and T_x in a straightforward manner:

$$
\begin{array}{l}
\models^{\mathcal{M}}_{\omega} A \quad \text{iff} \quad \mathcal{I}_S(\omega, A) = \top \\
\models^{\mathcal{M}}_{\omega} \alpha^x \quad \text{iff} \quad \mathcal{I}_D(\omega, \alpha^x) = \top \\
\models^{\mathcal{M}}_{\omega} \alpha \quad \text{iff} \quad \exists x \text{ such that } \models^{\mathcal{M}}_{\omega} \alpha^x \\
\models^{\mathcal{M}}_{\omega} \mathsf{S}_x A \quad \text{iff} \quad \|A\|^{\mathcal{M}} \in \mathcal{S}^x_{\omega} \\
\models^{\mathcal{M}}_{\omega} \mathsf{T}_x \alpha \quad \text{iff} \quad \|\alpha\|^{\mathcal{M}} \in \mathcal{T}^x_{\omega}
\end{array}
$$

Bearing in mind that the truth set is simply $\|\phi\|^{\mathcal{M}} = \{\omega \in \mathcal{M}$ s.t. $\models^{\mathcal{M}}_{\omega} \phi\}$, this cleanly propagates the action/state distinction from the Hamblinian core to the desired modalities. This semantics thus offers a simple, if restrictive, interpretation of the two modalities, sufficient to explicate interesting interactions in a range of delegation scenarios.

5 Delegation

Here we propose further axioms and theorems of our logic of agentative action that are relevant to delegation, discuss the issue of forbearance in some detail and then focus on the application of the theory to delegation in multi-agent systems.

5.1 Further axioms and theorems

Like the approaches of Chellas and Belnap *et al.*, (but contrary to von Wright's characterisation), the theory offers scope for nesting the two modalities in building a rich notion of responsibility. In contrast to the clean, minimalist account developed by Jones and Sergot, the current work is employed in characterising realistic exchanges in agent systems, and as such the precise nature of the action modality needs to be pinned down. Thus following Chellas *inter alia*, we accept the axiom schemas QS and QT (figure 3).

Schema QT is worthy of particular note: if agent x sees to it that agent y sees to it that action α is done, then x can be said to be responsible for seeing to it that α is done. The adoption of this schema is intuitively appealing: agent x, through seeing to it that y is responsible for α, is itself, by delegating, responsible for its performance.

We further accept the specialisations of the TS schema, TSS and TST (figure 4). These schemata lay the foundation for characterising acts of delegation, but before looking at that in more detail, a second type of nested modality must be addressed that relates to the non-adoption of the axiom schema 5 for S and T (see section 3).

TSS	$\mathsf{S}_x\mathsf{S}_y A \rightarrow \mathsf{S}_y A$
TST	$\mathsf{S}_x\mathsf{T}_y \alpha \rightarrow \mathsf{T}_y \alpha$

Fig. 4. Further theorems of delegation.

5.2 Forbearance

Pörn [25] claims that, "The proposition *i forbears to bring it about that p* is not synonymous with *it is not the case that i brings it about that p*", basing his notion of forbearance upon an agent's ability to, but restraint from, bringing about the state of affairs. The same idea is presented by von Wright [32], but in Pörn's [25] account, the ability to nest operators supports rendering forbearance simply as: $\mathsf{S}_x\neg\mathsf{S}_x A$.

As Pörn discusses, forbearance and its associated causal responsibility is intuitively a stronger notion than simply not-bringing-it-about, and the former entails the latter. It is appropriate therefore that by T, $\mathsf{S}_x\neg\mathsf{S}_x A$ does indeed entail $\neg\mathsf{S}_x A$. This account of forbearance is the same as that of refraining discussed by Horty and Belnap [16], where it is also demonstrated to be equivalent to von Wright's original formulation.

