Group Delegation and Responsibility

Timothy J. Norman
Department of Computing Science,
University of Aberdeen,
Aberdeen, AB24 3UE, Scotland, U.K.
tnorman@csd.abdn.ac.uk

Chris Reed
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 by one agent to another can be carried out by the performance of one or more communicative acts. Such acts may not only be directed to another individual, but also to a group of agents. In this paper, the semantics of imperatives are explored with reference to extant logics of agentative action, and in the context of the referent of an imperative being either an action or a state of affairs. The particular case of issuing of an imperative to a group of individuals is then discussed from both theoretical and practical perspectives, and the importance of distinguishing between groups with differing characteristics emphasised by analysis of an extended real-world example.

Categories and Subject Descriptors

I.2 [**Computing Methodologies**]: Artificial Intelligence; I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent systems, Coherence and coordination*; D.2.8 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods—*Modal Logic*; F.4 [**Theory of Computation**]: Mathematical Logic and Formal Languages; F.4.1 [**Mathematical Logic and Formal Languages**]: Mathematical Logic—*Modal Logic*

General Terms

Theory

1. INTRODUCTION

A natural extension to the theory of delegation and responsibility proposed in companion work [10] is the issuing of an imperative to a group of agents. Examples of such imperatives include "all of you stand up" and "one of you shut the door" both in the context of a teacher addressing a class. These examples illustrate the distinction between the group being addressed distributively (as a list of individuals) and as a collective (as a "meta-agent"). Rescher [13] uses the terms "distributive" and "collective" groups, and Kumar *et al.* [9] distinguish between groups as a "list of individuals" and as a "meta-agent". Here, we adopt the terms used by Rescher *ibid.*, but the concepts are equivalent. In the first case, the group is being addressed distributively — each student should stand up — and in the second case the group is being addressed as a collective — at least one student (but possibly all) should close the door.

An investigation into what it means for an imperative to be issued to a collective or distributive group must, however, be built upon a sound logic of agentative action, and, we argue, one that provides the flexibility of being able to refer to actions as well as states of affairs. For this reason, the preliminary exposition of the semantics of the modal operators $\mathsf{S}_x A$ — agent x sees to it that the state of affairs A holds — and $\mathsf{T}_x \alpha$ — agent x sees to it that the action α is done — presented in [10] is extended and their logics contrasted with extant logics of agentative action reported in the literature.

This paper extends existing work by providing a basic axiomatisation and intuitive understanding of a logic of agentative action (section 3), and a discussion of the merits of a possible-worlds semantic interpretation of the modalities S_x and T_x (section 4). The issuing of an imperative (with respect to an action or a state of affairs) to a (distributive or collective) group is then investigated (section 6). Finally, the theory is placed in a practical context through the analysis of a real-world scenario in which individual and group delegation is an integral part (section 7). The problem is introduced by briefly reviewing existing logics of agentative action.

2. LOGICS OF ACTION

In the literature dealing with logics of agentative action there are a number of theories that have illuminated and motivated the development of the current work of devising a characterisation suitable for application in multi-agent systems. Though many authors have contributed to the area, three theories stand out, representing recent distinct solutions to a common problem, each formulated in a similar style. These three are the work of Jones and Sergot [8], of Belnap, Horty and Perloff [1, 7], and of Chellas [3].

Chellas [3], building on earlier work, presents one of the first theories integrating communication (and, in particular, imperatives) and agentative action, where the $\triangle$ modality is formulated as a normal modal logic, based on a semantics of histories which map from time points to states of affairs.

Jones and Sergot [8] characterise the responsibility of an agent towards action using a small classical modal logic of

AAMAS'02, July 15-19, 2002, Bologna, Italy.

type ECT, in which semantics are of a conventional possible-worlds type.

Finally, Belnap and Perloff (1995) offer a characterisation of 'stit', or seeing to it that, based upon a semantics of 'stit-frames' founded on branching time. A full axiomatisation of this work on stit is provided by Xu [18]. One feature that Chellas [3], Jones and Sergot [8] and Belnap and Perloff [1] all have in common is that they do not distinguish between agentative actions and agent responsibility for states of affairs. Clearly though, each had concerns — Chellas voices his in a footnote [3, p. 489], and Jones and Sergot express theirs in an intriguing aside: "we have found reasons to be uneasy regarding this kind of dual employment [for both events and states of affairs]" [8, p. 435]. Belnap's concerns led to the development of a more refined model which distinguishes between the achievement stit and deliberative stit [7]. The former is a conventional notion of seeing-to-it-that, and the latter is more closely related to action; it refers to a single state, rather than the change occurring between two states, and is "evaluated at the very moment at which agent a sees to it that A" [7, p. 592].

