A Logical Analysis of the Relationship between Commitment and Obligation *

Churn-Jung Liau
Institute of Information Science
Academia Sinica, Taipei, Taiwan
E-mail: liaucj@iis.sinica.edu.tw

Abstract. In this paper, we analyze the relationship between commitment and obligation from a logical viewpoint. The principle of commitment implying obligation is proven in a specific logic of action preference which is a generalization of Meyer's dynamic deontic logic. In the proposed formalism, an agent's commitment to goals is considered as a special kind of actions which can change her deontic preference and her obligation to doing some action is based on the preference and the effects of the action. In the logic, it is shown that an agent has the obligation to doing any action which is necessary for achieving as many committed goals as possible. The semantics of our logic is based on the possible world models for dynamic logic of actions. A binary preference relation between possible worlds is associated with the model. Then the preference between actions are determined by comparing that of their consequences. According to the semantics, while the preference will influence the agent's choice of actions, commitment is a kind of actions that will change the agent's preference. Thus we can show how obligations arise from commitments via updating of deontic preference. The integrated semantics make it possible to express and reasoning about the mutual relationship among these mental attitudes in a common logic.

Keywords: Deontic logic, dynamic logic, logic of commitment, logic of preference, agent-oriented programming, logic in AI.

1. Introduction

Deontic logic is the logic for reasoning about norms. Deontic reasoning has been extensively exploited in ethics and legal philosophy since the ancient times. However, the first modern formal system for deontic logic is not established until the fifties(von Wright, 1951). Though the system is influential on the later work, there arise many paradoxes when it is applied to practical deontic reasoning, so alternative systems have also been proposed since then for the resolution of paradoxes(Åqvist, 1984). Among them, Meyer's approach is one of the most interesting(Meyer, 1988). His system is based on the reduction to dynamic logic(Harel, 1984), so both actions and propositions can be represented and the distinction of "ought-to-do" and "ought-to-be" can be made clearly. This

* A preliminary version of the paper has appeared as (Liau, 1998).

in turn clarifies much confusion caused by the inappropriate translation of practical deontic reasoning to formal systems.

While the different formal systems are mainly the consequence of fundamental studies of deontic reasoning, the applications of deontic logics to computer science and artificial intelligence has received more and more attention recently(Meyer and Wieringa, 1993). The applications include automated legal reasoning, electronic commerce, system specification, and so on. Since Meyer's system is strongly based on dynamic logic and the latter is the logic for reasoning about computer programs, it is in particular suitable for the potential application.

In addition to the reduction to dynamic logic, another important feature of Meyer's logic is the use of a special propositional atom V, meaning the violation of law (or something like sanction, punishment, etc.). This special atom is originally introduced by Anderson(Anderson, 1958) for reducing deontic logic to alethic modal logic. By using the special atom, an action is forbidden if the execution of it will necessarily lead to states in which V holds, and it is permitted if not forbidden. Moreover, an action is obligatory if failing to executing it will result in violation of law.

Though Meyer's logic is successful in reasoning about normative actions, it is inadequate in the representation of action preference. However, the norms are usually relative and conflict with other ones, so we may have to make some decision choices between conflicting actions. For example, the violation of constitution is considered more serious than that of regulation laws, so we will try to obey the former instead of the latter provided that it is impossible to enforce both in the same time. Under the situations, we will need the capability of reasoning about action preference. In (Liau, 1997), it is shown that Meyer's logic can be generalized to a logic of action preference(LAP) by using possibility theory constructs. Thus we can represent and reason about an agent's choice of action in LAP.

Recently, the notion of agent has served as a metaphor for computer systems and resulted in the development of agent-oriented programming(Shoham, 1993). The philosophical and logical analysis of the properties of natural agents, especially human beings, and their mutual relationships can provide insight for the design and implementation of artificial agents. In particular, the mental attitudes of agents including informational ones (e.g. knowledge and belief), motivational ones (e.g. commitment and intention), and social ones (e.g. obligation and permission) play important roles in such kind of systems. The analysis of these attitudes has been the traditional concern of philosophical logic, such as epistemic logic, doxastic logic(Hintikka, 1962), and deontic logic(Åqvist, 1984). The mutual relationships of these attitudes can also be analyzed by integrating these individual logics. For example, the relationship of knowledge and belief is considered in a logic where the epistemic operator satisfies the properties of S5 and the doxastic operator satisfies the axioms of KD45 in (Halpern, 1996b). Following the same spirit, we analyze the relationship between commitment and obligation from a logical viewpoint. The principle of commitment implying obligation is proven in an extension of $LAP2$ that is a variant of LAP. In the proposed formalism, an agent's commitment to goals is considered as a special kind of actions which can change her deontic preference and her obligation to doing some action is based on the preference and the effects of the action. In the logic, it is shown that an agent has the obligation to doing any action which is necessary for achieving as many committed goals as possible. The semantics of our logic is based on the possible world models for dynamic logic of actions. A binary preference relation between possible worlds is associated with the model. Then the preference between actions are determined by comparing that of their consequences. According to the semantics, while the preference will influence the agent's choice of actions, commitment is a kind of actions that will change the agent's preference. Thus we can show how obligations arise from commitments via updating of deontic preference. The integrated semantics make it possible to express and reasoning about the mutual relationship among these mental attitudes in a common logic.

Though commitment is usually seen as a directional act between two agents, for simplicity, we will only consider it from the viewpoint of the agent who makes the commitment. Thus we can say that an agent commits to some goals but in the same time omit the counterparty towards whom the agent is making the commitment. This kind of commitment is sometimes called internal commitment or goal commitment. In this paper, when we refer to commitment, we always mean internal commitment.

The rest of the paper is organized as follows. First, Meyer's deontic logic is reviewed in the next section. Then a logic of action preference generalizing LAP will be presented. Its relationship with deontic logics will also be discussed. In section 4, the notion of commitment is incorporated into our logic. In section 5, some related literatures are discussed. Finally, we give the conclusion and some perspective of the works.

2. Review of Dynamic Deontic Logic

The system of Meyer's logic is called PD_eL. The elementary symbols of the PD_eL language consist of

1. A set of propositional letters, $PV = \{p, q, r, \ldots\}$ and a special propositional letter V not in PV, and
2. a set of atomic actions, $A = \{a, b, c, \ldots\}$ and two distinguished action symbols $\emptyset$ and $\mathbf{u}$.

The set of well-formed formulas(Φ_1) and the set of action expressions(Π_1) are defined inductively in the following way.

1. Φ_1 is the smallest set such that
   - $PV \cup \{V\} \subseteq \Phi_1$, and
   - if $\varphi, \psi \in \Phi_1$ and $\alpha \in \Pi_1$, then $\neg\varphi, \varphi \vee \psi, [\alpha]\varphi \in \Phi_1$.
2. Π_1 is the smallest set such that
   - $A \cup \{\emptyset, \mathbf{u}\} \subseteq \Pi_1$, and
   - if $\alpha, \beta \in \Pi_1$ and $\varphi \in \Phi_1$, then $\alpha; \beta, \alpha \cup \beta, \alpha \& \beta, \overline{\alpha}, \varphi \to \alpha/\beta \in \Pi_1$.

Let $\Phi_0 = \{\varphi \in \Phi_1 \mid V \text{ does not occur in } \varphi\}$ denote the set of ordinary dynamic logic formulas, and $\Pi_0 = \{\alpha \in \Pi_1 \mid V \text{ does not occur in } \alpha\}$ be the set of pure action terms. The wff $\neg[\alpha]\neg\varphi$ is abbreviated as $\langle\alpha\rangle\varphi$ and the other classical connectives($\top, \bot, \wedge, \supset, \equiv$) are defined as usual. The wff $[\alpha]\varphi$ means that if action α is done, φ will hold. The action expressions $\alpha; \beta, \alpha \cup \beta, \alpha \& \beta$ denote the sequential composition, nondeterministic choice, and simultaneous execution of α and β respectively, whereas $\overline{\alpha}$ means the non-execution of α and $\varphi \to \alpha/\beta$ denotes that if φ holds then execute α else execute β.

