Rights and commitment in multi-agent agreements

Timothy J. Norman[1*] Carles Sierra[2†] Nick R. Jennings[1]

[1] Dept. of Electronic Engineering,
Queen Mary and Westfield College,
London E1 4NS, U.K.
{tim, nick}@elec.qmw.ac.uk

[2] Artificial Intelligence Research Institute,
Spanish Council for Scientific Research,
08193 Bellaterra, Barcelona, Spain.
sierra@iiia.csic.es

Abstract

For agents to act in collaboration, they often require an agreement that describes how they are to act, to which they are committed. Typically, agreements are characterised as an explicit course of action or a goal to be achieved. In this paper, it is argued such agreements may over specify the interaction required. To overcome this problem, a novel formalisation of agreements between agents is presented that is based on rights *as well as actions to be performed. Each agent that is involved in an agreement is bound to uphold the rights of others, and the implications of exercising rights and acting for others. It is argued that this approach provides agents with greater flexibility in the agreements they may reach, while retaining the necessary group commitment.*

1. Introduction

Agents are computational systems that inhabit and interact with dynamic, and not entirely predictable environments. They decide for themselves, on the basis of their individual beliefs, goals, etc., how to respond to the environment and other agents within it. However, it is often the case that an agent is motivated to achieve a goal that is only possible, made easier, or satisfied more completely by gaining the collaboration of others. For example, lifting a heavy table may not be possible without help. The process of gaining collaboration can take many forms. For a task such as lifting a table, it may be sufficient for one agent to simply ask another for help. However, many tasks require more detailed communication to generate an explicit mutually acceptable agreement through negotiation [14]. In general, the desired result of negotiation is an agreement to which each agent involved is committed that describes how they are to act; i.e. a state of group, social or joint commitment.

There are a number of views on how states of joint commitment between agents should be characterised (see Castelfranchi [2], Cohen *et al.* [4], and Jennings [12] for examples). A typical theory provides a characterisation of the state of group, social or joint commitment, and a description of the circumstances in which it should be reviewed. In short, agents are specified as being committed, as a group, to achieve some goal or to execute some plan. In many cases this is adequate, but in others it falls short [15]. Consider an agent that wishes to query the members of a special interest group (SIG). To do this, it requires permission from the representative of that group. If the agent queries the group without permission, it may subsequently be prevented from interacting with the SIG. In other words, the agent is capable of querying the group, but if it does so without permission, it may incur a penalty or sanction, such as having further messages blocked. Suppose that the agent gains the agreement of the representative to enable it to query the group, but if it does, it must provide a summary of the results of that query back to the group. In this case, the agent is permitted the right to query the group, and it commits to the group to provide a summary of the results in return. Here, neither the agent that wishes to query the members of the group nor the members themselves are necessarily committed to any specific course of action. However, *if* the agent does query the group (i.e. exercises its right) then it is committed to providing the group with a summary. It may be possible for this agreement to be described in terms of a joint commitment to a plan of action. However, for the

*Partially funded by EPSRC GR/L09714 (Trilogy).
†Partially funded by ESPRIT LTR 25500 (COMRIS) and the Spanish CICYT project TIC96-1138-C04-01 (SMASH).

goal to have satisfied the query to be achieved, the action of the representative, the members of the SIG, and the querying agent must be specified and scheduled in advance. Agents within the group that are capable of satisfying the query must be identified, their cooperation ensured, and their future action within the joint plan determined. However, between the generation of such a plan, and the completion of its execution, the world may change in such a way that re-planning is necessary. For example, a member of the SIG may become unavailable, or the querying agent may discover a more effective method of achieving its goal. Furthermore, without a notion of rights, as distinct from capabilities, the agent intending to satisfy its goals by querying the SIG must be considered 'incapable' without permission.

This paper addresses the question of how agreements between agents may be formulated so that the flexible interaction that is required may be specified. A novel formalisation of agreements between agents based on rights, as well as actions to be performed is presented, while the necessary commitment is captured through the agents, as a group, being bound to uphold the agreement. The concept of a right and the formation of agreements from rights are introduced in section 2. Then, a language for describing rights, agreements, and a notion of commitment to such agreements is specified in section 3. In section 4, it is shown how an agent theory may be developed using this language, which meets the desirable properties of such agreements identified in section 2. In section 5, related work is reviewed, and section 6 concludes and indicates directions for future work.

2. Rights and agreements

It is common for agents to be characterised in terms of their capabilities; i.e. the actions that they can perform. However, although an agent is capable, it may not have the right to perform an action. For example, members of parliament may be capable of asking questions of the government, but do not have the right to do so unless they are recognised by the speaker (or whomever has the power to permit such a right). An agent's role in a multi-agent system provides it with a set of rights to perform actions. Thus, the actions that an agent can use to achieve its goals without risking penalties or sanctions imposed by other members of the agent society are determined by those actions that it is capable of performing and has the right to perform. If it is not possible for an agent to achieve its goals within these restrictions, it may need to gain permission to perform an action that it does not normally have the right to do, or to gain the agreement of another agent to act for it. If an agent has the capability, and the right to perform some action, then it may also have the power to perform that action for another agent. It may also be possible for an agent to have the power to give another the right to perform some action. For example, the representative of a SIG is given the power by that group to give other agents that are not members of the group the right to communicate with the group's members. Without this permission, a message to the group from a non-member agent may incur sanctions on that non-member.