The current work differs from these three in several respects. Firstly, the semantic model is richer; in this paper, a simplified version is presented in terms of Kripkean frames, but in companion work, a complete model is developed using Hamblin's action state semantics, as described in [10]. Secondly, we argue that some of the logical properties of the systems presented in the three works are inappropriate for the task at hand, and that a regular modal logic is more appropriate. Finally, partly as a result of adopting Hamblin semantics, and partly as a result of adopting a regular logic, the system presented below clearly distinguishes between events and states of affairs in a solution which is more suited to multi-agent system development than the approach adopted by Horty and Belnap.

In this paper, we extend the discussion of the two new modalities, S_x and T_x, proposed in [10] as a coherent action logic that can distinguish between actions and states. (These modalities draw their names from von Wright's [15] distinction between Seinsollen, *ought to be*, and Tunsollen, *ought to be done*.) The modalities are introduced in two steps. First, through an informal presentation of an axiomatisation which is given an intuitive grounding and assumes a Kripkean underpinning. The formal semantics is then laid out more carefully using possible worlds, and shortcomings with this approach are identified to lay the groundwork for a more precise model to be developed using Action State Semantics.

3. AXIOMATISATION

Throughout the paper, we refer to actions (or deeds) by the symbols $\alpha, \beta, \ldots \in D$, states by $A, B, \ldots \in S$ and agents by $x, y, \ldots \in \mathcal{X}$. 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 [2]. 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.[1] 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.

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	$\mathsf{S}_x A \rightarrow \neg \mathsf{S}_x \neg A$

Figure 1: Rules of inference & axiom schemas of S.

R¬NS	$\dfrac{A}{\neg \mathsf{S}_x A}$
RNS	$\dfrac{A}{\mathsf{S}_x A}$
RMS	$\dfrac{A \rightarrow B}{\mathsf{S}_x A \rightarrow \mathsf{S}_x B}$
5S	$\neg \mathsf{S}_x A \rightarrow \mathsf{S}_x \neg \mathsf{S}_x A$
4S	$\mathsf{S}_x A \rightarrow \mathsf{S}_x \mathsf{S}_x A$

Figure 2: Further rules of inference and axiom schemas discussed.

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¬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" [8, p. 435]. Thus it could be argued that Jones and Sergot, like Belnap and Perloff (whose *negative condition* entails R¬N) do capture an element of the notion of responsibility, in the sense that no agent can be said to be 'responsible' for a tautology. Chellas' intuitions, by contrast run rather differently. He 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.

[1] 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.

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." [3, p. 508]. Chellas' decision, in particular, is motivated by the logical consequences of the rule, and 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 5.2. 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 5.2.

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 [2, p. 237].

4. SEMANTIC MODEL

Companion work [10] sketches a semantic model based upon Hamblin's Action State Semantics for imperatives [5]; here a simplified account is developed which is founded upon more traditional possible-worlds semantics.

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 [2].

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.)

$$\models^{\mathcal{M}}_{\omega} P_n \quad \text{iff} \quad \omega \in \mathcal{P}_n \text{ for } n = 1, 2, 3, \ldots \tag{1}$$
$$\models^{\mathcal{M}}_{\omega} \top \tag{2}$$
$$\not\models^{\mathcal{M}}_{\omega} \bot \tag{3}$$
$$\models^{\mathcal{M}}_{\omega} \neg A \quad \text{iff} \quad \not\models^{\mathcal{M}}_{\omega} A \tag{4}$$
$$\models^{\mathcal{M}}_{\omega} A \wedge B \quad \text{iff} \quad \models^{\mathcal{M}}_{\omega} A \text{ and } \models^{\mathcal{M}}_{\omega} B \tag{5}$$
$$\models^{\mathcal{M}}_{\omega} A \vee B \quad \text{iff} \quad \models^{\mathcal{M}}_{\omega} A \text{ or } \models^{\mathcal{M}}_{\omega} B \text{ or both} \tag{6}$$
$$\models^{\mathcal{M}}_{\omega} A \to B \quad \text{iff} \quad \text{if } \models^{\mathcal{M}}_{\omega} A \text{ then } \models^{\mathcal{M}}_{\omega} B \tag{7}$$
$$\models^{\mathcal{M}}_{\omega} A \leftrightarrow B \quad \text{iff} \quad \models^{\mathcal{M}}_{\omega} A \text{ if and only if } \models^{\mathcal{M}}_{\omega} B \tag{8}$$
$$\models^{\mathcal{M}}_{\omega} \mathsf{S}_x A \quad \text{iff} \quad \|A\|^{\mathcal{M}} \in \mathcal{N}^x_{\omega} \tag{9}$$