Forbearing from action (as opposed to forbearing from responsibility for a state of affairs) is constructed in an analogous way, so that not being responsible for action is captured by $\neg\mathsf{T}_x\alpha$, but forbearing from action is the stronger notion expressed by $\mathsf{S}_x\neg\mathsf{T}_x\alpha$.

There are several points of note in this stronger notion of forbearance. The first is to recall that the modal statements themselves are — just as in standard ontic logics — part of the state of the world, and can thus form the parameter to the S_x modality (but not the T_x modality, which is not referring to the contents of the state of a world at all). The second is to emphasise that $\mathsf{S}_x\neg\mathsf{T}_x\alpha$ is not equivalent to the statement "x forbears from performing action α". The T_x modality expresses responsibility for the execution of an action, not the agent of the action, so this notion of forbearance should more accurately be read as "x forbears from having action α carried out". With the S_x modality, it is easy to separate the notion of responsibility from a given agent's action; with the T_x modality it is easy to forget that it is responsibility for, rather than direct participation in, action that is being expressed. The symmetry between S_x and T_x, and the focus upon responsibility rather than direct participation in both cases is crucial for the development of notions of delegation.

5.3 Imperatives in multi-agent systems

The use of this theory of agentative action as a model for imperatives in agent communication is predicated on the idea that imperatives can be constructed using a deontic action logic. Note that this is not the same as claiming that a deontic logic can be reduced to imperatives or vice versa (cf. Hamblin [13, 113–127]). It is however, claimed that normative positions where both normative (obligation, permission, etc.) and action components are involved can be seen as imperatives.

In this way, the statement $\mathsf{S}_x \bigcirc \mathsf{T}_y \alpha$, can be read as "$x$ sees to it that the state of affairs holds in which it is obligatory for y to see to it that α is performed". Further, the statement might be issued as an imperative by some third party to x. A linguistic example of such an imperative might be: "Make sure your sister cleans her teeth!" There may be a range of means by which x might bring about this state of affairs (as with any other) but one obvious alternative is for x to issue an imperative to y of the form $\mathsf{T}_y \alpha$ (e.g. "Clean your teeth, sis!").

Thus, in general, the act of uttering an imperative can, in the right situation, bring about a normative state of affairs. Clearly, both the form and type of locutionary act employed, and the imperative's overall success, will be partly dependent upon a variety of contextual factors, including in particular the relationship between the utterer and hearer, and existing normative positions both personal and societal. The general form of the interaction, though, is that the utterer attempts to introduce a new norm (and it is this act which counts as the utterer working towards whole-hearted satisfaction at this point); this attempt, if combined successfully with contextual parameters will generate a new normative position (or a modification of an existing position).

$$\mathrm{utter}(s, h, i) \wedge \langle \mathrm{context} \rangle \rightarrow \bigcirc i$$

Here, 'utter' is an appropriate communicative primitive, such as 'request'. s is the speaker, h the hearer and i an imperative formed using the S and T action modalities. The consequent is the normative position in which the addressee is obliged with respect to the content of the imperative i.

As mentioned above, the imperatives $\mathsf{S}_x A$ and $\mathsf{T}_x \alpha$ implicitly admit the possibility that x further delegates the activity. This implicit assumption is based on the simple deontic inter-definition between obligation and permission: $\mathsf{P}\, p \leftrightarrow \neg \bigcirc \neg p$. This, combined with some notion of negation as failure, licenses any agent to bring about normative states of affairs (in the right context), unless expressly prohibited from so doing. Suppose that an agent x is obliged to see to it that the state of affairs A is achieved as a result of y issuing the imperative $\mathsf{S}_x A$ to x. As long as it is not the case that x is forbidden from seeing to it that some other agent, say z, sees to it that A, x is permitted to do so by further delegating the activity. This represents something of a simplification of Lindahl's [20] theory of normative position (see also Sergot [29]). In fact, there are seven distinct normative positions of an individual with respect to a state of affairs: an agent may have the freedom (or not) to bring about p, the freedom (or not) to bring about $\neg p$ and the freedom (or not) to remain passive towards p. The work presented in this paper does not address the range of freedoms described by Lindahl, but is consistent with it. The focus is on the distinction between an agent being free to act and being free to delegate a task. See Reed *et al.* [27] for an analysis of the semantics of various communicative acts where the full range of individual agent normative positions (among other dimensions of the semantics of agent communication languages) is considered.