Agreements between agents can now be created through the combination of *rights* and actions. However, it is essential to ensure that the agents involved in an agreement act in accordance with it — a notion of commitment must be an integral part of any theory of multi-agent agreements. Thus, an agent that is *bound to uphold* (i.e. committed to) an agreement should act in accordance with the agreement, uphold the rights of others, and accept the consequences of actions being performed and rights being exercised. Furthermore, to ensure the stability of group activity, the agents that are involved in an agreement must, *as a group*, be bound to uphold the agreement [1, 2]. For example, if an agent is bound to uphold the right of another to perform an action, then it should not act to prevent this right being exercised. For these reasons, in the language for agreements presented in the following section, individual and group commitment (i.e. Bound and G-Bound section 3.2) constrain the behaviour of any agent built using this specification.

Given these intuitions about rights and commitment in multi-agent agreements, what are the desirable properties of an agent that operates in these terms? Here, we consider the following example properties, which will be formalised in section 4.

'Moral' free will An agent is free to act for another if it is capable of performing the action, *and* has the right to do so.

Delegation Consider the following examples: (1) An elected member of a parliament has the right to ask questions of members of the government on behalf of an electorate. However, this does not imply that the member has the power to give another the right to do so. (2) In the futures market, a trader that has an option on (or the right to buy) a certain quantity of a commodity has the power to delegate this right to any other trader. These are two alternative views on the delegation of rights that agent theories may model.

Persistence Consider the following examples: (1) The right to buy a certain quantity of a commodity may only be exercised once. (2) The right to vote may be exercised many times once it is permitted until it is explicitly revoked. Again, there are two alternatives: one-shot and persistent rights.

3. A language for describing agreements

This section presents a formal characterisation of the types of agreements that may hold between agents, and how these are composed from rights and actions.

3.1. Syntax

Here, we use three basic sets: propositional variables, P, agents, *Agents*, and actions, *Actions*, to construct the syntax of our language for agreements, $\mathcal{L}$. The symbols p and q are used to denote propositional variables, x, y and z denote agents, and a and b denote individual actions. We assume a STRIPS-like action model [7], where an action $a \in Actions$ has a set of preconditions, denoted by $\mathsf{Pre}(a)$, and a set of effects, $\mathsf{Eff}(a)$. Each of these are sets of atomic propositions of our language $\mathcal{L}$ (i.e. members of the set Φ_0). The performance of an action in a state where the preconditions of that action are satisfied produces the effects specified; i.e. we assume that actions do not fail. Actions are either communicative or non-communicative. Non-communicative (or physical) actions are viewed as primitive. Communicative actions consist of an illocutionary particle, which is a member of the set $I = \{\mathsf{Permit}, \mathsf{Commit}, \ldots\}$, the sender and recipient, each of which are agents, and the content, which is a member of the set of atomic propositions, Φ_0 (def. 2). For example, the action of x permitting y the right to do a is expressed as $\mathsf{Permit}(x, y, \mathsf{Right}(y, x, a))$.

The language, $\mathcal{L}$, is based on dynamic logic, because we want to be able to talk about agents performing actions, action sequences, etc. The members of the set of atomic action expressions, Π_0, associate agents with actions, where a_x denotes the event of agent x doing action a.

Definition 1 Given a set of agents, *Agents*, and a set of actions, *Actions*, we define the set of atomic action expressions (or programs), $\Pi_0 = \{a_x : x \in Agents, a \in Actions\}$.

Agents are characterised in terms of their capabilities and rights. We denote that an agent, x, is capable of performing an action, a, by $\mathsf{Capable}(x, a)$, and C is used to refer to the set of potential capabilities of agents within the system. As discussed in section 2, rights characterise those actions that an agent can legally perform. $\mathsf{Right}(x, y, a)$ is used to denote the right, permitted by y, of agent x to do a, R is the set of potential rights. A right may be an inherent property of an agent (possibly defined by the agent's role in a society), or one permitted by some other agent. In addition to the capabilities and rights of agents, we wish to talk about actions actually having been performed by agents. $\mathsf{Done}(x, a)$ is used to denote that agent x has just performed action a, and D is used to denote the set of such formulae.

Actions to be performed and rights that may be exercised are combined to form agreements. A is used to refer to the set of agreements, and ag with annotations to denote members of this set. The agents involved in an agreement, denoted by $\mathsf{Agts}(ag)$, if they are to act in accordance with it, must be committed in some way to that agreement. Therefore, the notion of an agent being bound to an agreement, $\mathsf{Bound}(x, ag)$, is introduced. $\mathsf{Bound}(x, ag)$ is read

"x is bound to uphold the agreement ag". The formula, $\mathsf{Bound}(x, ag)$, captures x's commitment to act on the agreement, uphold the consequences of so acting (or not acting as the case may be), and to uphold the rights of others stated in that agreement. B is used to denote the set of such bindings.