Unfortunately, quite apart from practical difficulties in using such a model as the basis for implementation of a multi-agent system [17] 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 [3], Horty and Belnap [7] 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 [17]). 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 [16, p. 191]). This is defined recursively as follows:

$$j\angle_0\upsilon = \langle\emptyset, s_0 \in S(\upsilon), \delta_0^x \in D(\upsilon)\rangle$$
$$j\angle_t\upsilon = \langle j\angle_{t-1}\upsilon, s_t \in S(\upsilon), \delta_t^x \in D(\upsilon)\rangle$$

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), $D(\upsilon)$, (for more details, see Hamblin [5]).

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 [2], $\mathcal{S}^x_\omega$ 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.

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

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

The truth set is as follows:

$$\|\phi\|^\mathcal{M} = \{\omega \in \mathcal{M} \text{ s.t. } \models_\omega^\mathcal{M} \phi\}$$

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 and discuss the issue of forbearance.

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

Figure 3: Axioms of delegation.

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

Figure 4: Some theorems of delegation.

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 sees to it that α is done. The adoption of this schema is intuitively appealing: agent x, through it seeing to it that y is responsible for α is itself, through delegation of the act, 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).

These schemata lay the foundation for characterising acts of group delegation, but before looking at that problem 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.

5.2 Forbearance

Pörn [11] 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 [15], but in Pörn's [11] 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 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 [7], 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 the notions of group delegation and group interaction presented below.

6. DELEGATION TO A GROUP

In this section, the theory of delegation and responsibility is extended to the case of an imperative being issued to a group.

A group of agents is defined as a set of individuals, and upper case letters, $X, Y, \ldots \subseteq \mathcal{X}$, are used to denote groups. Here, we define a group simply as a set of agents rather than considering additional structure through defining a group as a set of groups. Providing additional structure by defining hierarchies of agent groups does not benefit a theory of delegation and responsibility, simply because an imperative, issued to a (collective or distributive) group of agents, is issued to every member of that group regardless of any additional organisational structure. This does not mean that organisational roles and relationships do not give weight to the delegation of an activity; this is an essential part of the context in which activities are delegated. Consider, for example, the CEO of a company issuing a directive regarding the company policy on (self-)certification of illness. This is directed to all employees of the company regardless of their position. It may be that this imperative is issued through the distribution of a memo — the mode of delivery is not important — but it applies to all those to whom it is directed. Suppose, in contrast, that the CEO issues the imperative to her heads of department that they each reduce the costs of their department by 10% over the following financial year. This is directed only to the heads of department. The instruction may provide weight to (give some justification for) any subsequent imperative issued by a head of department regarding the reduction of costs within their department. This subsequent imperative is, however motivated by the first, not the simple transmission of the CEO's imperative.

Delegating to a group can thus take several forms, depending upon whether it is a state of affairs or an action which is being delegated, and whether the group to whom the delegation is addressed is regarded as distributive or collective.

6.1 Distributive S

The imperative $\mathsf{S}_{[X]}A$ is read "all agents in the set X see to it that the state of affairs A holds". Such an imperative is *distributed* to each member of the group. Suppose that X is $\{x, y\}$. For the distributive group $[X]$ to see to it that A holds, both x and y must see to it that A. As a first attempt at capturing this type of group-directed imperative, let us consider the following definition:

