

Delegation and responsibility

Timothy J. Norman* Chris Reed†

* Department of Computing Science, University of Aberdeen,
Aberdeen, AB24 3UE, Scotland, U.K.

T.J.Norman@abdn.ac.uk

† Department of Applied Computing, University of Dundee,
Dundee, DD1 4HN, Scotland, U.K.

Chris.Reed@computing.dundee.ac.uk

Abstract. An agent may decide to delegate tasks to others. The act of delegating a task by one autonomous agent to another can be carried out by the performance of one or more imperative communication acts. In this paper, the semantics of imperatives are specified using a language of actions and states. It is further shown how the model can be used to distinguish between whole-hearted and mere extensional satisfaction of an imperative, and how this may be used to specify 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).

A number of questions are immediately apparent from the short characterisation of delegation and responsibility given above:

1. What is the nature of the relationship on which the ability to delegate is predicated?
2. Through what communicative acts can an agent delegate tasks, and how are they specified?
3. Under what conditions can it be said that delegation was successful?

In this paper, primary consideration is given to questions 2 and 3 (a reader interested in question 1 is referred to the model of roles and relationships proposed by Panzarasa *et al.* [18] and to the work of Castelfranchi [4]). Given that an agent has some rationale for attempting to delegate a task, how can this delegation be carried out and how can

it be considered a success? In answering these questions, the action-state semantics of imperatives proposed by Hamblin [11] is summarised, and the link between imperatives and normative positions is discussed. With this grounding, it is shown how the notion of delegation can be captured and how a clear distinction can be made between whole-hearted and mere extensional satisfaction. First, however, it is important to place imperatives in the context of other types of communicative acts, and relate this to existing literature on agent communication languages.

2 Imperatives and agent communication

Knowledge level [17] communication between agents through speech act based [2, 21] agent communication languages (ACLs) is both an active area of research [24, 20] and of standardisation [9, 8]. 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 [10]. 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 [21] description of the types of illocutionary (or communicative) act, Cohen and Levesque [6] 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, *a*, is an attempt to change the hearer’s mental state in such a way that it becomes an intention of the hearer to do *a*. The hearer, being an autonomous agent, may refuse. This is, of course, different from misunderstanding the speaker.

With this in mind, Cohen and Levesque [6] distinguish the goal of the speaker in performing the act and their intention. 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, however, is that the hearer believes that this is the intention of the speaker. 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.* [6, 24] 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 [9], and it is the specification of imperatives within FIPA that is returned to in section 5.

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. 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 [11, 26] and normative positions [15, 22].

2.1 The Nature of 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 [21, 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.* [13] among others, and these form an interesting avenue for future exploration. The focus of the current work, however, is on imperative logic. Hamblin's [11] 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 commands, 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 extralogical) factors affecting the subsequent effect of issuing imperatives.

2.2 Action State Semantics

Hamblin [11] offers an elegant investigation into imperatives, examining their role and variety, and developing an expressive syntax and semantics. Hamblin states [11, p. 137]

that to handle imperatives there are several features, “usually regarded as specialised”, which are indispensable: (1) a time-scale; (2) a distinction between actions and states; (3) physical and mental causation; (4) agency and action-reduction; and (5) intensionality. It is clear that any semantics, which competently integrates these five aspects should hold some significant appeal for those concerned with formalising the process by which agreements between agents are negotiated, specified, serviced, and verified.

The aim here is to equip the reader with a working grasp of action state semantics, so before the presentation of a formal summary, a brief overview is necessary. The first unusual feature of Hamblin’s model is the explicit representation of both events and states — that is, a world comprises a series of states connected by events. The states can be seen as collections of propositions; the 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 other, defining either states in terms of the sequences of events (true of most action and temporal deontic logics), or else, less commonly, events in terms of a succession of states (classical AI planning makes this assumption). It is interesting to note that the situation calculus [16] admits explicit representation of both states and events, but the commonly adopted “axioms of arboreality” [23] restrict the flexibility such that states can be defined as event sequences.¹

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 [7]. 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. Such whole-hearted satisfaction is based upon the notion of a strategy: an assignment of a deed to each time point for the specified agent. 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 execution of a deed from that strategy at every time point after the imperative is issued.