Now, the set of atomic propositional variables, Φ_0, of the language $\mathcal{L}$ is defined in terms of the set of propositional variables, P, atoms expressing that an action has just been done by an agent, D, the set of capabilities, C, rights, R, agreements, A, and bindings, B.

Definition 2 Given a set of agents, *Agents*, a set of actions, *Actions*, and a set of atomic programs, Π_0, the set of atomic propositions, Φ_0, of $\mathcal{L}$ is defined as follows.

1. P is a set of propositional variables.
2. $D = \{\mathsf{Done}(x, a) : x \in Agents, a \in Actions\}$ is a set of atoms expressing that an action has just been performed by an agent.
3. $C = \{\mathsf{Capable}(x, a) : x \in Agents, a \in Actions\}$ is a set of capabilities.
4. $R = \{\mathsf{Right}(x, y, a) : x, y \in Agents, a \in Actions\}$ is a set of rights.
5. A is a set of agreements, such that:
   (a) **if** $\Gamma \subseteq P \cup D$ **and** $\varphi \in \Pi_0 \cup R$ **then** $\mathsf{Agree}(\Gamma, \varphi) \in A$
   (b) **if** $ag, ag' \in A$ **then** $ag \,\&\, ag' \in A$
6. $B = \{\mathsf{Bound}(x, ag) : x \in Agents, ag \in A\}$ is a set of bindings of agents to agreements.

7. $\Phi_0 = P \cup D \cup C \cup R \cup A \cup B$.

We may now define the set Φ of compound formulae and the set Π of compound programs of $\mathcal{L}$.

Definition 3 Given a set of atomic propositions, Φ_0, and atomic programs, Π_0, the set of compound formulae, Φ, and compound programs, Π, of $\mathcal{L}$ is defined as follows.

1. $\mathsf{True}, \mathsf{False} \in \Phi$, $\Phi_0 \subseteq \Phi$.
2. **if** $\varphi_1, \varphi_2 \in \Phi$ **then** $\neg\varphi_1 \in \Phi$ **and** $(\varphi_1 \vee \varphi_2) \in \Phi$.
3. **if** $\varphi \in \Phi$ **and** $\alpha \in \Pi$ **then** $\langle\alpha\rangle\varphi \in \Phi$.
4. $\Pi_0 \subseteq \Pi$.
5. **if** $\alpha, \beta \in \Pi$ **then** $(\alpha;\beta), (\alpha \cup \beta), \alpha* \in \Pi$. (Meaning do α followed by β, do either α or β non deterministically, and repeat α a finite, but non deterministic number of times, respectively.)
6. **if** $\varphi \in \Phi$ **then** $\varphi? \in \Pi$. (Meaning proceed if φ is True else fail.)

The formula $\langle\alpha\rangle\varphi$ means that it is possible to execute α reaching a situation in which φ is true. $[\alpha]\varphi$ is the usual modal abbreviation for $\neg\langle\alpha\rangle\neg\varphi$, which means that every possible execution of α leads to a situation in which φ is true.

This section has introduced the syntax of a language, $\mathcal{L}$, for describing agreements between agents that are based on rights as well as actions. The semantics of capabilities, rights, agreements, and our notion of being bound to uphold an agreement (either individually or as a group) are presented in the following section.

3.2. Semantics

The semantics of the language of agreements, $\mathcal{L}$, is based on a possible worlds model [11]. We first define a class of models, $\mathcal{M}$.

Definition 4 $\mathcal{M} = \langle \mathcal{W}, \tau, \rho \rangle$ is a 3-tuple associating the possible multi-agent world states, $\mathcal{W}$, a function that assigns truth values to formulae, τ, and an accessibility relation associated to programs, ρ.

1. $\mathcal{W} = \mathcal{W}_{x_1} \times \mathcal{W}_{x_2} \times \ldots \times \mathcal{W}_{x_n}$ with $x_i \in \mathit{Agents}$ is the set of tuples representing the possible multi-agent worlds, or states, where each multi-agent state is a tuple of individual agent states. $\mathcal{W}_{x_i}$ is used to denote the set of possible worlds from x_i's perspective.
2. $\tau : \Phi \rightarrow 2^{\mathcal{W}}$ is a function that assigns to formulae, the set of multi-agent worlds in which they hold.
3. $\rho_x : \Pi \rightarrow 2^{(\mathcal{W}_x \times \mathcal{W}_x)}$ is a function, defined for each agent, such that for the atomic program $a_y \in \Pi_0$, $\rho_x(a_y)$ provides an accessibility relation over $\mathcal{W}_x$ associated with the event of y doing a. For instance, if there is a state $\omega_x \in \mathcal{W}_x$ in which the execution of action a by agent y produces a state $\omega'_x \in \mathcal{W}_x$, then $(\omega_x, \omega'_x) \in \rho_x(a_y)$. For compound programs in Π, this accessibility relation is defined as usual in the semantics of dynamic logic languages [10].