Note that the language of PD_eL is not the traditional one for dynamic logic. First, the Kleene star α^* is not in Π_1. This is because in the deontic reasoning domain, the repetition of some actions is not so usual as in computer programs. On the other hand, the simultaneous execution and non-execution of actions are nonstandard in dynamic logic. In fact, the two constructs significantly complicate the formal semantics of PD_eL. In particular, Meyer gives a two-level construction for the semantics of the action expressions. At the first level, each action expression is associated with a collection of so-called *synchronicity* sets(s-sets)[1]. Then at the second level, each s-set is mapped to a binary

[1] An s-set is a set of primitive actions and denotes the simultaneous execution of these actions.

transition relation on the set of possible worlds. The composition of these two mappings is roughly equivalent to the ordinary Kripke semantics for dynamic logic in which each action expression is directly mapped to a state transition relation. Since the precise description of Meyer's semantics unnecessarily complicates the matters, for the purpose of our current need, we will only present the denotation of an action expression as a state transition relation.

A Kripke model for PD_eL is a quadruple $\langle W, \models_1, \|\cdot\|_1, opt\rangle$, where W is a set of possible worlds and opt is a nonempty subset of W, meaning the best elements of W, whereas $\models_1 \subseteq W \times \Phi_1$ and $\|\cdot\|_1 : \Pi_1 \to \mathcal{P}(W \times W)$ define the truth relation and the action denotation function respectively. A Kripke mode for PD_eL is also called a *deontic model.* It is required that $\models_1$ and $\|\cdot\|_1$ must satisfy the following constraints.

– For all $w \in W$, $\varphi, \psi \in \Phi_1$, and $\alpha \in \Pi_1$,

  ($\models_1$0) $w \models_1 V \Leftrightarrow w \notin opt$,

  ($\models_1$1) $w \models_1 \varphi \vee \psi \Leftrightarrow w \models_1 \varphi$ or $w \models_1 \psi$,

  ($\models_1$2) $w \models_1 \neg\varphi \Leftrightarrow w \not\models_1 \varphi$,

  ($\models_1$3) $w \models_1 [\alpha]\varphi \Leftrightarrow \forall u \in \|\alpha\|_1(w), u \models_1 \varphi$, where $\|\alpha\|_1(w) = \{u \mid (w,u) \in \|\alpha\|_1\}$.

– For all $w \in W$, $\alpha, \beta \in \Pi_1$, and $\varphi \in \Phi_1$,

  ($\|\cdot\|_1$1): $\|\emptyset\|_1 = \emptyset$, $\|\mathbf{u}\|_1 = W \times W$.

  ($\|\cdot\|_1$2): $\|\alpha \cup \beta\|_1 = \|\alpha\|_1 \cup \|\beta\|_1$.

  ($\|\cdot\|_1$3): $\|\alpha;\beta\|_1 = \|\alpha\|_1 \circ \|\beta\|_1$.[2]

  ($\|\cdot\|_1$4): $\|\alpha\&\beta\|_1 \subseteq \|\alpha\|_1 \cap \|\beta\|_1$.

  ($\|\cdot\|_1$5): (negated actions)

  1. $\|\overline{\overline{\alpha}}\|_1 = \|\alpha\|_1$
  2. $\|\alpha \cup \overline{\alpha}\|_1 = \|\mathbf{u}\|_1$, $\|\alpha\&\overline{\alpha}\|_1 = \|\emptyset\|_1$.
  3. $\|\overline{\alpha \cup \beta}\|_1 = \|\overline{\alpha}\&\overline{\beta}\|_1$
  4. $\|\overline{\alpha\&\beta}\|_1 = \|\overline{\alpha} \cup \overline{\beta}\|_1$
  5. $\|\overline{\alpha;\beta}\|_1 = \|\overline{\alpha} \cup (\alpha;\overline{\beta})\|_1$
  6. $\|\overline{\varphi \to \alpha/\beta}\|_1 = \|\varphi \to \overline{\alpha}/\overline{\beta}\|_1$

  ($\|\cdot\|_1$6): (conditional actions) $\|\varphi \to \alpha/\beta\|_1(w) = \begin{cases} \|\alpha\|_1(w), & \text{if } w \models_1 \varphi, \\ \|\beta\|_1(w), & \text{if } w \not\models_1 \varphi. \end{cases}$

[2] i.e. the relational composition of $\|\alpha\|_1$ and $\|\beta\|_1$.

As usual, the denotation of an action is just a state transition relation. However, unlike standard compound actions in traditional dynamic logics, the denotations of $\overline{\alpha}$ and $\alpha\&\beta$ are not functionally determined by those of their component actions. This makes it impossible to define the denotation function only for primitive actions and then extend it to all actions. The reason that equality does not hold for ($\llbracket\cdot\rrbracket_1 4$) is due to the open specification of the action denotation in Meyer's semantics. As mentioned above, there are two mappings f_1 and f_2 which respectively assign a collection of s-sets to an action expression and a state transition relation to a collection of s-sets, and $\llbracket\cdot\rrbracket_1$ is equal to $f_2 \circ f_1$. Indeed, under some moderate condition, we have $f_1(\alpha\&\beta) = f_1(\alpha) \cap f_1(\beta)$. Moreover, let T_1, T_2 denote collections of s-sets, then we have $f_2(T_1) \subseteq f_2(T_2)$ if $T_1 \subseteq T_2$. Thus, the inequality of ($\llbracket\cdot\rrbracket_1 4$) holds. However, we don't have $f_2(T_1 \cap T_2) = f_2(T_1) \cap f_2(T_2)$ since different s-sets may have overlapping state transition relations. Therefore, when $T_1 \cap T_2 = \emptyset$ but there are some s-sets in T_1 and T_2 which are mapped to overlapped relations, then $f_2(T_1 \cap T_2)$ will be empty whereas $f_2(T_1) \cap f_2(T_2)$ be not. For example, assume under a state w, the door is closed, let α denote the action "close the door", then since $\alpha\&\overline{\alpha}$ is a failing action, $\llbracket\alpha\&\overline{\alpha}\rrbracket_1 = \emptyset$, but w is obviously in both sets of alternatives after the execution of α and $\overline{\alpha}$, i.e., $(w, w) \in \llbracket\alpha\rrbracket_1$ and $(w, w) \in \llbracket\overline{\alpha}\rrbracket_1$.

The deontic wffs are defined as abbreviations,

$$F\alpha = [\alpha]V,$$

$$P\alpha = \neg F\alpha = \langle\alpha\rangle\neg V,$$

$$O\alpha = F\overline{\alpha} = [\overline{\alpha}]V,$$

for all $\alpha \in \Pi_1$.

If $M = \langle W, \models_1, \llbracket\cdot\rrbracket_1, opt\rangle$ is a deontic model, then we write $M \models_1 \varphi$ if for all $w \in W$, $w \models_1 \varphi$. Let S be a subset of Φ_1 and $\varphi \in \Phi_1$, then $S \models_1 \varphi$ if for all models M and w in M, $w \models_1 \psi$ for all $\psi \in S$ implies $w \models_1 \varphi$. If $S = \emptyset$ (resp. $S = \{\psi\}$), then it is also written as $\models_1 \varphi$ (resp. $\psi \models_1 \varphi$).

3. A Logic of Action Preference

In this section, we introduce a logic of action preference, called $LAP2$, to distinguish it from the original LAP.

The language of $LAP2$ is an extension of PD_eL with two additional binary connectives $\succ$ and $\succeq$. The formation rules for well-formed formulas (Φ_2) and action expressions (Π_2) of $LAP2$ are as follows:

1. Φ_2 is the smallest set such that
   - $\Phi_1 \subseteq \Phi_2$,
   - if $\varphi, \psi \in \Phi_2$ and $\alpha \in \Pi_2$, then $\neg\varphi, \varphi \vee \psi, [\alpha]\varphi \in \Phi_2$, and
   - If α and $\beta \in \Pi_2$, then $\alpha \succ \beta, \alpha \succeq \beta \in \Phi_2$.
2. Π_2 is the smallest set such that
   - $\Pi_1 \subseteq \Pi_2$, and
   - if $\alpha, \beta \in \Pi_2$ and $\varphi \in \Phi_2$, then $\alpha;\beta, \alpha \cup \beta, \alpha \& \beta, \overline{\alpha}, \varphi \rightarrow \alpha/\beta \in \Pi_2$.