The summary presented below is the core of Hamblin’s model. For a more complete set-theoretic précis, the reader is referred to the appendix of Walton and Krabbe [26]. A world w in W is defined such that an assignment is made to every time point from T , (1) a state from the set of states S , (2) a member of the set H of ‘big happenings’ (each

¹ This conflation arises from associating a given sequence of events 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, there can be several different histories up to a given state and the histories are not themselves a part of those states.

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 $t' \geq t$.²

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.

Definition 1. 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*.

Definition 2. 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.

Definition 3. An agent x may be said to whole-heartedly satisfy an imperative i issued at t if and only if for every $t' \geq t$:

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

3 The Action Component

With this grounding in Hamblin's action-state semantics, the syntax is now extended to explicitly refer to agents performing actions and achieving goals. With the sets of deeds, $\alpha, \beta, \dots \in D$, states³ $A, B, \dots \in S$ and agents $x, y, \dots \in X$, action modalities for bringing about states of affairs and performing actions can be defined (section 3.2). Before this is done, however, it is useful to summarise Jones and Sergot's [14] action modality E_x on which the work reported here is based.

² 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.

³ In principle, states describe the entire universe at a given moment, so the elements referred to here are in fact portions of states; sets of propositions from which states are composed. The syntactic convenience adopted here is not important to the discussion.

3.1 The action modality E_x

In their model of institutionalised power, Jones and Sergot [14] specify an action modality E_x (read “ x sees to it that...”). E_x as the smallest system containing propositional logic and closed under the rule RE , with additional axiom schema T (necessary for a logic of successful action), and the rule of inference $R\neg N$ which is intended to capture the notion that an agent is somehow responsible for its actions.

$$\begin{array}{ll} \text{RE} & \frac{A \leftrightarrow B}{E_x A \leftrightarrow E_x B} \\ \\ \text{T} & E_x A \rightarrow A \\ \\ \text{R}\neg\text{N} & \frac{A}{\neg E_x A} \end{array}$$

In the context of specifying imperatives, the single most important shortcoming of the action modality E_x , as proposed by Jones and Sergot, is that it relies upon a state-based semantics. Events are viewed through what Hamblin terms ‘pseudo states’, such as the state of something having happened or of something having been done [11, p. 143]. Jones and Sergot themselves do not offer a precise characterisation of the reading of $E_x A$ — they typically refer to it as “ x seeing to it that state of affairs A holds” or “ x ’s act of seeing to it that A .” They remark [14, p. 435]: “we employ the same action modality E_x both for expressing that agent x creates/establishes states of affairs, and for expressing that x performs designated acts [...] We have found reasons to be uneasy regarding this kind of dual employment, and leave development to further work.”

A clean resolution to the issue of how to deal syntactically with both states of affairs and actions, while retaining the clear E-type properties and the seamless integration with deontic constructs, is crucial for providing a language of imperatives.

3.2 The Action Modalities S_x and T_x

Two new modalities, S_x and T_x , are proposed for a coherent action logic which can distinguish actions and states. $S_x A$ indicates that agent x sees to it that state of affairs A holds. Similarly, $T_x \alpha$ indicates that agent x sees to it that action α is carried out.⁴ Notice that in neither case is a specific action demanded of x — $T_x \alpha$ does not specify that x should necessarily perform action α , though clearly this would be one way in which it might be true. Each operator follows the relativised modal logic of type E [5], closed under rule RE : rule 1 being RE for the modal operator S_x and rule 2 being that for T_x .

$$\frac{A \leftrightarrow B}{S_x A \leftrightarrow S_x B} \quad (1)$$

$$\frac{\alpha \leftrightarrow \beta}{T_x \alpha \leftrightarrow T_x \beta} \quad (2)$$

⁴ The modalities draw their names from von Wright’s distinction in his seminal work on deontic logic [25, p. 59] between Seinsollen and Tunsollen, *ought to be* and *ought to be done*. Some of the links between the current work and deontic logic are explored below.

Following Jones and Sergot's exposition of the modality E_x , both S_x and T_x use the additional axiom schema T : axiom schema 3 being T for the modal operator S_x and axiom schema 4 being that for T_x .

$$S_x A \rightarrow A \quad (3)$$

$$T_x \alpha \rightarrow \alpha \quad (4)$$

These modalities are defined in terms of the action-state semantics summarised in section 2.2. Using the fact that, for a particular world, w , the states of that world are given by $S(w)$ and the deed-agent assignments are given by $D(w)$, the action modalities S_x and T_x are understood in the following way:

The modality S_x is tied to $S(w)$.