   From $\{\rho_x\}_{x \in Agents}$, we define $\rho : \Pi \rightarrow 2^{(\mathcal{W} \times \mathcal{W})}$ as: $\rho(\alpha) = \{(\langle \omega_{x_1}, \omega_{x_2}, \ldots, \omega_{x_n} \rangle, \langle \omega'_{x_1}, \omega'_{x_2}, \ldots, \omega'_{x_n} \rangle) : (\omega_{x_i}, \omega'_{x_i}) \in \rho_{x_i}(\alpha)\}$

Now, focusing on the actions performed by agents within a multi-agent system, we define the set of paths along which the state of the multi-agent system may pass. This set of paths is used to define our notion of commitment; i.e. what it means for an agent to be bound to uphold an agreement (see constraints 6–11 in table 1).

Definition 5 The set of possible paths starting at state ω along which the multi-agent system may pass, Paths_ω, is defined as:

$$\mathit{Paths}_\omega = \{\langle \omega, \ldots, \omega^i, \omega^{i+1}, \ldots \rangle : \omega^i, \omega^{i+1} \in \mathcal{W}, \textbf{ and } \exists \alpha \in \Pi_0 \textbf{ s.t. } (\omega^i, \omega^{i+1}) \in \rho(\alpha)\}$$

The precedence relation $\prec_p$ is defined for multi-agent states along a path p as $\omega^i \prec_p \omega^j$ **iff** $i < j$ ($\preceq_p$ is similarly defined). We say that $\omega \prec \omega'$ **iff** $\exists p \in \mathit{Paths}_\omega$ such that $\omega \prec_p \omega'$.

For our purposes the class of models of $\mathcal{L}$ that we are interested in are those that satisfy constraints 1–11 (see table 1). To determine whether some action (or action sequence) has actually been performed, the history of what actions have been performed and by what agents must be recorded in each state [5]. Therefore, as each state has a unique history, state transitions must be strictly diverging (C1), so we have a branching structure of states. For our purposes, we require that the action that has just been done is recorded within each state (C2), and that actions never fail (i.e. the effects of an action hold whenever it has just been done) (C3).

C4 means that agreements involving an action that must be done may be true only in those states from which there is at least one course of action leading to the performance of that action. In other words, agents cannot agree on the performance of impossible actions. C5 is the obvious constraint for conjuncts of such agreements. Note that we do not restrict agreements concerning rights in this way. The reason for this is that it is not necessary for rights to be exercised. This may

C1 **if** $(\omega', \omega) \in \rho(a)$ **and** $(\omega'', \omega) \in \rho(b)$ **then** $\omega' = \omega''$ **and** $a = b$
C2 $\omega \in \tau(\mathsf{Done}(x, a))$ **iff** $\exists \omega' \in \mathcal{W} : \omega' \prec \omega$ **and** $(\omega', \omega) \in \rho(a_x)$
C3 **if** $\omega \in \tau(\mathsf{Done}(x, a))$ **then** $\omega \in \tau(\mathsf{Eff}(a))$
C4 $\omega \in \tau(\mathsf{Agree}(\Gamma, a_x))$ **iff** $\exists \omega', \omega'' \in \mathcal{W} : \omega \preceq \omega' \prec \omega''$ **and** $\omega' \in \tau(\Gamma)$ **and** $\omega'' \in \tau(\mathsf{Done}(x, a))$
C5 $\omega \in \tau(\mathsf{Agree}(\Gamma, a_x) \,\&\, \mathsf{Agree}(\Gamma', a_y))$ **iff** $\omega \in \tau(\mathsf{Agree}(\Gamma, a_x)) \cap \tau(\mathsf{Agree}(\Gamma', a_y))$
C6 $\omega \in \tau(\mathsf{Bound}(x, \mathsf{Agree}(\emptyset, a_x)))$ **iff** $\forall p \in Paths_\omega \; \exists \omega' \in p : \omega \prec_p \omega'$ **and** $\omega' \in \tau(\mathsf{Done}(x, a))$
C7 $\omega \in \tau(\mathsf{Bound}(x, \mathsf{Agree}(\emptyset, a_y)))$ **iff** $\exists \omega' \in \mathcal{W} : \omega \prec \omega'$**and** $\omega' \in \tau(\mathsf{Done}(y, a))$ **and** $\neg\exists \omega'', \omega''' \in \mathcal{W} : \omega \preceq \omega'' \prec \omega''' \prec \omega'$ **and** $(\omega'', \omega''') \in \rho(a_x)$ **and** $\exists \varphi \in \mathsf{Pre}(a) : \omega'' \in \tau(\varphi), \omega''' \notin \tau(\varphi)$
C8 **if** $\omega \in \tau(\mathsf{Bound}(x, \mathsf{Agree}(\emptyset, \mathsf{Right}(x, y, a))))$ **then** $\omega \in \tau(\mathsf{Right}(x, y, a))$
C9 **if** $\omega \in \tau(\mathsf{Bound}(x, \mathsf{Agree}(\emptyset, \mathsf{Right}(y, z, a))))$ **then** $\omega \in \tau(\mathsf{Bound}(x, \mathsf{Agree}(\emptyset, a_y)))$
C10 $\omega \in \tau(\mathsf{Bound}(x, \mathsf{Agree}(\Gamma, \varphi)))$ **iff** $\forall \omega' \in \mathcal{W}$ **if** $\omega \prec \omega'$ **and** $\omega' \in \tau(\Gamma)$ **then** $\omega' \in \tau(\mathsf{Bound}(x, \mathsf{Agree}(\emptyset, \varphi)))$
C11 $\omega \in \tau(\mathsf{Bound}(x, ag \,\&\, ag'))$ **iff** $\omega \in \tau(\mathsf{Bound}(x, ag)) \cap \tau(\mathsf{Bound}(x, ag'))$

Table 1. Constraints on the class of models of $\mathcal{L}$.

be useful if rights can be delegated (see section 4); an agent that cannot exercise a right may delegate that right to another that can.