The wff $\alpha \succ \beta$(resp. $\alpha \succeq \beta$) is also written as $\beta \prec \alpha$ (resp. $\beta \preceq \alpha$). The wff $\alpha \succ \beta$ means that α is strictly preferred to β, whereas $\alpha \succeq \beta$ denote that α is preferred to β. Formally, a *preferential model* for $LAP2$ is a quadruple $\langle W, \models_2, \llbracket \cdot \rrbracket_2, \geq \rangle$, where W is a set of possible worlds and $\geq$ is a reflexive and transitive binary relation on W. In the preferential models, let us define $Opt(W, \geq) = \{w \in W \mid \not\exists u \in W(u > w)\}$, i.e., the set of optimal worlds. Furthermore, we assume in any preferential model, the set of optimal worlds is nonempty (i.e. the set W has at least a $\geq$-maximal element). Let $u > v$ denote $u \geq v$ and $v \not\geq u$ for $u, v \in W$. For $X, Y \subseteq W$, we also write $X \geq Y$(resp. $X > Y$) if there exists $u \in X$ such that for all $v \in Y$, $u \geq v$(resp. $u > v$). Note that the ordering between subsets of possible worlds is induced from that between individual worlds by some optimistic criterion. This is not the only way to define the set ordering. For other possible alternatives and related issues, see section 5.

Then $\llbracket \cdot \rrbracket_2 : \Pi_2 \rightarrow \mathcal{P}(W \times W)$ must still satisfy ($\llbracket \cdot \rrbracket_1 1$)–($\llbracket \cdot \rrbracket_1 6$), whereas $\models_2 \subseteq W \times \Phi_2$ now have to satisfy ($\models_1 1$)–($\models_1 3$) and the following constraints:

($\models_2 0$') $w \models_2 V \Leftrightarrow \exists u \in W(u > w)$,

($\models_2 4$) $w \models_2 \alpha \succeq \beta$ iff $\llbracket \alpha \rrbracket_2(w) \geq \llbracket \beta \rrbracket_2(w)$,

($\models_2 5$) $w \models_2 \alpha \succ \beta$ iff $\llbracket \alpha \rrbracket_2(w) > \llbracket \beta \rrbracket_2(w)$,

The definition of $S \models_2 \varphi$ is analogous to that for $\models_1$ when $S \cup \{\varphi\} \subseteq \Phi_2$. $LAP2$ has some interesting properties similar to LAP. These properties all easily follow from the semantics.

PROPOSITION 1.

1. $\models_2 (\alpha_1 \cup \alpha_2 \succ \beta) \equiv (\alpha_1 \succ \beta) \vee (\alpha_2 \succ \beta)$

2. $\models_2 (\beta \succ \alpha_1 \cup \alpha_2) \supset (\beta \succ \alpha_1) \wedge (\beta \succ \alpha_2)$

3. $\models_2 (\alpha_1 \& \alpha_2 \succ \beta) \supset (\alpha_1 \succ \beta) \wedge (\alpha_2 \succ \beta)$

4. $\models_2 \neg(\alpha \& \beta \succ \beta)$

5. $\models_2 \neg(\alpha \succ \alpha \cup \beta)$

6. $\models_2 (\alpha;\beta_1 \succ \alpha;\beta_2) \supset \langle\alpha\rangle(\beta_1 \succ \beta_2)$

7. $\models_2 (\varphi \to \alpha_1/\alpha_2 \succ \beta) \equiv (\varphi \supset \alpha_1 \succ \beta) \wedge (\neg\varphi \supset \alpha_2 \succ \beta)$

Furthermore, the same results (except 4. and 5.) hold by replacing $\succ$ *with* $\succeq$*.*

The action preference wffs can be combined with the conditional action wffs to denote a decision choice action. More specifically, define

$$\alpha \oplus \beta = (\alpha \succ \beta) \to \alpha/(\alpha \prec \beta \to \beta/(\alpha \cup \beta)).$$

Then doing $\alpha \oplus \beta$ will mean doing α or β selectively according to the strict preference relation $\succ$. The same definition can be carried out for $\succeq$.

It can easily be seen that a preferential model is a generalization of deontic model. Just like LAP, $LAP2$ is also a conservative extension of PD_eL.

PROPOSITION 2. *If* $S \cup \{\varphi\} \subseteq \Phi_1$*, then* $S \models_1 \varphi$ *iff* $S \models_2 \varphi$*.*

Proof: Since a deontic model can be seen as a special case of preferential model, it can be easily shown that if $S \models_2 \varphi$ then $S \models_1 \varphi$ by the definition of logical consequence. On the other hand, if $S \not\models_2 \varphi$, then there exist a preferential model $\langle W, \models_2, \llbracket\cdot\rrbracket_2, \geq\rangle$ and a $w \in W$ such that for all $\psi \in S$, $w \models_2 \psi$ but $w \not\models_2 \varphi$. Let $M = \langle W, \models_1, \llbracket\cdot\rrbracket_1, opt\rangle$, where $\llbracket\cdot\rrbracket_1$ is a restriction of $\llbracket\cdot\rrbracket_2$ to Π_1, $opt = Opt(W, \geq) = \{w \in W \mid \not\exists u \in W(u > w)\}$, and $\models_1$ is the restriction of $\models_2$ to the set $W \times \Phi_1$, then it can be verified that M is indeed a deontic model by induction on the complexity of formulas and action expressions. By the definition, M is a counter-model of $S \models_1 \varphi$. ■

This result shows that $LAP2$ is at least as expressive as PD_eL in deontic reasoning. Furthermore, $LAP2$ facilitates the expression of comparative preference reasoning directly in its language.

In the preferential models, the relation $\geq$ is a preorder. A preorder is also called a ranked order (total preorder) if for all $u, v \in W$, we have either $u \geq v$ or $v \geq u$. In the LAP model, since the preferential relation is represented by a possibility distribution(Zadeh, 1978), it is really a ranked order with upper and lower bounds. Though ranked orders may produce more interesting properties, the commitment actions we will discuss in the next section may sometimes induce incomparability between worlds, so the generalization from LAP to $LAP2$ models is indeed necessary for the logic of commitment.

4. A Logic of Commitment

Our logic of commitment is called LC. The language of LC is an extension of $LAP2$ with two additional action-forming operators ! and @. The formation rules for well-formed formulas (Φ_3) and action expressions (Π_3) of LC are as follows:

1. Φ_3 is the smallest set such that
   - $\Phi_2 \subseteq \Phi_3$,
   - if $\varphi, \psi \in \Phi_3$ and $\alpha \in \Pi_3$, then $\neg\varphi, \varphi \vee \psi, [\alpha]\varphi \in \Phi_3$, and
   - If α and $\beta \in \Pi_2$, then $\alpha \succ \beta, \alpha \succeq \beta \in \Phi_3$.
2. Π_3 is the smallest set such that
   - $\Pi_2 \subseteq \Pi_3$,
   - if $\varphi \in \Phi_0$, then $!\varphi, @\varphi \in \Pi_3$, and
   - if $\alpha, \beta \in \Pi_3$, then $\alpha;\beta, \alpha \cup \beta \in \Pi_3$.

The actions of types $!\varphi$ and $@\varphi$ are called commitment and fulfillment actions respectively. The intended meaning of $!\varphi$ is the agent commits to the goal φ, whereas $@\varphi$ means that the commitment to φ is deleted since φ has been achieved.

Note that we do not allow the comparison between commitments since they are not state transition actions. Also, only commitments to Φ_0 goals are legal actions under our syntax. In other words, we will not allow the commitments to an action preference statement. Furthermore, though commitments to a formula $\neg\varphi$ is allowed, the non-commitment to φ (i.e. $\overline{!\varphi}$) is not a legal action expression. We do not need the simultaneous commitments $!\varphi\&!\psi$ either since it is equivalent to $!(\varphi \wedge \psi)$.