$$\mathsf{S}_{[X]}A =_{\text{def}} \bigwedge_{x \in X} \mathsf{S}_x A$$

Suppose that two employees, Alice and John, are requested to make sure that the department's financial report is completed for a meeting tomorrow. $X = \{\text{Alice}, \text{John}\}$ is, therefore, a group to which the following imperative is issued (where 'r' indicates the report): $\mathsf{S}_{[\{\text{Alice},\text{John}\}]}\text{r}$. In accordance with the above definition, this is equivalent to $\mathsf{S}_{\text{Alice}}\text{r} \wedge \mathsf{S}_{\text{John}}\text{r}$, and, if uttered in an appropriate context, this imperative will lead to the establishment of the normative state of affairs: $\bigcirc(\mathsf{S}_{\text{Alice}}\text{r} \wedge \mathsf{S}_{\text{John}}\text{r})$. This state of affairs could be satisfied, for example, through Alice collating the statistics for the current year and John developing a proposal for the financial management of the year ahead. Exactly how the delegated state of affairs is achieved is left to the members of the group addressed — in this case, Alice and John.

It is worth considering at this point the following axiom schemas of the logic of $\bigcirc$.

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

If both are accepted, as they are in a standard deontic logic (the smallest normal system containing the axiom D$\bigcirc$ [2]), uttering an imperative to a distributive group with n members, if successful, will produce the same normative state of affairs as n successful imperatives uttered, one to each member of the group. In the case of our example, if M$\bigcirc$ is accepted, it would yield the following: $\bigcirc\mathsf{S}_{\text{Alice}}\text{r} \wedge \bigcirc\mathsf{S}_{\text{John}}\text{r}$.

Intuitively there seems to be more to the issuing of an imperative to a distributive group than there would be in issuing a number of individual imperatives. Suppose that John fails to contribute to the report, leaving all the effort to Alice. If Alice does, in fact, complete the report on schedule can it be said that the original group-directed imperative has been whole-heartedly satisfied? We believe that this is not the case.

Could it be said, however, that Alice has "done her part" in the satisfaction of the imperative? Simply due to the fact that she has whole-heartedly satisfied the imperative, it can be seen that she has, in fact, done her part. It can, similarly, be seen that John has not. Therefore, this formulation of the group-directed imperative does not enforce group responsibility — Alice is not culpable for the fact that John did not contribute to the completion of the report.

Here, it is argued that the fact that this imperative is issued to the group, rather than individual imperatives issued to each member of the group, indicates that there should be a notion of *group responsibility* for the activity concerned. For this reason, the following stronger definition of an imperative directed to a distributive group is adopted:

$$\mathsf{S}_{[X]}A =_{\text{def}} \bigwedge_{x \in X} \mathsf{S}_x \left(\bigwedge_{y \in X} \mathsf{S}_y A \right)$$

Using the same example and this alternative formulation, the imperative issued to Alice and John expands to (reducing $\mathsf{S}_x\mathsf{S}_x A$ to $\mathsf{S}_x A$ by T throughout):

$$\mathsf{S}_{\mathrm{Alice}}\mathrm{r} \land \mathsf{S}_{\mathrm{Alice}}\mathsf{S}_{\mathrm{John}}\mathrm{r} \land \mathsf{S}_{\mathrm{John}}\mathrm{r} \land \mathsf{S}_{\mathrm{John}}\mathsf{S}_{\mathrm{Alice}}\mathrm{r}$$

This introduces a flavour of group responsibility for the satisfaction of the state of affairs concerned — as well as Alice and John being individually responsible for the completion of the report, Alice is responsible for seeing to it that John (at least contributes to) the completion of the report and vice versa. This is exactly what is required in a model of group-directed imperatives.

6.2 Collective S

The imperative $\mathsf{S}_{\langle X\rangle}A$ is read "at least one agent in the set X sees to it that the state of affairs A holds". This imperative is directed to a collective group and, for it to be satisfied, at least one of the group members (possibly all of them) must see to it that A holds. A first attempt at a definition of this group-directed imperative, therefore, may be to equate it with the disjunction of individual imperatives for each member of the group.

$$\mathsf{S}_{\langle X\rangle}A =_{\mathrm{def}} \bigvee_{x\in X} \mathsf{S}_x A$$

Consider an example that is similar to that discussed in section 6.1: Alice and John are instructed that at least one of them must prepare a presentation on next year's financial prospects for the meeting tomorrow. (In this example, although not in general, it would be sensible for only one of them to see to it that the presentation is prepared.) Following the initial definition above, this imperative, directed to an collective group, would expand to (where 'p' indicates the presentation):

$$\mathsf{S}_{\mathrm{Alice}}\mathrm{p} \lor \mathsf{S}_{\mathrm{John}}\mathrm{p}$$