C6 to C11 further constrain the set of models of $\mathcal{L}$ that we permit by characterising the state of an agent being bound to uphold an agreement. First, consider an agent that is bound to uphold an agreement that is conditional on Γ, where $\Gamma \subseteq P \cup D$ (see def. 2). Once these conditions hold, then the agent becomes unconditionally bound to that agreement (C10). For example, if agents x and y are bound to uphold an agreement in which x has the right to demand payment for doing b that is conditional on it having done b, then once $\mathsf{Done}(x, b)$ is true both agents are unconditionally bound to an agreement in which x has that right.

Consider an agreement that is unconditional, and which states that agent x must perform action a. From x's perspective, this means that its possible future paths are constrained to those in which it does a (C6). From the perspective of some other agent bound to uphold this agreement, it will not act to prevent x fulfilling its commitment. Constraint 7 states that an agent, y (where it is assumed that $x \neq y$), will not perform some action that deletes a precondition of the action, a, that must be done by x (this is often referred to as a clobbering action in the planning literature). Now suppose that the agreement is unconditional, and provides agent x with the right to do a. From the perspective of some other agent, y, it is similarly constrained not to perform a clobbering action (C9). From the perspective of the agent that is permitted the right to do some action under an agreement, the agent, x, has the right to do the action, a, (C8). For example, an agent that is unconditionally bound to an agreement in which it has the right to demand payment for doing b, will have that right. Finally, an agent that is bound to uphold a conjunction of agreements is bound to uphold each conjunct (C11).

A1 $[a_x]\mathsf{Done}(x, a)$
A2 $[a_x]\neg\mathsf{Done}(y, b)$, if $b \neq a$ or ($b = a$ and $x \neq y$)
A3 $\mathsf{Done}(x, a) \rightarrow \mathsf{Eff}(a)$
A4 $[\alpha]\neg\mathsf{Done}(x, a) \rightarrow \neg\mathsf{Agree}(\Gamma, a_x)$ for all $\alpha \in \Pi$ and $\Gamma \subseteq P \cup D$
A5 $\mathsf{Bound}(x, \mathsf{Agree}(\emptyset, \mathsf{Right}(x, y, a))) \rightarrow \mathsf{Right}(x, y, a)$
A6 $\mathsf{Bound}(x, \mathsf{Agree}(\emptyset, \mathsf{Right}(y, z, a))) \rightarrow \mathsf{Bound}(x, \mathsf{Agree}(\emptyset, a_y))$
A7 $\mathsf{Bound}(x, \mathsf{Agree}(\Gamma, \varphi)) \rightarrow ([\alpha]\Gamma \rightarrow [\alpha]\mathsf{Bound}(x, \mathsf{Agree}(\emptyset, \varphi)))$
A8 $\mathsf{Bound}(x, ag \,\&\, ag') \leftrightarrow \mathsf{Bound}(x, ag) \wedge \mathsf{Bound}(x, ag')$

Table 2. Some Axioms of $\mathcal{L}$.

To ensure the stability of group activity, the agents that are involved in an agreement require some kind of group commitment to act in accordance with the agreement. Each agent that is involved in an agreement, should be committed to do its part and not to prevent others from so doing. Thus, we introduce the notion of a group of agents being bound to an agreement. The formula $\mathsf{G\text{-}Bound}(\mathsf{Agts}(ag), ag)$ denotes the state of group commitment that will ensure that each agent involved in ag will be bound to uphold that agreement. Hence, the state of the agents involved in an agreement being bound as a group to uphold the agreement is a distributed state in which each agent in the group is bound to that agreement.

Definition 6 A group of agents, $X \subseteq Agents$, being $\mathsf{G\text{-}Bound}$ by an agreement, ag, is defined as the conjunction of the propositions $\mathsf{Bound}(x, ag)$ for each x in X.

$$\mathsf{G\text{-}Bound}(X, ag) \stackrel{\text{def}}{=} \bigwedge_{x \in X} \mathsf{Bound}(x, ag)$$

3.3. Axiomatics

It remains now to provide an axiomatic system for $\mathcal{L}$. The language is defined upon the sets of atomic formulae Φ_0 (def. 2) and atomic programs Π_0 (def. 1). The

additional axioms of $\mathcal{L}$, each of which correspond to a constraint on the class of models we are considering, are A1–8 in table 2. In appendix A, we show certain formal properties of this language. This completes the introduction to the language, $\mathcal{L}$.