Before giving the formal semantics, we need some notations. As usual, let Φ_0^* denote the set of all finite sequences of wffs in Φ_0. We will use $\sigma, \tau, \ldots$ to denote the elements of Φ_0^*. The symbol λ will denote the empty sequence. Let $\sigma = \varphi_1 \cdot \varphi_2 \cdots \varphi_n$ and $\varphi \in \Phi_0$, then $\sigma\backslash\varphi$ is defined as $\varphi'_1 \cdot \varphi'_2 \cdots \varphi'_k$, where φ'_i is the ith element of σ such that $\varphi \not\models_1 \varphi'_i$. In other words, $\sigma\backslash\varphi$ is the result of removing all formulas implied by φ from σ. Let W be a set of possible worlds and $U \subseteq W \times \Phi_0^*$, then $P_W(U)$

is the projection of U on W, defined by $P_W(U) = \{w \mid (w,\sigma) \in U\}$. Let $\langle W, \models_1, [\![\cdot]\!]_1, opt\rangle$ be a deontic model, then for a preorder $\geq$ on W, and a sequence $\sigma \in \Phi_0^*$, the preorder induced from $\geq$ and σ by *fair* strategy, denoted by $\geq_\sigma^{fair}$, is defined inductively as follows:

1. $\geq_\lambda^{fair} = \geq$,
2. $\geq_{\sigma\cdot\varphi}^{fair} = \{(u,v) \mid u \geq_\sigma^{fair} v \wedge (v \models_1 \varphi \Rightarrow u \models_1 \varphi)\}$

Analogously, the preorder induced by *first-come-first-serve* strategy, denoted by $\geq_\sigma^{fcfs}$, is defined inductively as follows:

1. $\geq_\lambda^{fcfs} = \geq$,
2. $\geq_{\sigma\cdot\varphi}^{fcfs} = \{(u,v) \mid u >_\sigma^{fcfs} v \text{ or } u \geq_\sigma^{fcfs} v \wedge v \geq_\sigma^{fcfs} u \wedge (v \models_1 \varphi \Rightarrow u \models_1 \varphi)\}$

In what follows, we will write $\geq_\sigma$ for $\geq_\sigma^{fair}$ or $\geq_\sigma^{fcfs}$ and distinguish them only when it is necessary.

The intended use of $\geq_\sigma$ is to update the preference ordering by a sequence of unfulfilled goals. In the preceding sections, a state is just a possible world, so the state transition relation corresponding to an action just changes a possible world into another one among the possible alternatives when the action is performed. However, because commitment and fulfillment are not actions in a traditional sense, we will extend a state to a pair consisting of a possible world and a sequence of unfulfilled goals. While the execution of ordinary actions change the possible world component of the extended state, commitment and fulfillment may add or delete formulas to the sequence of unfulfilled goals. Then the preference ordering under an extended state is determined by the initial one $\geq$ and the sequence of goals σ. The fair and fcfs strategies represent different ways by which the preference ordering can be determined.

To illustrate the difference of the two strategies, we consider a sequence consisting of only two goals φ and ψ and assume the initial preference ordering $\geq = W \times W$. The evolution of the preference ordering after the commitment of φ and ψ is shown in figure 1, where (a→b→c) and (a→b→d) represent respectively the preference updating by fair and fcfs strategies. The circles represent nonempty sets of possible worlds satisfying the formulas contained in them. The arrows represent preferences for all worlds between two circles connected by them. Thus a unidirectional arrow represents that the preferences are strict and the bidirectional ones mean that the two classes of possible worlds are in fact in the same rank of the preference ordering. Since our initial ordering is $W \times W$, all worlds are equivalent under the ordering. This is shown in figure 1(a), where there exists a bidirectional arrow between any two circles. Then, after the commitment of a goal φ, the worlds satisfying φ are preferred to those not. The effects of fair and fcfs strategies are of no difference when only one goal is committed. This is shown in figure 1(b) by a unidirectional arrow from the circle labelling $\neg\varphi, \neg\psi$ to that labelling φ, ψ. For the sake of clarity, we intentionally omit the arrows that can be derived by the transitivity of the ordering from figure 1(b), (c), (d). The things get different when another goal ψ is also committed. For the fair strategy, both unfulfilled goals are considered equally important, so essentially a world is preferred to another one, if the goals satisfied in the latter are contained in those satisfied in the former. The resultant ordering is shown in figure 1(c). The direction of the arrows is consistent with the subset relation between sets of positive formulas in the circles. In particular, note that the two circles labelling $\varphi, \neg\psi$ and $\neg\varphi, \psi$ are not comparable. On the other hand, for the fcfs strategy, the things are quite different. Since φ is committed before ψ, it is considered more important under the strategy. Thus a world satisfying $\varphi, \neg\psi$ is preferred to those satisfying $\neg\varphi, \psi$. This is shown in figure 1(d), where the ordering is linear between the circles.

Now, we are ready to define the semantics of *LC*. A *commitment model* is a sextuple $\langle W, \models_1, \llbracket \cdot \rrbracket_1, \models_3, \llbracket \cdot \rrbracket_3, \geq \rangle$, where

1. W is a set of possible worlds and $\geq$ is a preorder on W,

2. $\models_1 \subseteq W \times \Phi_1$ satisfies ($\models_1$ 1)–($\models_1$ 3),

3. $\llbracket \cdot \rrbracket_1 : \Pi_1 \to \mathcal{P}(W \times W)$ satisfies ($\llbracket \cdot \rrbracket_1 1$)–($\llbracket \cdot \rrbracket_1 6$),

4. $\models_3 \subseteq (W \times \Phi_0^*) \times \Phi_3$ must satisfy

   (coherence): if $\varphi \in \Phi_0$, then $w, \sigma \models_3 \varphi \Leftrightarrow w \models_1 \varphi$,

   ($\models_3 0$) $w, \sigma \models_3 V \Leftrightarrow \exists u \in W (u >_\sigma w)$,

   ($\models_3 1$) $w, \sigma \models_3 \varphi \vee \psi \Leftrightarrow w, \sigma \models_3 \varphi$ or $w, \sigma \models_3 \psi$,

   ($\models_3 2$) $w, \sigma \models_3 \neg\varphi \Leftrightarrow w, \sigma \not\models_3 \varphi$,

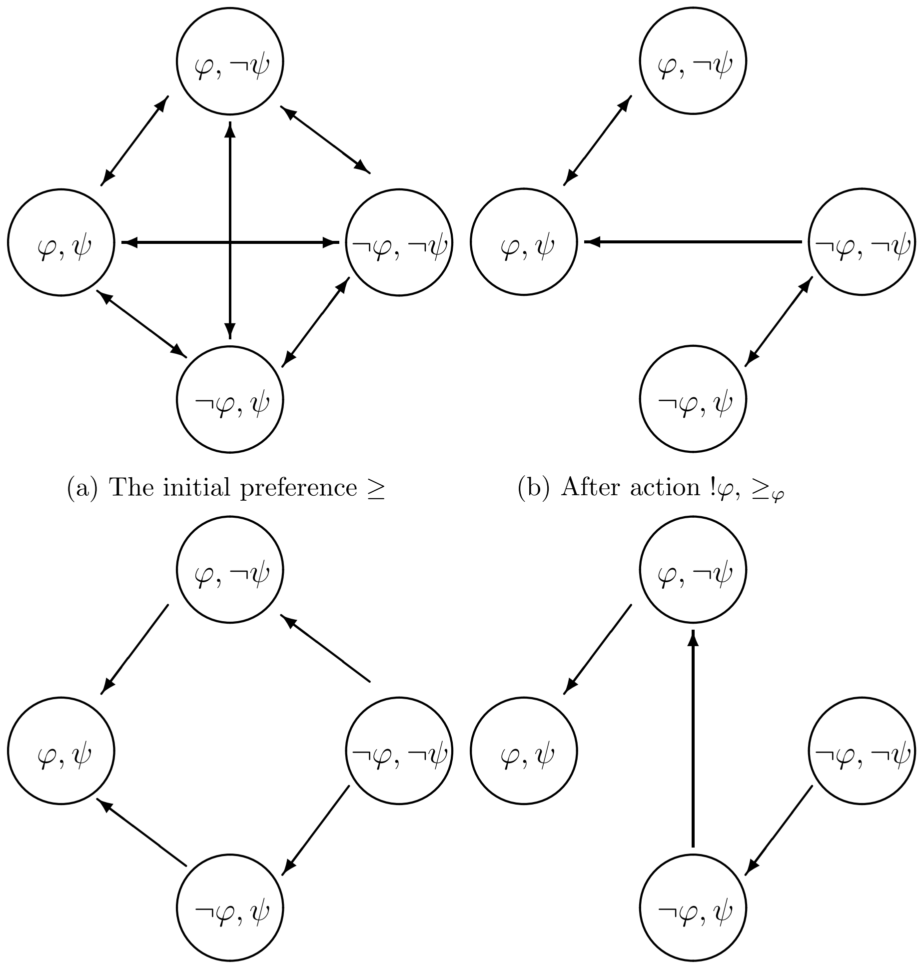
   ($\models_3 3$) $w, \sigma \models_3 [\alpha]\varphi \Leftrightarrow \forall (u, \tau) \in \llbracket\alpha\rrbracket_3(w, \sigma), (u, \tau \models_3 \varphi)$, where $\llbracket\alpha\rrbracket_3(w, \sigma) = \{(u, \tau) \mid (((w, \sigma), (u, \tau)) \in \llbracket\alpha\rrbracket_3\}$,