Rescher [13, p. 59] discusses the issuing of an imperative to such a collective group. Using the example of a group of students being instructed to close the door, Rescher considers the following alternative formulation: "do not let it occur that no one in the group [. . .] closes the door" addressed distributively to the group (or "*forbear* from letting it occur that no one in the group closes the door", ref. section 5.2). Consider this alternative formulation of the example considered here: "do not let it occur that neither Alice nor John sees to it that the presentation is prepared". This can be expressed as follows:

$$\begin{aligned}&\mathsf{S}_{\mathrm{Alice}}\neg\left(\neg\mathsf{S}_{\mathrm{Alice}}\mathrm{p} \land \neg\mathsf{S}_{\mathrm{John}}\mathrm{p}\right) \quad \land\\ &\mathsf{S}_{\mathrm{John}}\neg\left(\neg\mathsf{S}_{\mathrm{Alice}}\mathrm{p} \land \neg\mathsf{S}_{\mathrm{John}}\mathrm{p}\right)\end{aligned}$$

This alternative reading does, in common with the definition of $\mathsf{S}_{[X]}A$ discussed in section 6.1, include a flavour of group responsibility for the state of affairs referred to in the group-directed imperative. The general case, therefore, of an imperative issued to a collective group $\langle X\rangle$ and a state of affairs A that includes this *group responsibility* is:

$$\mathsf{S}_{\langle X\rangle}A =_{\mathrm{def}} \bigwedge_{x\in X} \mathsf{S}_x\left(\bigvee_{y\in X}\mathsf{S}_y A\right)$$

Returning to the example, when uttered in an appropriate context, the imperative will lead to the establishment of the following normative state of affairs:

$$\begin{aligned}&\bigcirc\mathsf{S}_{\mathrm{Alice}}\left(\mathsf{S}_{\mathrm{Alice}}\mathrm{p} \lor \mathsf{S}_{\mathrm{John}}\mathrm{p}\right) \quad \land\\ &\bigcirc\mathsf{S}_{\mathrm{John}}\left(\mathsf{S}_{\mathrm{Alice}}\mathrm{p} \lor \mathsf{S}_{\mathrm{John}}\mathrm{p}\right)\end{aligned}$$

It is obligatory that Alice sees to it that either Alice or John (or both) sees to it that the presentation is prepared and it is obligatory that John sees to the same. For either of these members of the group to whole-heartedly satisfy this group-directed imperative, they must take into account the activity of the other. This, therefore, enforces cooperation in the satisfaction of the imperative issued.

6.3 Distributive & collective T

Definitions of the distributive and collective group-directed imperatives with respect to action, $\mathsf{T}_{[X]}\alpha$ and $\mathsf{T}_{\langle X\rangle}\alpha$ respectively, follow those of $\mathsf{S}_{[X]}A$ and $\mathsf{S}_{\langle X\rangle}A$. Again, to capture the notion of group responsibility, $\bigwedge_{x\in X}\mathsf{S}_x$ precedes the distributive and collective cases for the performance of actions.

$$\mathsf{T}_{[X]}\alpha =_{\mathrm{def}} \bigwedge_{x\in X} \mathsf{S}_x\left(\bigwedge_{y\in X}\mathsf{T}_y\alpha\right)$$

$$\mathsf{T}_{\langle X\rangle}\alpha =_{\mathrm{def}} \bigwedge_{x\in X} \mathsf{S}_x\left(\bigvee_{y\in X}\mathsf{T}_y\alpha\right)$$

Finally, in this discussion of group-directed imperatives, it is worth noting that the issuing of an imperative to a (distributive or collective) group of agents is a generalisation of the issuing of an imperative to an individual. This can be seen by considering the special case of issuing an imperative to a group containing a single agent: the case where $X = \{x\}$. These follow simply from the definitions of each group-directed imperative definition and the specialisations of the T schema discussed in section 5.1. Thus:

$$\begin{aligned}&\mathsf{S}_{[\{x\}]}A \rightarrow \mathsf{S}_x A \qquad && \mathsf{S}_{\langle\{x\}\rangle}A \rightarrow \mathsf{S}_x A\\ &\mathsf{T}_{[\{x\}]}\alpha \rightarrow \mathsf{T}_x\alpha && \mathsf{T}_{\langle\{x\}\rangle}\alpha \rightarrow \mathsf{T}_x\alpha\end{aligned}$$

7. CONSTRUCTION SCENARIO

Construction engineering sites have been recognised as offering a rich domain filled with diverse and complex inter- and intra-organisational relationships and communication channels [12]. The examples presented here exploit this richness by examining a number of situations occurring as part of the scenario captured in the precis presented in figure 5 (adapted from Perry [12]).