- (a) If $A \in S(w) \leftrightarrow B \in S(w)$ then $S_x A \in S(w) \leftrightarrow S_x B \in S(w)$.
- (b) If $S_x A \in S(w)$ then $A \in S(w)$.

The modality T_x is tied to both $S(w)$ and $D(w)$.

- (a) If $\alpha \in D(w) \leftrightarrow \beta \in D(w)$ then $T_x \alpha \in S(w) \leftrightarrow T_x \beta \in S(w)$.
- (b) If $T_x \alpha \in S(w)$ then $\alpha \in D(w)$.

Note that no equivalent to the rule of inference $R\neg N$ is defined here ($R\neg N$ was included by Jones and Sergot [14] to capture the notion that an agent is somehow responsible for its actions). In contrast, responsibility for seeing to it that states of affairs hold (for S_x) and actions are done (for T_x) is captured by the definition of whole-hearted satisfaction. An agent's responsibility for bringing about a state of affairs or for seeing to it that an action is done is therefore defined by both the maintenance of a partial i -strategy, Q'_i , and the selection of deeds specified by this Q'_i (see definitions 2 and 3).

It is worth re-emphasising at this point that there might be a temptation to "simplify" the logic, and define one of the two action modalities in terms of the other; to do so, however, would lose the attractive distinction provided by the semantic model, and thereby spoil the possibility of reasoning about commitment and the satisfaction of imperatives.

4 Delegation

There are two simple developments leading from the foundation introduced above that facilitate the introduction of machinery to handle delegation in a straightforward manner. First, the recognition that deontic statements can themselves be seen as states of affairs (see, for example Pörn [19]). Such states of affairs, like any other, can be the subject of the S_x modality. Second, imperatives can be constructed using the resulting deontic action logic.⁵

⁵ Note that it is not being claimed that deontic logic can be reduced to imperatives or vice versa (cf. Hamblin, 1987, pp. 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 $S_x O T_y \alpha$, 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 $T_y \alpha$ (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 either personal or 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).

$$\text{utter}(S, H, I) \wedge \langle \text{context} \rangle \rightarrow O I \quad (5)$$

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 then that the addressee is obliged with respect to the content of the imperative I .

As mentioned above, the imperatives $S_x A$ and $T_x \alpha$ implicitly admit the possibility that x delegates responsibility for their achievement. This implicit assumption is based on the simple deontic inter-definition between obligation and permission.

$$P p \leftrightarrow \neg O \neg p \quad (6)$$

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. This represents something of a simplification of Lindahl’s [15] theory of normative position (see also Sergot [22]). 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.

In fact, 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 imperative $S_x O T_y \alpha$ captures this enforced delegation: that x brings it about that the state of affairs holds in which y is responsible for ensuring that the action α is performed.

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 it is inappropriate for a logic to be founded upon some arbitrary action definition for delegation. Thus, it is undesirable to specify 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. The agent to whom an action may be delegated may himself be subject to a number of imperatives, the status of which should not be impinged upon by restrictions on x 's power to delegate. Thus, if y has an obligation towards action i.e. $OT_y\alpha$ then x 's inability to delegate responsibility for α should not be expressed using $\forall y \neg OT_y\alpha$.

The solution relies upon Hamblin's notion of whole-hearted satisfaction, and, in particular, upon interpreting the locutionary act, of $T_x\alpha$, as a request for whole-hearted satisfaction of the imperative. Trivially, the locution $\neg T_x\alpha$ is a request for not satisfying the imperative $T_x\alpha$ whole-heartedly; i.e. not adopting partial i -strategies at each t , and avoiding deeds that ensure extensional satisfaction of $T_x\alpha$. Crucially, extensional satisfaction of α is not thereby precluded (x may not 'whole-heartedly bring it about', but it may happen any way). This negated imperative can then be used to restrict the license to delegate: $\neg S_x P T_y\alpha$. Thus x must not whole-heartedly satisfy the imperative that the state of affairs is reached in which some y is permitted to be responsible for the performance of α . This not only avoids problems (1) and (2) by referring to a range of states of affairs after delegation, but also circumvents (3) by leaving open the possibility that $P T_y\alpha$, or even $O T_y\alpha$, 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). To enjoin x to perform some action α which he does not have the power to delegate can therefore be expressed as $T_x\alpha \wedge \forall y \neg S_x P T_y\alpha$.