   ($\models_3 4$) $w, \sigma \models_3 \alpha \succeq \beta$ iff $P_W(\llbracket\alpha\rrbracket_3(w, \sigma)) \geq_\sigma P_W(\llbracket\beta\rrbracket_3(w, \sigma))$, and

   ($\models_3 5$) $w, \sigma \models_3 \alpha \succ \beta$ iff $P_W(\llbracket\alpha\rrbracket_3(w, \sigma)) >_\sigma P_W(\llbracket\beta\rrbracket_3(w, \sigma))$, and

5. $\llbracket \cdot \rrbracket_3 : \Pi_3 \to \mathcal{P}((W \times \Phi_0^*) \times (W \times \Phi_0^*))$ must satisfy

   (coherence): if $\alpha \in \Pi_0$, then $\llbracket\alpha\rrbracket_3(w, \sigma) = \{(u, \sigma) \mid u \in \llbracket\alpha\rrbracket_1(w)\}$,



(a) The initial preference $\geq$ (b) After action $!\varphi$, $\geq_\varphi$

(c) After action $!\varphi; !\psi$, $\geq^{fair}_{\varphi\cdot\psi}$ (d) After action $!\varphi; !\psi$, $\geq^{fcfs}_{\varphi\cdot\psi}$

Figure 1. The comparison of fair and fcfs strategies

(commitment): $[\![!\varphi]\!]_3(w, \sigma) = \{(w, \sigma \cdot \varphi)\}$,

(fulfillment): $[\![@\varphi]\!]_3(w, \sigma) = \{(u, \sigma \backslash \varphi) \mid u \models_1 \varphi\}$,

and ($[\![\cdot]\!]_1$2)–($[\![\cdot]\!]_1$6) with $[\![\cdot]\!]_1$ being replaced by $[\![\cdot]\!]_3$(in ($[\![\cdot]\!]_1$6), w is also replaced by w, σ).

For all $S \cup \{\varphi\} \subseteq \Phi_3$, $S \models_3 \varphi$ if for all commitment model $M = \langle W, \models_1, [\![\cdot]\!]_1, \models_3, [\![\cdot]\!]_3, \geq\rangle$, w in W, and $\sigma \in \Phi_0^*$, $w, \sigma \models_3 \psi$ for all $\psi \in S$ implies $w, \sigma \models_3 \varphi$. The given preorder $\geq$ is the initial preference of the agent and the commitment action $!\varphi$ updates it according to the fair or fcfs strategy, so we in fact define two kinds of commitment models. To distinguish between them, sometimes we will add the superscript $fair$ or $fcfs$ to the truth relation $\models_3$ and action denotation function $\llbracket \cdot \rrbracket_3$. Note that $\models_3$ and $\llbracket \cdot \rrbracket_3$ only agree with $\models_1$ and $\llbracket \cdot \rrbracket_1$ respectively in the set Φ_0 and Π_0 since the evaluation of the symbol V relies on the preference relation induced by the commitment actions. Thus, the wffs in $\Phi_1 \backslash \Phi_0$ and the action expressions in $\Pi_1 \backslash \Pi_0$ will be taken care of by $\models_3$ and $\llbracket \cdot \rrbracket_3$ directly.

It can be seen that for all $\alpha \in \Pi_2$, if $((w, \sigma), (u, \tau)) \in \llbracket \alpha \rrbracket_3$, then $\sigma = \tau$. In other words, the action expressions in Π_2 are purely state-transition ones, so the conditions ($\models_3$ 4) and ($\models_3$ 5) are well-motivated. On the other hand, the commitment actions keep the state unchanged though they indeed update the preference relation. The fulfillment actions may both change the states and update the preference relation. However, we do not specify how the world state is changed when the goal is fulfilled, so any (ordinary) action that can change the current state into one satisfying the goal may be taken. Thus, if the current extended state is (w, σ) and $@\varphi$ is performed, we will possibly reach $(u, \sigma \backslash \varphi)$ for any u which is accessible from w by some ordinary action and satisfies φ. Since there exists some universal action $\mathbf{u}$ in our logic, any world $u \in W$ is accessible from w by $\mathbf{u}$, so we do not have to add the restriction of accessibility to the semantics for $@\varphi$. However, the existence of $\mathbf{u}$ is somewhat over-idealized since in a real situation we may have no way to reach another given state. To accommodate this case, we may change the semantics of $\mathbf{u}$ into $\llbracket \mathbf{u} \rrbracket_1 \subseteq W \times W$ to mean the accessibility relation by any possible actions and the semantic restriction for fulfillment action is modified as $\llbracket @\varphi \rrbracket_3(w, \sigma) = \{(u, \sigma \backslash \varphi) \mid u \models_1 \varphi, (w, u) \in \llbracket \mathbf{u} \rrbracket_\mathbf{1}\}$.

Furthermore, an action $\alpha \in \Pi_2$ is called *commitment-invariant* if it does not have any subexpression of the form $\varphi \rightarrow \beta_1 / \beta_2$ with occurrence of V, $\succ$, or $\succeq$ in φ. It can be shown by induction that if α is a commitment-invariant action, then for any $\sigma, \tau \in \Phi_0^*$, $((w, \sigma), (u, \sigma)) \in \llbracket \alpha \rrbracket_3$ iff $((w, \tau), (u, \tau)) \in \llbracket \alpha \rrbracket_3$, i.e., the results of doing α only depend on the current world situation but not on the unfulfilled goals since it does not contain any conditional choice which has to be evaluated according to the current preference.

The following proposition illustrates how the commitment actions change an agent's preference between other actions.

PROPOSITION 3. *Let $\models_3$ denote either $\models_3^{fair}$ or $\models_3^{fcfs}$ and α, β be commitment-invariant actions, then*

1. $\models_3 (\alpha \succeq \beta) \wedge (([\alpha]\varphi \wedge [\beta]\varphi) \vee ([\alpha]\neg\varphi \wedge [\beta]\neg\varphi)) \supset [!\varphi](\alpha \succeq \beta)$

2. $\models_3 ((\alpha \succeq \beta) \wedge [\alpha]\varphi \wedge [\beta]\neg\varphi) \supset [!\varphi](\alpha \succ \beta)$

3. $\models_3 (\alpha \succ \beta) \wedge (([\alpha]\varphi \wedge [\beta]\varphi) \vee ([\alpha]\neg\varphi \wedge [\beta]\neg\varphi) \vee ([\alpha]\varphi \wedge [\beta]\neg\varphi)) \supset [!\varphi](\alpha \succ \beta)$

4. $\models_3^{fcfs} \alpha \succ \beta \supset [!\varphi]\alpha \succ \beta$

Proof: We will prove the first one and the remaining ones are analogous. Let $\langle W, \models_1, \llbracket \cdot \rrbracket_1, \models_3, \llbracket \cdot \rrbracket_3, \geq \rangle$ be a commitment model, $w \in W$ and $\sigma \in \Phi_0^*$. According to ($\models_3$ 4),

$$\begin{aligned} & w, \sigma \models_3 \alpha \succeq \beta \\ \Leftrightarrow\ & P_W(\llbracket \alpha \rrbracket_3(w,\sigma)) \geq_\sigma P_W(\llbracket \beta \rrbracket_3(w,\sigma)) \\ \Leftrightarrow\ & \exists u(((w,\sigma),(u,\sigma)) \in \llbracket \alpha \rrbracket_3 \wedge \forall v(((w,\sigma),(v,\sigma)) \in \llbracket \beta \rrbracket_3 \Rightarrow u \geq_\sigma v)), \end{aligned}$$

and since $\varphi \in \Phi_0^*$, according to coherence

$$\begin{aligned} & w, \sigma \models_3 (([\alpha]\varphi \wedge [\beta]\varphi) \vee ([\alpha]\neg\varphi \wedge [\beta]\neg\varphi)) \\ \Rightarrow\ & (u, \sigma \models_3 \varphi \wedge v, \sigma \models_3 \varphi) \vee (u, \sigma \not\models_3 \varphi \wedge v, \sigma \not\models_3 \varphi) \\ \Leftrightarrow\ & (u \models_1 \varphi \wedge v \models_1 \varphi) \vee (u \not\models_1 \varphi \wedge v \not\models_1 \varphi) \\ \Rightarrow\ & u \geq_{\sigma\cdot\varphi} v. \end{aligned}$$