The example begins with a site visit by the Senior Engineer to her foremen, during which she delegates the task of completing the formwork to her foremen. There may be several foremen, each in charge of a team; only one of these teams need carry out the work, but the Senior Engineer wants to leave it to the foremen to negotiate amongst themselves which team should be responsible for the work. Her locution is captured in L1, figure 6.

The Highways Agency has contracted "ConsCo" to build a new bridge over a railway line as a part of a large civil engineering project. At one stage, the plans call for a crane to lift wooden formwork into place to support the pouring of concrete. There are six distinct roles utilised in the examples:

i The *Senior Engineer*, with overall responsibility for the construction team (the example conflates several layers of management for the sake of clarity).

ii The *Foreman*, under direct charge of the Senior Engineer.

iii The *Ganger*, under direct charge of a Foreman, and in turn, in charge of a team of labourers.

iv The *Resident Engineer*, responsible for ensuring that the contract laid out in the design documentation is successfully executed on the ground by the construction team.

v The *Temporary Works Coordinator*, part of the production support team, and responsible for designing "temporary works", such as formwork.

vi The *Crane Plant Supplier*, responsible for supply and advice of plant material on site The examples examine a scenario in which the constraints recognised by the CP supplier lead to a request for a change in the design of the formwork.

Figure 5: The construction engineering scenario

L1 Senior Engineer to {foreman1, foreman2, foreman3}
$\mathsf{S}_{\langle\{\text{foreman1,foreman2,foreman3}\}\rangle}$formwork_complete

L2 Foreman to {ganger1, ganger2}
$\mathsf{S}_{[\text{ganger1,ganger2}]}$formwork_complete

L3 Senior Engineer to Crane Plant Supplier
$\mathsf{S}_{\text{CP_supplier}}$formwork_in_place

L4 Resident Engineer to Crane Plant Supplier
$\mathsf{S}_{\text{CP_supplier}}\neg\mathsf{S}_{\text{CP_supplier}}$unsafe

L5 Crane Plant Supplier to Temporary Works Coordinator
$\mathsf{S}_{\text{TW_coordinator}}$lighter_formwork_design

L6 Temporary Works Coordinator to Senior Engineer
$\mathsf{T}_{\text{Senior_engineer}}$lop_formwork

L7 Senior Engineer to {foreman1, foreman2, foreman3}
$\mathsf{T}_{\langle\{\text{foreman1,foreman2,foreman3}\}\rangle}$lop_formwork

L8 Foreman to {ganger1, ganger2}
$\mathsf{T}_{[\{\text{ganger1,ganger2}\}]}$lop_formwork

Figure 6: Locations in the CE scenario.

L1 is an imperative to a collective group, and (given that it is correctly addressed to each member of the group and the appropriate organisational context holds) leads as a result to a complex set of norms holding for each member of the group (where 'fc' indicates formwork_complete and 'f1' is foreman1, etc.):

$$\begin{aligned}&\bigcirc\mathsf{S}_{f1}\,(\mathsf{S}_{f1}\mathrm{fc}\vee\mathsf{S}_{f2}\mathrm{fc}\vee\mathsf{S}_{f3}\mathrm{fc})\quad\wedge\\&\bigcirc\mathsf{S}_{f2}\,(\mathsf{S}_{f1}\mathrm{fc}\vee\mathsf{S}_{f2}\mathrm{fc}\vee\mathsf{S}_{f3}\mathrm{fc})\quad\wedge\\&\bigcirc\mathsf{S}_{f3}\,(\mathsf{S}_{f1}\mathrm{fc}\vee\mathsf{S}_{f2}\mathrm{fc}\vee\mathsf{S}_{f3}\mathrm{fc})\end{aligned}$$

What these norms achieve is the stimulation of negotiation: if foreman2, say, is already fully committed, then meeting his disjunction demands being responsible for seeing to it that one of the other two brings about the desired state of affairs. As in [10], the simplest way for this to be achieved is by communicating with either of the other two, to persuade them appropriately. Of course, the same constraints would also be impinging upon the other two foremen, who may be trying to persuade one another. It is clear from this example that dealing with three-party persuasive, deliberative and negotiative dialogues is, contrary to an oft claimed assumption, not simply a matter of scaling up a two-party dialogue. A richer characterisation of the locutions and locution-witnessing that is demanded by the binding enforced by collective imperatives is the subject of current research, but the current example serves to demonstrate that the obligations lead naturally to the most obvious (though not sole) solution, namely a three-way meeting.