It can further be seen that, from axioms QS and QT (figure 3) and theorems TSS and TST (figure 4), that the further delegation of the activity will mean that the agent, x in this case, will be successful in fulfilling its responsibility for the completion of the activity. For example, from axiom QS and theorem TSS, if agent x sees to it that some

other agent y brings about the state of affairs A, then y brings about A (TSS) and x brings about A (QS).

It may be necessary to restrict the freedom of an agent to delegate, and to ensure that it carries out some action or brings about a state by his own, direct, intervention. Equally, there are, rarer, cases in which delegation is demanded. Taking this second and simpler case first, the imperatives $\mathsf{S}_x\mathsf{T}_y\alpha$ and $\mathsf{S}_x\mathsf{S}_yA$ capture this enforced delegation. $\mathsf{S}_x\mathsf{T}_y\alpha$ states that x brings it about that the state of affairs holds in which y is responsible for ensuring that the action α is performed. Similarly, $\mathsf{S}_x\mathsf{S}_yA$ states that x brings it about that the state of affairs holds in which y is responsible for ensuring that the state of affairs A is achieved.

The first case is slightly more complex. The implicit freedom of T_x (and identically for S_x) must be restricted by ensuring that x does not delegate. There are three important problems with an interpretation of this restriction:

1. Delegation is not a specified action. There are many ways of delegating, and most logics of action are not built around such template actions, in which place-holders such as 'delegate' can then subsequently be instantiated by some real action which constitutes delegation. It is certainly not a feature of the logic of S and T, and is not supported in the underlying semantics either, for good philosophical reasons [13]. It has been argued that delegation might be captured as a single, distinct communicative action [18]. The problem with this approach is that within any single given theory, definitions of other communicative acts already cover all the ground that constitutes delegation. In other words, delegation might be achieved through the application of any number of other primitives. To build on the approach by then predicating such action, and, in particular, abrogating the use of such delegation action, is doomed to failure, since on purely rational grounds, agents would simply employ these other means to their ends. It is thus indefensible to specify the prohibition of a delegation action.
2. As explained above, the distinction between states and events is a key component of action state semantics and to tie states to event postconditions would conflate this distinction, loosing much of the power of the semantics. Therefore, it is also undesirable to prohibit a state of affairs which can be uniquely identified with the postcondition of delegation.
3. An agent, say y, may be subject to a number of imperatives including, for example, the obligation to bring about that α is done, the status of which should not be impinged upon by restrictions on x's power to delegate responsibility for the performance of α. All that we wish to do is to restrict x's licence to delegate.

The solution lies in the notion of forbearance discussed in section 5.2. Intuitively, we wish to ensure that agent x forbears from seeing to it that some other agent becomes responsible for the activity. Suppose that the imperative concerned is $\mathsf{T}_x\alpha$. We wish to ensure that x forbears from bringing it about that another agent, say y, sees to it that α is done: $\mathsf{S}_x\neg\mathsf{S}_x\mathsf{T}_y\alpha$. Thus, the following imperative can be used to ensure that agent x carries out action α by its own, direct, intervention:

$$\mathsf{T}_x\alpha \wedge \mathsf{S}_x\neg\mathsf{S}_x\mathsf{T}_y\alpha$$

A simple linguistic gloss on this imperative runs, "x, do α and forbear from delegating responsibility for doing α!" — our adapted version of Pörn's [25] forbearance is thus being reconstructed in the imperative. Similarly, the following imperative may be used to ensure that agent x forbears from bringing it about that another agent, say y, sees to it that A is achieved:

$$\mathsf{S}_x A \wedge \mathsf{S}_x \neg \mathsf{S}_x \mathsf{S}_y A$$

These are simply special cases of forbearance from bringing about some state of affairs — $\mathsf{S}_x \neg \mathsf{S}_x A$, section 5.2 — where the state of affairs concerned is that some other agent becomes responsible for some activity.