Since α, β are commitment-invariant, we have $u \in P_W(\llbracket \alpha \rrbracket_3(w, \sigma \cdot \varphi))$ and $v \in P_W(\llbracket \beta \rrbracket_3(w, \sigma \cdot \varphi))$, so

$$\begin{aligned} & P_W(\llbracket \alpha \rrbracket_3(w, \sigma\cdot\varphi)) \geq_{\sigma\cdot\varphi} P_W(\llbracket \beta \rrbracket_3(w, \sigma\cdot\varphi)) \\ \Leftrightarrow\ & w, \sigma\cdot\varphi \models_3 \alpha \succeq \beta \\ \Leftrightarrow\ & w, \sigma \models_3 [!\varphi](\alpha \succeq \beta). \end{aligned}$$

■

Thus a commitment action may change a preference into a strict one either by fair or fcfs strategy. However, strict preference may be destroyed by fair strategy but not by the fcfs one.

On the other hand, the change of preference caused by fulfillment actions may be arbitrary since they may have global effect on the preference relation. In general, a fulfillment action will flatten some ordering by achieving some previously committed goals. To understand the total effect of a sequence of commitment and fulfillment actions, we will consider a special kind of models $M_0 = \langle W, \models_1, \llbracket \cdot \rrbracket_1, \models_3, \llbracket \cdot \rrbracket_3, \geq_0 \rangle$, where $\geq_0 = W \times W$. Such models are also called *neutral* models. Let $\mathcal{M}_0$ denote the class of all neutral models.

Given a $\sigma \in \Phi_0^*$, a maximal consistent subset of σ in M_0 is a maximal subset of σ, S, such that $|S|_{M_0} = \{w \mid w \models_1 S\}$ is nonempty. Let

$Con_{M_0}(\sigma)$ denote the set of all maximal consistent subsets of σ in M_0, then φ_σ is defined as

$$\varphi_\sigma = \bigvee\{\bigwedge S \mid S \in Con_{M_0}(\sigma)\}.$$

Let S be a subset of the sequence σ, then the characteristic formula of S with respect to σ is defined by

$$\chi(S,\sigma) = \bigwedge(S \cup \{\neg\varphi \mid \varphi \in \sigma, \varphi \notin S\}).$$

LEMMA 1. *If $M_0 \in \mathcal{M}_0$, w is a world of M_0, and $\sigma \in \Phi_0^*$, then*

1. *for all $\beta \in \Pi_2$, $w,\sigma \models_3^{fair} [\overline{\beta}]\neg\varphi_\sigma \equiv O\beta$,*
2. *for all $S_2 \subseteq S_1 \subseteq \sigma$ and $\alpha,\beta \in \Pi_2$, $w,\sigma \models_3^{fair} \langle\alpha\rangle\chi(S_1,\sigma) \wedge [\beta]\chi(S_2,\sigma) \supset \alpha \succeq \beta$*
3. *for all $S_2 \subset S_1 \subseteq \sigma$ and $\alpha,\beta \in \Pi_2$, $w,\sigma \models_3^{fair} \langle\alpha\rangle\chi(S_1,\sigma) \wedge [\beta]\chi(S_2,\sigma) \supset \alpha \succ \beta$*

Proof: First, for a possible world u in M_0, let us define $u(\sigma) = \{\varphi \in \sigma \mid u \models_1 \varphi\}$. Then, by induction on the length of σ, we can prove the following properties for any possible worlds u, v (we omit the superscript "fair" for simplicity):

(i) if $u \geq v$ and $v(\sigma) \subseteq u(\sigma)$, then $u \geq_\sigma v$

(ii) if $v(\sigma) \not\subseteq u(\sigma)$, then $u \not\geq_\sigma v$.

Thus, in the neutral model M_0, we have $u \geq_\sigma v$ iff $v(\sigma) \subseteq u(\sigma)$ since $u \geq v$ holds from the initial, so for any u, $u \models_1 \bigwedge S$ for some $S \in Con_{M_0}(\sigma)$ iff $\not\exists v(v >_\sigma u)$. Furthermore, since $S \subseteq \Phi_0$, $u \models_1 \bigwedge S$ iff $u,\sigma \models_3 \bigwedge S$.

We are now ready to prove the lemma.

1. For any possible world u, we have

$$\begin{aligned} & u,\sigma \models_3 \neg\varphi_\sigma \\ \Leftrightarrow\ & u,\sigma \not\models_3 \bigwedge S, \forall S \in Con_{M_0}(\sigma) \\ \Leftrightarrow\ & \exists v(v >_\sigma u) \\ \Leftrightarrow\ & u,\sigma \models_3 V. \end{aligned}$$

This implies $u,\sigma \models_3 \neg\varphi_\sigma \equiv V$ for any possible world u, so $w,\sigma \models_3 [\overline{\beta}]\neg\varphi_\sigma \equiv [\overline{\beta}]V$, i.e., $w,\sigma \models_3 [\overline{\beta}]\neg\varphi_\sigma \equiv O\beta$.

2. The proof is

$$\begin{array}{ll} & w,\sigma \models_3 \langle\alpha\rangle\chi(S_1,\sigma)\wedge[\beta]\chi(S_2,\sigma) \\ \Leftrightarrow & (\exists u\in P_W(\|\alpha\|_3(w,\sigma)), u,\sigma\models_3 \chi(S_1,\sigma))\wedge(\forall v\in P_W(\|\beta\|_3(w,\sigma)), v,\sigma\models_3\chi(S_2,\sigma)) \\ \Rightarrow & \exists u\in P_W(\|\alpha\|_3(w,\sigma))\forall v\in P_W(\|\beta\|_3(w,\sigma)).v(\sigma)\subseteq u(\sigma),\ \text{ since } S_2\subseteq S_1 \\ \Leftrightarrow & \exists u\in P_W(\|\alpha\|_3(w,\sigma))\forall v\in P_W(\|\beta\|_3(w,\sigma)).u\geq_\sigma v \\ \Leftrightarrow & P_W(\|\alpha\|_3(w,\sigma))\geq_\sigma P_W(\|\beta\|_3(w,\sigma)) \\ \Leftrightarrow & w,\sigma\models_3 \alpha\succeq\beta. \end{array}$$

3. The proof is analogous to 2. except that $\subseteq$ (resp. $\geq_\sigma$) is replaced by $\subset$ (resp. $>_\sigma$).

■

Note that in the proof 2. above, the derivation of $v(\sigma)\subseteq u(\sigma)$ from $S_2\subseteq S_1$ depends on the fact that u(resp. v) satisfies *exactly* all goals in S_1(resp. S_2) but not any one outside it, so the characteristic formula $\chi(S_1,\sigma)$(resp. $\chi(S_1,\sigma)$) cannot be replaced by $\bigwedge S_1$(resp. $\bigwedge S_2$). That is, the negative part in the definition of the characteristic formulas is indeed necessary.

Let $s=\alpha_1\cdot\alpha_2\cdots\alpha_n$ be a sequence of action expressions such that each α_i is in the forms of $!\varphi$ or $@\varphi$. Define α_s as the result of replacing all "$\cdot$" in s by the action composition operator ";". The *unfulfilled commitment sequence* corresponding to s, denoted by σ_s, is defined inductively by

1. $\sigma_\lambda=\lambda$,
2. $\sigma_{s\cdot !\varphi}=\sigma_s\cdot\varphi$,
3. $\sigma_{s\cdot @\varphi}=\sigma_s\backslash\varphi$.