Once one of the foreman has agreed to carry out the formwork task, he must propagate the instructions down to his gangers, all of whom must then commit to working on the project. The foreman's locution is of the form given in L2. Once again, this leads (as a result of the foreman being in a position to issue orders to his gangers) obligations on each ganger, and upon the ganger team as a whole. Note that by his locution (or locutions), it can be said that the foreman sees to it that the gangers see to it that the formwork is complete and thus, by the axiom QS, the foreman sees to it that the formwork is complete. This, of course, is the essence of delegation.

Meanwhile, the Senior Engineer must also instruct the Crane Plant Supplier to lift the formwork into place once it is complete. Notice that the Senior Engineer neither knows nor cares *how* this is to be achieved — it is only the final state that she is interested in. Her locution, addressed to the CP Supplier, is given in L3.

The Resident Engineer is to ensure the safety of the same lifting procedure. There are several ways in which this might be cast, but for the sake of example let us view this as an imperative, issued by the Resident Engineer for forbearance from 'unsafety' on the part of the CP supplier (L4).

The CP supplier has specialised knowledge, distinct from the other site workers. Given the restrictions on the location and size of the job, the supplier realises that the by lifting the formwork in its current design he would be responsible for a lack of safety - he would renege on his commitment to to forbear from 'unsafety'. His reasoning suggests a solution - to require the formwork to be lighter, and he communicates this to the Temporary Works Co-ordinator (L5).

Again, we are not focusing upon the reasoning of individual agents (a simple deductive KB appropriately populated would suffice for this example), so we are not interested in how the TW coordinator reaches the solution of lopping off one section, but only in how he communicates this to the Senior Engineer. Notice that the TW coordinator is proscribing a specific task, rather than a state of affairs (such as "make it lighter"), since only she has the requisite knowledge to reason about how to bring about the state of affairs. Her locution to the Senior Engineer would be captured by L6. The delegation route would then run from Senior Engineer to foremen (collectively) and from foreman to gangers (distributively) as before, through L7 and L8.

8. DISCUSSION AND FUTURE WORK

The model proposed by Kumar *et al.* [9] builds upon the logic of intention (as a persistent, relativised goal) proposed by Cohen & Levesque [4] (henceforth C&L). In C&L's model, a world is a "sequence (or course) of events, temporally extended infinitely in the past and future", and each event is associated with a specific person (the agent of the event). Formulae are then interpreted with respect to a particular point in a world; this interpretation gives a "snapshot" of the world at a particular time. An action is a sequence of events, and formulae of the form (DONE a) state that the performance of action a immediately preceded the point in a world at which it is interpreted. Each action is either a primitive action or a program composed from primitive actions using the normal operators of first-order dynamic logic [6] such as sequential action and iteration. The use of these special DONE predicates means that the program that was executed to attain this state is included within each snapshot of the world. Thus, each state has a unique history enforcing an arboreal structure of states [14].

In contrast, a Hamblinian world is a sequence of world states (sets of propositions in the case of this paper rather than predicates) connected by agents' deeds. These deed-agent assignments are the events of a Hamblinian world ("happenings" [5], or events that are unconnected to agents, are not considered here, but they could be viewed as deeds done by a special agent). The model presented in this paper uses two interpretation functions, $\mathcal{I}_S$ and $\mathcal{I}_D$, to obtain the state and the deed-agent assignments at a particular point in a Hamblinian world. Within this model, there is no need to have a record of the actions that have been performed by each agent to attain the state of affairs at some time point within the state. A particular state may appear in many histories simply because the clear ontological distinction between actions (or events) and states of affairs is reflected in the structure of a Hamblinian world — there is no need to conflate these concepts by recording special DONE predicates within each state.