So far in this discussion, the operator $\bigcirc$ representing the concept of deontic necessity is introduced with little discussion regarding its logic. Though the properties of deontic logic in general are not a focus of this paper, it is worth bearing in mind the following axiom schemas:

$$\begin{array}{ll} \mathrm{M}\bigcirc & \bigcirc(\phi \wedge \psi) \rightarrow \bigcirc\phi \wedge \bigcirc\psi \\ \mathrm{C}\bigcirc & \bigcirc\phi \wedge \bigcirc\psi \rightarrow \bigcirc(\phi \wedge \psi) \end{array}$$

If both are accepted, as they are in a standard deontic logic (the smallest normal system containing the axiom D$\bigcirc$ [6]), uttering the imperative $\mathsf{T}_x\alpha \wedge \mathsf{S}_x \neg \mathsf{S}_x \mathsf{T}_y \alpha$, if successful, will produce the normative state of affairs: $\bigcirc (\mathsf{T}_x\alpha \wedge \mathsf{S}_x \neg \mathsf{S}_x \mathsf{T}_y \alpha)$. The inward distributivity of M$\bigcirc$ then yields: $\bigcirc \mathsf{T}_x \alpha \wedge \bigcirc \mathsf{S}_x \neg \mathsf{S}_x \mathsf{T}_y \alpha$. The second conjunct is precisely what is required to restrict x's licence to further delegate the activity: x must refrain from establishing the state of affairs in which y is responsible for the performance of action α. This not only avoids problems (1) and (2) by referring to the imperative $\mathsf{T}_y\alpha$, but also circumvents (3) by leaving open the possibility that $\mathsf{P}\,\mathsf{T}_y\alpha$, or even $\bigcirc \mathsf{T}_y \alpha$, is (or will) in fact be the case — but not as a result of anything x has done (this, after all, is the definition of extensional satisfaction).

5.4 Examples

A couple of examples will serve to demonstrate not only the syntax of imperatives, the normative positions they engender, and the means by which whole-hearted satisfaction can be determined, but also to show clearly that the formalisation is intuitive and uncluttered.

Example 1 A lecturer is told by her head of department to prepare copies of her lecture notes for her class. She may, for example, copy the notes herself or request that the departmental secretary copy the notes.

The initial instruction refers to the action of copying the lecture notes; the Head of Department's locution is captured in L1, figure 5. This, because of the nature of the relationship between the Head of Department and the Lecturer, results in the normative state of affairs: $\bigcirc \mathsf{T}_{\mathrm{Lecturer}} \mathrm{copy\_notes}$.

This imperative may be whole-heartedly satisfied if the lecturer copies the notes herself; i.e. a world in which the deed-agent assignment $\mathrm{copy\_notes}^{\mathrm{Lecturer}}$ is present.

L1 Head of Department to Lecturer
$\mathsf{T}_{\text{Lecturer}}\text{copy\_notes}$
L2 Lecturer to Secretary
$\mathsf{T}_{\text{Secretary}}\text{copy\_notes}$
L3 Course Director to Lecturer
$\mathsf{T}_{\text{Lecturer}}\text{write\_exam} \wedge \forall y \in \{\mathcal{X} \setminus \{\text{Lecturer}\}\}\ \mathsf{S}_{\text{Lecturer}} \neg \mathsf{S}_{\text{Lecturer}} \mathsf{T}_y \text{write\_exam}$
L4 Course Director to Lecturer
$\mathsf{S}_{\text{Lecturer}} \mathsf{T}_{\text{Senior Secretary}} \text{print\_exam}$
L5 Lecturer to Student
$\mathsf{S}_{\text{Student}} \text{has\_paper}$
L6 Student to Librarian
$\mathsf{S}_{\text{Librarian}} \text{has\_paper}$
L7 Student to Librarian
$\mathsf{T}_{\text{Librarian}} \text{complete\_ILL}$

Fig. 5. Locutions in the University examples.