4. Meeting the requirements

The language, $\mathcal{L}$, enables the specification of agents that may generate, commit to, and once committed, act in accordance with agreements. In this language, the bindings between a set of agents, $\mathsf{Agts}(ag)$, and an agreement, ag, that is a formula of the form $\mathsf{G\text{-}Bound}(\mathsf{Agts}(ag), ag)$, represents a form of social commitment between those agents. Here, we propose a number of axioms that capture our intuitions concerning the relationship between capabilities and rights, morality, and the delegation and persistence of rights (see section 2). Note that the constraints on models of $\mathcal{L}$ (C6–11) are designed so that any agent that conforms to this specification will act to uphold agreements once bound to them.

First, we introduce the communicative action Commit, and indicate how agents can place conditions on their commitments (see Norman & Jennings [15] for more detail on the notion of conditional commitment in negotiation). Suppose that a necessary effect of the action $\mathsf{Commit}(x, y, ag)$ is to commit the speaker, x, to the hearer, y, to uphold the agreement, ag.

A9 $[\mathsf{Commit}(x, y, ag)]\mathsf{Bound}(x, ag)$

Now, we can express that an agent, which is capable of acting for another under some set of conditions, Γ, is free to set extra conditions on such an agreement (remember that $\varphi \in \Pi_0 \cup R$, see def. 2).

A10 $\mathsf{Right}(x, y, \mathsf{Commit}(x, z, \mathsf{Agree}(\Gamma, \varphi))) \rightarrow$
$\mathsf{Right}(x, y, \mathsf{Commit}(x, z, \mathsf{Agree}(\Gamma', \varphi)))\ \forall \Gamma' \supseteq \Gamma$

4.1. Capabilities and Rights

The capabilities of agents (written $\mathsf{Capable}(x, a)$) correspond to those actions that an agent is capable of performing whenever the preconditions of the action hold. So we define $\mathsf{Capable}(x, a) \equiv \mathsf{Pre}(a) \rightarrow \langle a_x \rangle \mathsf{True}$. Note that capabilities are properties of an agent (they do not depend on the world state), and hence capabilities are defined as part of the theory of each agent. Then, in the theory of some agent x (i.e. T_x), it will be true, that if the preconditions of an action hold and the agent is capable of doing the action then the action may be executed by that agent, i.e. $T_x \vdash \mathsf{Pre}(a) \wedge \mathsf{Capable}(x, a) \rightarrow \langle a_x \rangle \mathsf{True}$. The theory, T_x, will then contain the equivalence of this as an axiom.

A11 $\mathsf{Capable}(x, a) \leftrightarrow (\mathsf{Pre}(a) \rightarrow \langle a_x \rangle \mathsf{True})$

The rights of agents are distinct from capabilities in that they describe those actions that the agent can legally perform. Therefore, a moral agent, if it does not have the right to perform some action, will not do so (axiom 12).

A12 $\neg\mathsf{Right}(x, y, a) \rightarrow [\alpha]\neg\mathsf{Done}(x, a)\ \forall \alpha \in \Pi$

It may be useful in a particular agent theory for the rights of agents to be persistent; i.e. once an agent receives a right, it keeps it forever. This can be modelled by axiom 13. Note that this is different from the persistence of rights expressed in axiom 17, section 4.4.

A13 $\mathsf{Right}(x, y, a) \rightarrow [\alpha]\mathsf{Right}(x, y, a)$, for all $\alpha \in \Pi$

4.2. 'Moral' free will

The property of moral free will described in section 2 can be expressed by an axiom that states that if the agent is capable of doing a *and* has the right to do so, then it has the right to commit to do a for some other agent.

A14 $\mathsf{Capable}(x, a) \wedge \mathsf{Right}(x, y, a) \rightarrow$
$\mathsf{Right}(x, y, \mathsf{Commit}(x, z, \mathsf{Agree}(\emptyset, a_x)))$

4.3. Delegation

Suppose that a necessary effect of the communicative action $\mathsf{Permit}(x, y, \mathsf{Right}(y, x, a))$ is to give permission to the hearer, y, the right to perform a. (Note that in committing to an agreement, an agent may implicitly permit another the right to perform some action.) Therefore, to express the property of automatic delegation of rights, axiom 15 may be added to the theory. If automatic delegation is to be prevented within the theory, the the negation of this axiom is added.

A15 $\mathsf{Right}(x, y, a) \rightarrow$
$\mathsf{Right}(x, y, \mathsf{Permit}(x, z, \mathsf{Right}(z, x, a)))$

4.4. Persistence

We wish to allow theories in which the binding to uphold rights is either one-shot (i.e. may only be exercised once), or persistent for the duration of the commitment (i.e. may be exercised an unlimited number of times while the agent is bound to uphold the agreement in which it is permitted). Axioms 16 and 17 respectively express these two alternatives.