The characterisation is isomorphic for $S_x A$. The basic semantic interpretation permits delegation, and that power can be restricted where necessary, resulting in the conjunction of imperatives $S_x A \wedge \forall y \neg S_x P S_y A$.

4.1 Worked 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 then (a) copy the notes herself (b) request that the departmental secretary copy the notes.

The initial request concerns actions, so the appropriate modality is T_x , and the whole imperative is specified in equation 7.

$$i_1 = T_{lecturer} copy_notes \quad (7)$$

The worlds in which this imperative is extensionally satisfied, W_{i_1} are given by equation 8.

$$W_{i_1} = \{w \mid \exists x \cdot copy_notes^x \in D(w)\} \quad (8)$$

That is, all those worlds in which anyone (any x) is assigned the deed of `copy_notes` ($D(w)$ gives deed-agent assignments, see section 2.2).

Thus a world in which the deed-agent assignment `copy_noteslecturer` is present would represent one in which i_1 is extensionally satisfied.

Alternatively, following example (1)b, the lecturer could issue the imperative i_2 to the secretary:

$$i_2 = T_{secretary} copy_notes \quad (9)$$

This should, in the given context, lead to a normative state of affairs in which

$$O T_{secretary} copy_notes \quad (10)$$

i.e. 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 i_2 , but also of i_1 in (4) (of course, the worlds of extensional satisfaction of i_2 are identical to those of i_1 in this case). Notice also that the secretary could further delegate the task to the tea-boy, etc.

Whole-hearted satisfaction of i_1 is defined as usual using W_{i_1} — the lecturer must select a deed from a partial i_1 -strategy, i.e. a partial allocation of deeds at time points to ensure that at least some of the worlds W_{i_1} remain possible. Both direct action and delegation thus keep extensional satisfaction within the bounds of possibility, and could thus figure in whole-hearted satisfaction of i_1 .

A similar situation holds for the secretary, and the tea-boy (presumably there would eventually be only the option of performing the task, since any alternative would lead to the imperative lapsing).

Example 2 A lecturer is told by her mentor that she must, herself, write an exam paper.

The initial request again concerns action, so the positive part of the imperative is captured by

$$i_3 = T_{lecturer} write_exam \quad (11)$$

There is, however, the non delegation component, captured by the second conjunct:

$$i_3 = T_{lecturer} write_exam \wedge \forall y \neg S_{lecturer} P T_y write_exam \quad (12)$$

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 $PT_y write_exam$ 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 A colleague asks the lecturer to get hold of a paper for him. She 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 a secretary.

The imperative issued to the lecturer concerns a state of affairs, having a copy of the paper, and can be captured thus:

$$i_4 = S_{lecturer}has_paper \quad (13)$$

If the paper is on-line, the deed-agent assignment $download_paper^{lecturer}$ is sufficient to introduce has_paper into the state of the world, thereby extensionally (and whole-heartedly) satisfying the imperative i_4 .

The alternative is to delegate the task to the secretary, perhaps by issuing the imperative i_5

$$i_5 = S_{secretary}has_paper \quad (14)$$

The secretary would then be responsible (through the new normative position $OS_{secretary}has_paper$) 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 lecturer how her secretary finds the paper; the lecturer's task is (in this case) done on creating the obligation on her secretary.

Alternatively, the lecturer 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 secretary of her goal; (ii) to provide the secretary with more detailed instructions (as might be appropriate if the secretary had been recently appointed, say). Delegating the action is formulated, as can be seen from (5), in as natural a way as delegating states of affairs — in this case, the lecturer would utter i_6 :

$$i_6 = T_{secretary}complete_ILL \quad (15)$$

5 Specifying communicative acts

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 [9] provides four primitives that can be clearly understood as imperatives: 'request', 'request-when', 'request-whenever'

and ‘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.