This is, however, only one possibility for the Lecturer. The Lecturer could issue the imperative represented by locution L2, figure 5. This should, in the given context, lead to a normative state of affairs: $\bigcirc \mathsf{T}_{\text{secretary}} \text{copy\_notes}$; i.e. the state of affairs in which the secretary is obliged to see to it that the copy_notes action is carried out. The action of the secretary carrying out copy_notes would fulfil the definition of extensional satisfaction not only of L2, but also of L1 in figure 5 (of course, the worlds of extensional satisfaction of L2 are identical to those of L1 in this case). Notice also that the secretary could further delegate the task to the tea-boy, etc.

Example 2 A lecturer is told by the Course Director that she must, herself, write an exam paper.

The initial request again concerns action, so the positive part of the imperative is captured by the first conjunct of locution L3 in figure 5. There is, however, the non delegation component, captured by the second conjunct. This states that the Lecturer is obliged to forbear from bringing about the state of affairs in which any agent (in the set of agents $\mathcal{X}$) with the exclusion of itself brings it about that the exam is written.

Thus the Lecturer may not be responsible for bringing about that any other agent is permitted to write her exam for her. Of course, it is conceivable that if, for example, she were to fall ill, her head of department might grant exam-writing permission to someone else in her place. Or, at a stretch of the imagination, there might be a role in a higher echelon of exam administration in which someone has the authority to write any exam paper they choose. Thus the normative position $\mathsf{P}\,\mathsf{T}_y \text{write\_exam}$ may either exist or come into existence for some agent y — this is extensional satisfaction. It may not, however, come about as the result of whole hearted satisfaction on the part of the lecturer.

Example 3 The Lecturer is told by the Course Director to ensure that the senior secretary prints the exam.

This is an example in which further delegation is demanded — the Senior Secretary is the only person in the department who should print exam papers, so the Lecturer must delegate this action to the Senior Secretary. The locution L4 captures this imperative, and will, if successful, produce the following normative state of affairs:

$$\bigcirc \mathsf{S}_{\text{Lecturer}} \mathsf{T}_{\text{Senior Secretary}} \text{print\_exam}$$

The Lecturer will then, with a view to whole-heartedly satisfying this imperative, issue the imperative that is captured by the locution $\mathsf{T}_{\text{Senior Secretary}} \text{print\_exam}$ to the Senior Secretary.

Example 4 The Lecturer asks her PhD student to get hold of a paper for her. The student may be able to download the paper right away, or, if it is not available online, to delegate the task of getting hold of the paper via an Inter-Library Loan request to the Librarian.

The imperative issued to the lecturer concerns a state of affairs, having a copy of the paper, and can be captured by locution L5 in figure 5. If the paper is on-line, the deed-agent assignment $\text{download\_paper}^{\text{Student}}$ is sufficient to introduce has_paper into the state of the world, thereby extensionally (and whole-heartedly) satisfying the imperative.

The alternative is to delegate the task to the Librarian (if possible), perhaps by issuing the imperative captured by locution L6 in figure 5. The Librarian would then be responsible (through the new normative position $\bigcirc \mathsf{S}_{\text{Librarian}} \text{has\_paper}$) for getting hold of the paper by whatever means she might see fit — by filling in an inter-library loan form, by ringing the British Library or whatever. It is of no concern to the PhD Student how the Librarian finds the paper; the Student's task is (in this case) done on creating the obligation on the Librarian.