PROPOSITION 4. *If M_0, w, and s are defined as above, then*

1. *for all $\beta\in\Pi_2$, $w,\lambda\models_3^{fair}[\alpha_s]([\overline{\beta}]\neg\varphi_{\sigma_s}\equiv O\beta)$.*
2. *for all $S_2\subseteq S_1\subseteq\sigma_s$ and $\alpha,\beta\in\Pi_2$, $w,\lambda\models_3^{fair}[\alpha_s](\langle\alpha\rangle\chi(S_1,\sigma_s)\wedge[\beta]\chi(S_2,\sigma_s)\supset\alpha\succeq\beta)$*
3. *for all $S_2\subset S_1\subseteq\sigma_s$ and $\alpha,\beta\in\Pi_2$, $w,\lambda\models_3^{fair}[\alpha_s](\langle\alpha\rangle\chi(S_1,\sigma_s)\wedge[\beta]\chi(S_2,\sigma_s)\supset\alpha\succ\beta)$*

Proof: The proof follows easily from the preceding lemma since $w,\lambda\models_3^{fair}[\alpha_s]\varphi$ iff $w,\sigma_s\models_3^{fair}\varphi$ for any $\varphi\in\Phi_3$ by definition of σ_s and the semantics of commitment and fulfillment actions. ■

The first part of the proposition says that if executing β is necessary to fulfill a maximal consistent subset of commitments, then it is obligatory to do β. This is the principle of commitment implying obligation. On the other hand, it also states that the only obligation is to achieve as many committed goals as possible. This is due to the neutrality of the model, so the only obligation is caused by the commitment action. If the initial model is not neutral, then we may have some obligations even before any goal commitment.

However, that an agent ought to do β does not mean that she will really do it. She may just not have the capability or opportunity to do it(for a formal model of capability and opportunity, see (van der Hoek et al., 1994)). In this case, the second and third parts of the proposition means that the agent should do her best to fulfill the commitments. In other words, we are modeling a sincere agent in the sense that she prefers the actions that can achieve as many commitments as possible.

The preceding proposition states the principle of commitment implying obligation under the fair strategy. For the fcfs strategy, an analogous result holds and it can also be proved in a similar way. Given $\sigma = \varphi_1 \cdot \varphi_2 \cdots \varphi_n \in \Phi_0^*$, define

1. $S_0 = \emptyset$,
2. $S_i = \begin{cases} S_{i-1} \cup \{\varphi_i\} & \text{if } |S_{i-1} \cup \{\varphi_i\}|_{M_0} \neq \emptyset, \\ S_{i-1} & \text{otherwise,} \end{cases}$ for $1 \leq i \leq n$.

Let $\psi_\sigma = \bigwedge S_n$, then we have

LEMMA 2. *If $M_0 \in \mathcal{M}_0$, w is a world of M_0, and $\sigma \in \Phi_0^*$, then $w, \sigma \models_3^{fcfs} [\overline{\beta}]\neg\psi_\sigma \supset O\beta$.*

PROPOSITION 5. *For all w in M_0 and $\beta \in \Pi_2$, $w, \lambda \models_3^{fcfs} [\alpha_s]([\overline{\beta}]\neg\psi_{\sigma_s} \supset O\beta)$.*

The second and third parts of Lemma 1 and Proposition 4 also hold for the fcfs case provided that the condition $S_2 \subset S_1$ (resp. $S_2 \subseteq S_1$) is replaced by $S_2 \sqsubset S_1$ (resp. $S_2 \sqsubseteq S_1$), where $S_2 \sqsubset S_1$ is defined with respect to a $\sigma = \varphi_1 \cdot \varphi_2 \cdots \varphi_n$ in the following way:

$$S_2 \sqsubset S_1 \Leftrightarrow \exists 1 \leq i \leq n(\varphi_i \in S_1 \wedge \varphi_i \notin S_2 \wedge \forall j < i(\varphi_j \in S_2 \Rightarrow \varphi_j \in S_1)),$$

and $S_2 \sqsubseteq S_1$ is $S_2 \sqsubset S_1$ or $S_1 = S_2$.

5. Related Works

The idea of considering actions updating the preference relation comes originally from the update semantics for default logic(Veltman, 1996). In the semantics, each subset of worlds (i.e. a proposition) is associated with a preference relation and a default of the form $\varphi \rightsquigarrow \psi$, where φ and ψ are classical logic formulas, will cause the modification of the preference associated with the subset of φ-worlds. Then if the agent knows that φ holds, she will prefer to believe ψ rather than $\neg\psi$. Thus, in update semantics for default logic, the default will change the cognitive preference of an agent, whereas in our case, the commitment actions will update the action preference and have influence on the agent's choice of actions. Furthermore, in the update semantics, only the fair strategy is considered while we consider both fair and fcfs ones.

The motivation for the logical semantics of commitment is to provide a formal account to some mental attitudes used in agent-oriented programming (Shoham, 1993). However, in (Shoham, 1993), commitment and obligation are not distinguished. Here, we model both of them in a common logic, and the first part of Proposition 4 in fact provide a formal relationship between them. That is, for a value-neutral agent, the unfulfilled commitment decides her obligations.

Some motivational attitudes, including preference, goal, and commitment, have been also formalized in (van Linder et al., 1995b). However, in (van Linder et al., 1995b), the preference is a unary operator applied to wffs, whereas the commitment is a operator applied to action expressions. On the contrary, in our logic, the preference is a binary connective between action expression and the commitment operator is applied to a wff. So the semantics of *LC* and that in (van Linder et al., 1995b) are quite different. Moreover, the action terms is more restrictive in (van Linder et al., 1995b) since negated and simultaneous actions are not allowed there.

Essentially, the logics of commitment can be divided into four categories according to the syntax of commitment representation as follows:

1. "Commitment to goal" as an act,
2. "Commitment to act" as an act (or meta-act),
3. "Commitment to goal" as a proposition,
4. "Commitment to act" as a proposition.

The logic *LC* reported here belongs to the first category since the action-forming operator "!" is applied to some goal (expressed as a wff in Φ_0) and the resultant expression represents an act. While the logic in (van Linder et al., 1995b) belongs to the second category, there are also some alternative formalisms in the same category. For example, in (Dignum et al., 1996; Dignum and van Linder, 1996), an integrated semantics for different mental attitudes are developed. In those papers, an expression of the form "$COMMIT(\alpha)$" is a meta-act if α is an act. The principle "commitment entails obligations" also hold in those logics. However, their logics are based on deontic instead of preference semantics, so the effect of "$COMMIT(\alpha)$" is to update the set of ideal worlds. This kind of semantics, as argued above, would not allow committing to conflicting actions, so according to (Dignum et al., 1996)(p.92), if $COMMIT(\alpha)$ would bring us into a structure in which some world has no ideal successor, it is defined to be equivalent to **fail** (denoted by $\emptyset$ in PD_eL). For commitment to act, it seems reasonable that commitment to conflicting actions simultaneously is not expectable, however, commitment to conflicting goals seems not so unusual in the real situation.

Furthermore, there are also logics belonging to the third category. This kind of logics usually have the well-known BDI architecture as their semantic basis(Cavedon et al., 1997; Singh, 1997b; Singh, 1997a). In (Cavedon et al., 1997), a wff of the form $SCOM_I(\tau, \mu, \varphi)$ means the agent team τ has a social intention- commitment to μ with respect to the goal φ. If not taking the syntactic difference of action and proposition into account, our commitment action can be seen as a special case of the form $SCOM_I(\tau, \tau, \varphi)$, where τ is a fixed individual agent. However, the semantics between these two logics are very different. Since the semantics of social commitment in (Cavedon et al., 1997) is based on standard Kripke semantics, it also suffers from the notorious logical omniscience problem(Fagin et al., 1996) (or more correctly, side-effect problem in the context of commitment), i.e., if a team agent commits to the goal φ, then it also commits to all logical consequence of φ. While in our semantics, $!\varphi$ indeed makes all worlds satisfying φ (and any formulas logically equivalent to φ) preferred to those not, it does not make all worlds satisfying the logical consequence of φ change in the same way. On the other hand, the constructs in (Cavedon et al., 1997) facilitate the comparison of different levels of social commitment from a subservient agent who always fulfill her commitment to vindictive agent who refuses to adopt a team goal. These different constructs, providing a basis to modeling of social attitudes of agents, are lacking in our logic, so to model the multi-agent environment, our logic must be extended further. (Also see the next section for some suggestions.).