The informal description of ‘request-whomever’ (no formal definition is provided within the 1997 FIPA specification [9]) and the difference between this primitive and ‘request’ is of particular interest here. The primitive ‘request-whomever’ is 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.” Ignoring the ambiguity of these two sentences, an interpretation of the ‘request-whomever’ primitive could be that the message is an attempted delegation of an action where the freedom to further delegate the action is unrestricted. Presumably, this means that the recipient can: (1) not understand the message; (2) refuse the request; (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. The fact that agents can continue to “pass the buck” by forwarding this ‘request-whomever’ means that determining whether the request (or the imperative) has been satisfied is difficult. A further problem arises: What if agent x requests that y or whomever does a , then y requests that z or whomever does a and then z requests x to do a . Neither y nor z has refused the request, and both have passed it on to an agent other than the agent that requested that they do a , but the buck has passed back to x ! The action-state semantics of the model of imperatives presented in this paper provides a means to tie down this *delegation*.

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 here:⁶

$$\begin{aligned} &\langle i, \text{Request}(j, a) \rangle \\ \text{FP} : & B_i \text{Agent}(j, a) \wedge \neg B_i I_j \text{Done}(a) \\ \text{RE} : & \text{Done}(a) \end{aligned}$$

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 is the agent of the action to which the request refers! Through the notion of whole-hearted satisfaction, the model presented in this paper provides a more justifiable and robust characterisation of *responsibility*.

⁶ There is a further feasibility condition defined in the FIPA specification [9], 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.

6 Conclusions and future work

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 delegated task of doing α ; i.e. it involves the imperative $T_x\alpha$. 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.

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 [12, 15]. 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 [14], it is essential that an account of delegation is not dependent upon the detailed choices for the logic of the underlying action component.

The work presented in this paper has shown the application of the elegant action-state semantics proposed by Hamblin [11] to agent communication. Building on this, the paper contributes by: (1) giving a clear account of the distinction between doing and achieving through the introduction of the modalities S_x and T_x ; and (2) showing how delegation and responsibility can be cleanly captured using this novel framework.

References

1. L. Åqvist. *A new approach to the logical theory of interrogatives*. Tübingen, 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. B. F. Chellas. *Modal logic: An introduction*. Cambridge University Press, 1980.
6. 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.
7. F. Dignum. Using transactions in integrity constraints: Looking forward or backwards, what is the difference? In *Proceedings of the Workshop on Applied Logics*, 1992.
8. 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.

9. Foundation for Intelligent Physical Agents. *FIPA specification part 2: Agent communication language*, 1997. <http://www.fipa.org/>.
10. H. P. Grice. Meaning. *Philosophical review*, 66:377–388, 1957.
11. C. L. Hamblin. *Imperatives*. Basil Blackwell, Oxford, 1987.
12. 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.
13. J. Hintikka, I. Halonen, and A. Mutanen. Interrogative logic as a general theory of reasoning. unpublished manuscript, 1996.
14. A. I. J. Jones and M. J. Sergot. A formal characterisation of institutionalised power. *Journal of the IGPL*, 4(3):429–445, 1996.
15. L. Lindahl. *Position and change: A study in law and logic*. D. Reidel Publishing Company, Dordrecht, 1977.
16. 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.
17. A. Newell. The knowledge level. *Artificial Intelligence*, 18:87–127, 1982.
18. P. Panzarasa, T. J. Norman, and N. R. Jennings. Modeling sociality in the BDI framework. In J. Lui and N. Zhong, editors, *Proceedings of the First Asia-Pacific Conference on Intelligent Agent Technology*. World Scientific Publishing, 1999.
19. I. Pörn. *The logic of power*. Basil Blackwell, 1970.
20. C. A. Reed. Dialogue frames on agent communication. In *Proceedings of the Third International Conference on Multi-Agent Systems*, pages 246–253, 1998.
21. J. R. Searle. *Speech acts: An essay in the philosophy of language*. Cambridge University Press, 1969.
22. M. J. Sergot. Normative positions. In P. McNamara and H. Prakken, editors, *Norms, Logics and Information Systems*. ISO Press, 1998.
23. M. Shanahan. *Solving the frame problem*. MIT Press, 1997.
24. 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.
25. 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.
26. D. N. Walton and E. C. W. Krabbe. *Commitment in dialogue: Basic concepts of interpersonal reasoning*. SUNY, New York, 1995.