Alternatively, the Student may decide to specify not the state of affairs that is desired, but rather the means by which they might be achieved. There are two key reasons why she might do this: (i) to avoid informing the Librarian of her goal (not relevant in this example); or (ii) to provide the Librarian with more detailed instructions (as might be appropriate if the PhD Student has already established that the library doesn't have a subscription to the journal in which the paper appears). Delegating the action is formulated, as can be seen from the locution L7 in figure 5, in as natural a way as delegating states of affairs.

6 Discussion

It now remains to discuss the consequences of using the model described in this paper in the practical task of specifying the primitives of an agent communication language. Following the distinction between actions and states, which has proven so useful in this discussion of imperatives, it is proposed that the primitives of an agent communication language should reflect this distinction. The FIPA ACL [11] provides three primitives that can be clearly understood as imperatives: 'request', 'request-when' and 'request-whenever'. A further primitive was included in earlier versions of this specification, but does not appear in the latest version: 'request-whomever'. Each of these primitives

refer to actions to be performed. The rationale for this choice being that they may refer to other communication primitives. For example, the primitive 'query-if' is defined in terms of the imperative 'request' and the indicative 'inform' — 'query-if' is a request that the recipient either inform the sender that some proposition is true (according to the beliefs of the recipient of the request) or that it is false.

The communicative act 'request-whomever' was given an informal description in earlier versions of the FIPA specification; it does not appear within the 2000 FIPA specification [11], but it is worth discussing here because of its clear relation to the theory of delegation presented in this paper. The primitive 'request-whomever' was described as "The sender wants an action performed by some agent other than itself. The receiving agent should either perform the action or pass it on to some other agent." This may be interpreted as an attempt (following Cohen and Levesque's [8] terminology) to delegate an action where the freedom to further delegate the action is unrestricted. This means that the recipient can: (1) not understand the message; (2) refuse the request;[3] (3) accept the request and perform the action itself; (4) accept the request and 'request' some other agent to perform it; or (5) accept the request and 'request-whomever' some other agent to perform it. This is, essentially, the same as case (b) mentioned in section 1 — "I don't care who does (action) α, but it must be done" — and is, therefore, the imperative $\mathsf{T}_x\alpha$, where x is the recipient and α is the action that is the message content.

A formal specification of the communicative act 'request' is provided. This is, in fact, a primitive communicative act in the FIPA specification, and the other imperatives (mentioned above) are specialisations. In common with the majority of action languages, the formal specification of the primitive 'request', and all other communicative acts within the FIPA specification, provides a set of 'feasibility preconditions' (FP) and a set of 'rational effects' (RE). The definition of request is reproduced in figure 6.[4]

$$\langle i, \mathrm{Request}(j, a)\rangle$$
$$\mathrm{FP} : B_i\,\mathrm{Agent}(j, a) \wedge \neg B_i\, I_j\,\mathrm{Done}(a)$$
$$\mathrm{RE} : \mathrm{Done}(a)$$

Fig. 6. The FIPA request communicative act.

There are two issues in this definition that are important to this discussion. First, the model relies on 'pseudo-states': the state of some action a having been done. As discussed, the model presented in this paper avoids this problem: it provides a means through which the primitives of an agent communication language can refer to the delegation of both *actions* and *goals*.

Second, and more importantly, to capture the notion of responsibility for satisfying the request, the preconditions include the belief of the message sender that the recipient

[3] (1) and (2) are appropriate responses for all FIPA messages.