A16 $\mathsf{Bound}(x, \mathsf{Agree}(\emptyset, \mathsf{Right}(y, z, a))) \rightarrow$
$([\alpha]\mathsf{Done}(y, a) \rightarrow$
$[\alpha]\neg\mathsf{Bound}(x, \mathsf{Agree}(\emptyset, \mathsf{Right}(y, z, a))), \forall \alpha \in \Pi$

A17 $\mathsf{Bound}(x, \mathsf{Agree}(\emptyset, \mathsf{Right}(y, z, a))) \rightarrow$
$[\alpha]\mathsf{Bound}(x, \mathsf{Agree}(\emptyset, \mathsf{Right}(y, z, a))), \forall \alpha \in \Pi$

5. Related work

Castelfranchi [2] distinguishes between individual and social commitment, arguing that "a social commitment is not an individual commitment shared by many agents" (see also Bratman [1]). Social commitment (or S-COMM) is defined as a relation between three agents and an action: (S-COMM x y a z), where x is committed to y to do a, and z is the agent before whom x is committed. The important relationship between Castelfranchi's notion of social commitment and G-Bound (def. 6) is the idea that the agents involved are bound to uphold certain *rights* of others. For example, in the state (S-COMM x y a z) [2], x is committed to uphold y's right to expect it to do a, and to protest if a is not done. This notion of committing as a "rights-producing act" is related to our notion of an agent, y, committing to an agreement in which it must uphold the right of another, x. In the work presented here, y is constrained not to act against x exercising its right (see section 3). However, we consider rights that are expressed explicitly within an agreement. Rights and actions are combined to form complex agreements so that the consequences of exercising a right may commit the agent to performing some other action, or it may permit another agent some other right. Thus, not only may committing be a rights-producing act, but exercising rights or acting within the context of an agreement may also be a rights-producing act because the agent is bound to uphold that agreement. Castelfranchi also discusses the act of giving permission as a kind of passive help [3]. Within the model presented here, an agent that is bound to uphold an agreement in which another has the right to perform some action is committed not to prevent this right from being exercised. Thus, states of commitment and the act of permitting are closely related.

Deontic logic [13] is a branch of modal logic that is concerned with reasoning about normative behaviour, using modalities such as prohibition, permission and obligation. Hence, this is related to the intuitions about rights and commitments discussed here. However, a number of problems have been identified in using such a system for characterising the act of giving permission [3]. Permission is commonly defined as the dual of obligation [13]; i.e. an agent that is not obliged not to do a is permitted to do a. In contrast, we assume that an agent has the rights to perform some action only if its theory contains the appropriate formula. Furthermore, if an agent does not have the right to do a, it must gain the permission of an agent that may permit it to do a. An agent will not have the right to do this action simply because it is not obliged not to do it.

It was mentioned in section 1 that theories of joint commitment typically bind the agents involved to achieving goals or executing actions in the pursuit of a shared goal. There is a great deal of research on the recognition of states in which there is the potential for cooperation (e.g. Grosz & Kraus [8], Haddadi [9] and Wooldridge & Jennings [18]), on team formation (e.g. Cohen *et al.* [4] and Tambe [16]), and on characterisation of states of joint commitment (e.g. Cohen *et al.* [4], Dunin-Kęplicz & Verbrugge [6] and Jennings [12]). Our work does not attempt to provide another theory of joint commitment, but focuses on what agents are committed to. The work cited above shares the idea that agents are committed (either jointly, socially, or as a group or collective, depending on the terminology used by various authors)[1] to achieve a goal, or to execute a plan. Here, we suggest that the agreements to which a group of agents are jointly committed may be extended by introducing explicit rights within agreements. This provides a greater degree of flexibility in agreements, while retaining the necessary commitment to group activity.

6. Conclusions and future work

This paper has introduced a novel formulation of agreements between agents as combinations of rights and actions to which the agents involved are bound *as a group*. A language for agreements, $\mathcal{L}$ has been presented in which agents are constrained to act to uphold the rights of others and act in accordance with an agreement to which it is bound. A number of properties (*morality*, *delegation* and *persistence*) are discussed, and it is shown how these may be introduced as axioms of a theory of agency. It is argued that the intuitions captured in this model provide a flexible way of describing agreements between agents, while retaining a notion of joint commitment, which is widely recognised as necessary to ensure that agents act on their agreements.

There are a number of avenues for future development of this model. We do not consider an agent's motivation for seeking collaboration, the generation of commitment, or the monitoring and revision of such commitments. At present, the specification does not fully account for the loss of commitment [17]. Reasons for a loss of commitment may be that it is no longer

[1] The different terms reflect (sometimes subtly) different concepts.

possible to perform an action or exercise a right, that the agents involved jointly agree to no longer be bound by the agreement, or that the agreement has expired for some other reason (e.g. a deadline). For example, it is unreasonable to expect the right to cash a cheque to be valid indefinitely.

Acknowledgements. We would like to thank Lluis Godo, Mike Wooldridge and the anonymous reviewers for their useful comments on earlier drafts of this paper.

References

[1] M. E. Bratman. Shared cooperative activity. *The Philosophical Review*, 101(2):327–341, 1992.

[2] C. Castelfranchi. Commitments: From individual intentions to groups and organisations. In *Proc. 1st Int. Conf. on Multi-Agent Systems*, pages 41–48, 1995.

[3] C. Castelfranchi. Practical "Permission": Dependence, power, and social commitment. In *Proc. 2nd Workshop on Practical Reasoning and Rationality*, Manchester, UK., 1997.

[4] P. R. Cohen, H. J. Levesque and I. A. Smith. On team formation. In J. Hintikka and R. Tuomela, eds., *Contemporary Action Theory*, Synthese, in press.