As both Rescher [13] and Kumar *et al.* [9] point out, not only is it possible that the recipient of an imperative (or any other communicative act) be a group, but the issuer (or source) of that act may be a group (collective or distributive). Rescher [13] uses a number of examples to illustrate the various possibilities. These include: "*Group (Collective) to Group (Collective)* Court order to a corporation to divest itself of certain holdings (in violation of antitrust statutes)" [13, p. 13]. Jones and Sergot [8] use similar examples to illustrate their "counts as" connective; for example, "x's uttering the words 'I pronounce you man and wife' counts (in [society] s) as a means of guaranteeing that s sees to it that [the recipients of the declaration] are married". Jones and Sergot *ibid.* do not, however, confine their theory to communicative acts, but present a general theory of agents acting on behalf of a group. An analysis of the utterance of an imperative by an individual on behalf of a group may, therefore, be related to Jones and Sergot's [8] notion of "counting as in a society". This is a necessary element of a complete theory of delegation and responsibility that is outside the scope of this paper.

On commands in dialogue and the context of issuing commands, Rescher [13] states that, "[g]enerally speaking, the source should have some entitlement or authority for giving a command to its recipient". This means that a command (or any imperative) could be questioned by its recipient regarding the authority of the source and the grounds for it being issued. Understanding how the issuing of an imperative fits into the wider structure of inter-agent dialogue may influence the design of flexible agent communication protocols, and is, therefore, a relevant avenue for future research.

9. CONCLUSION

This paper develops earlier work on delegation and responsibility in two important ways. First, in developing a basic theory of agentative action, a complete axiomatisation of the two modalities S_x and T_x is presented and the applicability of providing these modalities with possible-worlds semantics is explored in detail. Second, the theory of delegation and responsibility is extended to account for the issuing of imperatives to (distributive and collective) groups. Thirdly, group responsibility is identified, and the situations that count as group fulfilment of imperatives and their resultant norms are characterised. Finally, the theory is demonstrated in an extended real-world scenario, and it is shown to be sufficiently flexible to handle a wide range of complex communicative and normative situations, whilst remaining sufficiently well defined and theoretically unadorned to support implementable agent reasoning capabilities.

10. REFERENCES

[1] N. Belnap and M. Perloff. The way of the agent. *Studia Logica*, 51(3/4):463–484, 1992.

[2] B. F. Chellas. *Modal logic*. CUP, 1980.

[3] B. F. Chellas. Time and modality in the logic of agency. *Studia Logica*, 51(3/4):485–517, 1992.

[4] P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artif. Intell.*, 42:213–261, 1990.

[5] C. L. Hamblin. *Imperatives*. Basil Blackwell, 1987.

[6] D. Harel. Dynamic logic. In *Handbook of Philosophical Logic*, volume II, pages 497–604. D. Reidel, 1984.

[7] J. F. Horty and N. Belnap. The deliberative stit: A study of action, omission, ability, and obligation. *J. of Philosophical Logic*, 24:583–644, 1995.

[8] A. I. J. Jones and M. J. Sergot. A formal characterisation of institutionalised power. *J. of the IGPL*, 4(3):429–445, 1996.

[9] S. Kumar, M. J. Huber, D. R. McGee, P. R. Cohen, and H. J. Levesque. Semantics of agent communication languages for group interaction. *AAAI*, pages 42–47, 2000.

[10] T. J. Norman and C. A. Reed. Delegation and responsibility. In *Intelligent Agents VII*, volume 1986 of *LNAI*. Springer-Verlag, 2001.

[11] I. Pörn. *The logic of power*. Basil Blackwell, 1970.

[12] M. J. Perry. *Distributed cognition and computer supported collaborative design: The organisation of work in construction engineering*. PhD thesis, Brunel University, 1997.

[13] N. Rescher. *The logic of commands*. Routledge, 1966.

[14] M. Shanahan. *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia*. MIT Press, 1997.

[15] G. H. von Wright. *An essay in deontic logic and the general theory of action*, volume 21 of *Acta philosophica Fennica*. North-Holland, 1968.

[16] D. N. Walton and E. C. W. Krabbe. *Commitment in dialogue: Basic concepts of interpersonal reasoning*. SUNY, 1995.

[17] M. J. Wooldridge. *Reasoning about rational agents*. MIT Press, 2000.

[18] M. Xu. On the basic logic of STIT with a single agent. *J. of Symbolic Logic*, 60(2):459–483, 1995.