[4] There is a further feasibility condition defined in the FIPA specification [11], but this refers to the feasibility conditions of the action a. Although this is itself problematic, it is not relevant to this discussion, and is therefore omitted.

is the agent of the action a. This is stated in the FIPA specification as follows [11, p. 32]: "$\mathrm{Agent}(i, a)$ means that i denotes the only agent that ever performs (in the past, present or future) the actions which appear in action expression a". This means that the semantics of this communicative act imposes a significant restriction on the action language that may be used as content to a FIPA message — all actions must be exclusive to the agent that performs the act. This is not a problem if the content is another FIPA message because the specification would include reference to the sender of the message, but the action used to illustrate the use of request in the FIPA specification [11, p. 25] is `"open \"db.txt\" for input"`! Leaving aside this difficulty, the request communicative act is close to case (d) mentioned in section 1: "You must, through your own, direct, intervention do α". An example of this case has been discussed in section 5.4: locution L5 in figure 5, where the Lecturer is instructed by the Course Director to write an exam paper and forbear from delegating responsibility for writing the paper.

Although this discussion has been restricted to the FIPA agent communication language, similar limitations can be identified in other ACLs such as KQML; see Reed *et al.* [27] for a more detailed analysis. This does, however, illustrate the fact that delegation cannot be captured as a single, distinct communicative act. There is a real need to develop flexible agent communication languages to support the complex dialogues that are required by agents interacting at the knowledge level.

In the discussion on delegation, it is assumed that getting someone else to act on your behalf is a valid means to the satisfaction of a commitment. This avoids the need to restrict the action component, and hence tie ends to sets of means. The restriction that delegation is forbidden (it is forbidden because the agent is obliged not to delegate) must then be explicitly stated within an agreement. This has some parallel with the notion of the protective perimeter of rights [14, 20]. The protective perimeter contains those actions that can be used to fulfil an obligation. This requires that the action component is extended to indicate that set of acceptable methods of achieving the goal. However, in parallel with Jones and Sergot [17], it is essential that an account of delegation is not dependent upon the detailed choices for the logic of the underlying action component.

7 Conclusion

There are several key advantages that can be gained through adopting the model presented in this paper. First, it becomes possible, in a single formalism, to distinguish an agent doing something, being responsible for getting something done, and being responsible for bringing about a state of affairs. This model provides a clear semantic interpretation for each. Second, it becomes possible to consider an agent's actions with regard to its commitment to a future obligation, and to determine whether or not it is behaving reasonably with respect to that commitment. Suppose that x accepts the task of doing α; i.e. it receives the imperative $T_x\alpha$ under the right context. Under this agreement, x is at all times obliged to perform deeds which ensure that it can carry out α, or at least it is forbidden from performing deeds which will remove the extensional satisfaction of $T_x\alpha$ from the bounds of possibility.

Thirdly, the language used for describing states of affairs in which agents have responsibilities and commitments can be used by those agents in ascribing such responsibilities through imperative- (rather than indicative-) based communicative acts of delegation. Finally, the intuitive simplicity of the approach has been demonstrated to be easily applied to real world examples, and to capture cleanly our intuitive understanding of responsibility and delegation.