[5] F. Dignum and B. van Linder. Modelling social agents: Communication as action. In J. P. Müller, M. J. Wooldridge, and N. R. Jennings, eds, *Intelligent Agents III (ATAL '96)*, vol. 1193 of *LNAI*, pp. 205–218. Springer-Verlag, 1996.

[6] B. Dunin-Kęplicz and R. Verbrugge. Collective commitments. In *Proc. 2nd Int. Conf. on Multi-Agent Systems*, pages 56–63, 1996.

[7] R. E. Fikes and N. J. Nilsson. STRIPS: A new approach to the application of theorem proving to problem solving. *Artif. Intell.*, 2:189–208, 1971.

[8] B. J. Grosz and S. Kraus. Collaborative plans for complex group activity. *Artif. Intell.*, 86(2):269–357, 1996.

[9] A. Haddadi. *Communication and cooperation in agent systems: A pragmatic approach*, vol. 1056 of *LNAI*. Springer-Verlag, 1996.

[10] D. Harel. Dynamic logic. In D. Gabbay and F. Guenther, eds, *Handbook of Philosophical Logic Volume II*, pages 497–604. D. Reidel Publishing Company, 1984.

[11] J. Hintikka. *Knowledge and belief*. Cornell University Press, 1962.

[12] N. R. Jennings. Commitments and conventions: The foundations of coordination in multi-agent systems. *Knowledge Engineering Review*, 8(3):223–250, 1993.

[13] J.-J. Ch. Meyer and R. J. Wieringa, eds. *Deontic logic in computer science: Normative system specification*. Wiley, 1993.

[14] H. J. Müller. Negotiation principles. In G. M. P. O'Hare and N. R. Jennings, eds, *Foundations of Distributed Artificial Intelligence*, pages 211–229. Wiley, 1996.

[15] T. J. Norman and N. R. Jennings. Generating states of commitment between autonomous agents. In *Proc. 3rd Australian Workshop on Distributed Artificial Intelligence*, Perth, WA, 1997. to appear.

[16] M. Tambe. Towards flexible teamwork. *Journal of Artificial Intelligence Research*, 7:83–124, 1997.

[17] D. N. Walton and E. C. W. Crabbe. *Commitment in dialogue*. State University of New York Press, 1995.

[18] M. J. Wooldridge and N. R. Jennings. Cooperative problem solving. *Journal of Logic and Computation*, 1998. to appear.

A. Soundness

Theorem 1 Let $\mathcal{M} = \langle \mathcal{W}, \tau, \rho \rangle$ belong to the class $\mathcal{C}^{\mathcal{L}}$ of standard PDL models. Then axioms (1)-(8) are sound.

Proof

1. Axiom 1 is valid in $\mathcal{M}$ iff for all $\omega \in \mathcal{W}$ $\mathcal{M}, w \models [a_x]\mathsf{Done}(x, a)$, and this is so iff for all $\omega' \in \mathcal{W}$ if $(\omega, \omega') \in \rho(a_x)$ then $\omega' \in \tau(\mathsf{Done}(x, a))$. This is exactly C2 read from right to left. Thus, $\omega' \in \tau(\mathsf{Done}(x, a))$, and hence $\mathcal{M}, w \models [a_x]\mathsf{Done}(x, a)$.
2. Axiom 2 is valid in $\mathcal{M}$ iff for all $\omega \in \mathcal{W}$, $\mathcal{M}, \omega \models [a_x]\neg\mathsf{Done}(y, b)$ if $b \neq a$ or ($b = a$ and $x \neq y$). Let us assume the contrary (i.e. $\mathcal{M}, \omega \not\models [a_x]\neg\mathsf{Done}(y, b)$) and make an analysis by cases:
   (a) $b \neq a$. With our assumption $\mathcal{M}, \omega \not\models [a_x]\neg\mathsf{Done}(y, b)$ we have $\mathcal{M}, \omega \models \neg[a_x]\neg\mathsf{Done}(y, b)$ which is equivalent to $\mathcal{M}, \omega \models \langle a_x \rangle \mathsf{Done}(y, b)$ which means that $\exists \omega' \in \mathcal{W}$ such that $(\omega, \omega') \in \rho(a_x)$ and $\mathcal{M}, \omega' \models \mathsf{Done}(y, b)$, but then we have by C2 that $\exists \omega'' \in \mathcal{W}$ such that $(\omega'', \omega') \in \rho(b_y)$, but by C1 we have that $b = a$ obtaining a contradiction.
   (b) $b = a$ and $x \neq y$. Similar to the previous case.
3. The proof of axiom 3 is straight forward from constraint 3.
4. Axiom 4 is valid in $\mathcal{M}$ iff for all $\alpha \in \Pi$ and for all $\omega \in \mathcal{W}$ $\mathcal{M}, \omega \models [\alpha]\neg Done(x, a) \rightarrow \neg Agree(\Gamma, a_x)$, and this is true iff $\exists \alpha \in \Pi$ such that if $\mathcal{M}, \omega \models Agree(\Gamma, a_x)$ then $\mathcal{M}, \omega \models \langle \alpha \rangle Done(x, a)$. And this is immediate from C4.
5. The proof of axioms 5, 6, 7, and 8, are straight forward from C8, C9, C10 and C11 respectively. □