References

1. L. Åqvist. *A new approach to the logical theory of interrogatives*. Tubingen, TBL Verlag Gunter Barr, 1975.
2. J.L. Austin. *How to do things with words*. Oxford University Press, 1962.
3. M. Barbuceanu and M. S. Fox. Integrating communicative action, conversations and decision theory to coordinate agents. In *Proceedings of the Second International Conference on Autonomous Agents*, pages 47–58, 1997.
4. C. Castelfranchi. Modelling social action for AI agents. *Artificial Intelligence*, 103:157–182, 1998.
5. C. Castelfranchi and R. Falcone. Principles of trust for MAS: Cognitive anatomy, social importance, and quantification. In *Proceedings of the Third International Conference on Multi-Agent Systems*, pages 72–79, 1998.
6. B. F. Chellas. *Modal logic: An introduction*. Cambridge University Press, 1980.
7. B. F. Chellas. Time and modality in the logic of agency. *Studia Logica*, 51(3/4):485–517, 1992.
8. P. R. Cohen and H. J. Levesque. Communicative actions for artificial agents. In *Proceedings of the First International Conference on Multi-Agent Systems*, pages 65–72, 1995.
9. F. Dignum. Using transactions in integrity constraints: Looking forward or backwards, what is the difference? In *Proceedings of the Workshop on Applied Logics*, 1992.
10. T. Finin, D. McKay, R. Fritzson, and R. McEntire. KQML: An information and knowledge exchange protocol. In K. Funchi and T. Yokoi, editors, *Knowledge Building and Knowledge Sharing*. Ohmsha and IOS Press, 1994.
11. Foundation for Intelligent Physical Agents. *FIPA communicative act library specification: XC00037H*, 2000. http://www.fipa.org/.
12. H. P. Grice. Meaning. *Philosophical review*, 66:377–388, 1957.
13. C. L. Hamblin. *Imperatives*. Basil Blackwell, 1987.
14. H. L. A. Hart. Bentham on legal rights. In A. W. B. Simpson, editor, *Oxford Essays in Jurisprudence*, 2, pages 171–201. Oxford University Press, 1973.
15. J. Hintikka, I. Halonen, and A. Mutanen. Interrogative logic as a general theory of reasoning. unpublished manuscript, 1996.
16. J. F. Horty and N. Belnap. The deliberative stit: A study of action, omission, ability, and obligation. *Journal of Philosophical Logic*, 24:583–644, 1995.
17. A. I. J. Jones and M. J. Sergot. A formal characterisation of institutionalised power. *Journal of the IGPL*, 4(3):429–445, 1996.
18. L. Kagal, T. Finin, and Y. Peng. A delegation based model for distributed trust. In *Proceedings of the IJCAI 2001 Workshop on Autonomy, Delegation and Control: Interacting with Agents*, 2001.
19. S. Kumar, M. J. Huber, D. R. McGee, P. R. Cohen, and H. J. Levesque. Semantics of agent communication languages for group interaction. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence*, pages 42–47, 2000.

20. L. Lindahl. *Position and change: A study in law and logic*. D. Reidel Publishing Company, Dordrecht, 1977.
21. J. McCarthy and P. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In D. Michie and B. Meltzer, editors, *Machine Intelligence*, volume 4, pages 463–502. Edinburgh University Press, 1969.
22. A. Newell. The knowledge level. *Artificial Intelligence*, 18:87–127, 1982.
23. T. J. Norman and C. A. Reed. Group delegation and responsibility. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems*, 2002.
24. P. Panzarasa, N. R. Jennings, and T. J. Norman. Formalising collaborative decision making and practical reasoning in multi-agent systems. *Journal of Logic and Computation*, 12(1):55–117, 2002.
25. I. Pörn. *The logic of power*. Basil Blackwell, 1970.
26. C. A. Reed. Dialogue frames in agent communication. In *Proceedings of the Third International Conference on Multi-Agent Systems*, pages 246–253, 1998.
27. C. A. Reed, T. J. Norman, and N. R. Jennings. Negotiating the semantics of agent communication languages. *Computational Intelligence*, to appear.
28. J. R. Searle. *Speech acts: An essay in the philosophy of language*. Cambridge University Press, 1969.
29. M. J. Sergot. Normative positions. In P. McNamara and H. Prakken, editors, *Norms, Logics and Information Systems*. ISO Press, 1998.
30. M. Shanahan. *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia*. MIT Press, Cambridge, MA, 1997.
31. 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*, pages 269–276, 1998.
32. G. H. von Wright. *An essay in deontic logic and the general theory of action*, volume 21 of *Acta philosophica Fennica*. North-Holland, Amsterdam, 1968.
33. D. N. Walton and E. C. W. Krabbe. *Commitment in dialogue: Basic concepts of interpersonal reasoning*. SUNY, New York, 1995.
34. M. J. Wooldridge. *Reasoning about rational agents*. MIT Press, 2000.