In (Singh, 1997b), the notion of commitment is related with economic rationality, so a wff of the form $C_x(p, c)$ means the agent x commits to the goal p to the level c, where c is a real number. In that framework, c may be the utility of achieving p, and after some attempt to execution of action for achieving p, the cost should be subtracted from c, so if after repeated attempts to fulfilling some goal without success, the commitment level of that goal will decrease to zero and become not intended any more. In our logic, we cannot model the utility of goals directly. However, we can imagine to associate a priority level to each of our commitment. Though in our semantics of commitment, we only consider the fair and fcfs strategy, the priority mechanism can be easily added to the fcfs strategy. We only need to rearrange the unfulfilled commitment sequence in the decreasing ordering of priority and then apply the fcfs strategy.

The semantics of commitment is also investigated in a diachronic deontic logic(DDL) proposed in (Brown, 1996). In DDL, commitment is called incurring of obligation, so **incur**A is a wff representing the obligation A is incurred. Syntactically, this kind of commitment belongs to the third category, however, its semantics is very similar in spirit to that of (Dignum et al., 1996). For DDL, each state is associated with a set of propositions called obligation set, and then **incur**A is true in a state s if the proposition A is in the obligation set of s but not in the previous moment of s. The essential difference between DDL and LC from the viewpoint of semantics for commitment is that the incurring of an obligation in DDL just change the obligation set of that state, whereas in LC, the commitment action update preference ordering between possible worlds. However, unlike the logic in (Dignum et al., 1996), the obligation set is not required to be consistent, so simultaneous incurring of conflicting obligations is allowed.

Since our logic emphasizes the influence of goal commitment on preference, it has some relationship with qualitative decision theory(Wellman and Doyle, 1991). In (Wellman and Doyle, 1991), it is shown that goals can be defined in terms of preference and preferences can be derived from sets of goals. Roughly speaking, the former corresponds to the logic of preference, whereas the latter to LC. However, the process of inducing goals from preference ordering in (Wellman and Doyle, 1991) is quite different from the semantics of $LAP2$. According to Wellman and Doyle's semantics, a proposition φ is a goal if φ is preferred to $\neg\varphi$ *ceteris paribus*. In other words, φ is a goal if for any possible worlds w_1 and w_2, when the only difference of w_1 and w_2 is that $w_1 \models \varphi$ but $w_2 \models \neg\varphi$, then $w_1 \geq w_2$. The definition also induces a preference between a proposition and its negation, however, it does not say anything about the preference between any two propositions. In fact, how to induce a preference ordering between any two sets of possible worlds from that imposed on the individual worlds remains a controversial issue. In $LAP2$, we adopt a quite weak definition. A stronger definition is to require that $X > Y$ if all worlds in X are preferred to each in Y. However, this definition runs into the so-called strong preference problem(von Wright, 1963), so a revised definition is proposed in (van der Torre and Tan, 1997). In the revised version, φ is preferred to ψ if every φ world is preferred to or incomparable with any ψ world and φ is obligatory if φ is preferred to $\neg\varphi$ and the optimal worlds are all φ worlds. Even other alternatives are possible, for example, in (Halpern, 1996a), it is defined that $X \succ^* Y$ is for each $v \in Y$, there exists $u \in X$ such that $u > v$ and u dominates Y, where u dominates Y iff for no $w \in Y$ is it the case that $w > u$. On the other hand, in (Bell and Huang, 1997), $X \succ Y$ is defined as all the most normal worlds in X are preferred to each of the most normal worlds in Y, where the most normal worlds are selected according to another normality criteria. All these definitions show that no agreement has been achieved about how to induce a goal from preference yet and we do not claim ours is the better one. Rather we believe each of these definitions may be better in some aspects to others and worse in other aspects. However, we remark that though the formalization of the relationship between commitment and obligation relies somewhat on our adoption of the optimistic set ordering, the principle of commitment implying obligation itself is not influenced by the choice of different definitions. Thus, slightly modified versions of the main results in section 4 will hold if some alternative definitions of the set ordering are used.

Another recent development of qualitative decision theory is the possibilistic logic framework advocated by Dubois et al. in a series of papers[3] (Dubois and Prade, 1995; Dubois and Prade, 1997; Dubois and Prade, 1998; Dubois et al., 1997a; Dubois et al., 1997b; Dubois et al., 1997c; Dubois et al., 1998). The possibilistic logic framework meets with our logics at least at the following two points for some special cases.

First, in the possibilistic logic framework, both uncertainty of knowledge and priorities of goals can be represented by a possibilistic logic base. A possibilistic logic base $K = \{(\varphi_i, c_i) \mid 1 \leq i \leq n\}$ is a set of possibilistic logic formulas, where φ_i is a classical logic formulas and c_i is a level in a finite set $C \subset [0, 1]$. When representing a set of prioritized goals, (φ_i, c_i) means that the priority of achieving φ_i is c_i and the semantics of possibilistic logic naturally providing an approach of deriving preference from a set of prioritized goals.

[3] Thanks to Dr. Prade for calling my attention to this line of development and generously providing me with their papers.

For example, if we start from a neutral model and assume our commitment actions are of the form $!p$, where p is a classical propositional formula, then after a sequence of commitment actions $!p_1; !p_2; \cdots ; !p_i$, we will have a sequence of goals $G = (p_1 \cdot p_2 \cdot \cdots p_n)$. If we adopt the fair strategy, then the most straightforward possibilistic logic base corresponding to G will be $K = \{(p_i, 1) \mid 1 \leq i \leq n\}$ since all goals are equally important. However, in the possibilistic framework, the violation of any goal is not acceptable, whereas in the LC semantics, we consider that even if some goals are violated, the situation in which more goals are satisfied is definitely better than the others. There is however an alternative way to encode a set of equally important goals in a possibilistic logic base. Let $\{S_j \mid 1 \leq j \leq k\}$ be the set of all consistent subsets of G and m be the maximal cardinality of the sets S_j's. Then we can define a set of modified goals

$$\varphi_i = \bigvee \{\chi(S_j, G) \mid \ |S_j| = i\},$$

for all $1 \leq i \leq m$, i.e., φ_i is the characteristic formula of exactly i goals in G being satisfied. Let $K' = \{(\varphi_i, \frac{1}{i}) \mid 1 \leq i \leq m\}$, then the induced ordering be closer to that induced by LC semantics since it ranks the possible worlds according to the number of goals in G being satisfied. The only difference is that since partial order is allowed in our logic, two worlds satisfying same number but not the same set of goals are incomparable, whereas in possibilistic framework, the two worlds will be put in the same rank.

On the other hand, the fcfs strategy is a kind of priority strategy since the goal committed earlier has higher priority. Its special feature is that the priorities of any two goals are distinct. Thus, in this case, the possibilistic logic base corresponding to G is naturally $K = \{(p_i, c_i) \mid 1 \leq i \leq n\}$, where $c_1 > c_2 > \cdots > c_n$. The induced rank order is then somewhat similar to our ordering, however, the main difference is that when the most important goal is violated, our ordering will consider the worlds satisfying the secondly important goal are better than those not, whereas possibilistic semantics will rank them as the same. According to possibilistic logic, it seeks to satisfy the secondly important goals only when the most important one has been satisfied and the solution is totally unacceptable if the most important one is violated.

The second meeting point of our logics with possibilistic logic framework will be the comparison of $LAP2$ semantics with the pessimistic and optimistic utility functions. In the possibilistic logic framework, when the knowledge about the real world is uncertain, the agent can not know exactly what the consequence is after doing some action α, so the possible consequences of doing α is represented as a possibility distribution π_α. On the other hand, the agent's preference (possibly determined by her committed goals according to the way described above) is represented by another possibility distribution μ, then the pessimistic utility of α, a counterpart to expected utility in classical decision theory, is defined as

$$E_*(\alpha) = \min_{w \in \Omega} \max(\mu(w), 1 - \pi_\alpha(w)),$$

and the optimistic one is defined as

$$E^*(\alpha) = \max_{w \in \Omega} \min(\mu(w), \pi_\alpha(w)).$$

Since in $LAP2$, the preference is a partial ordering, we will only consider the special case where the preference is a rank ordering, i.e., the case of LAP. Since in our logics, the epistemic aspect is not considered, the only uncertainty about the consequence after the execution of α is due to the nondeterminism of α, so π_α will be replaced by a subset of Ω in our case. Let us denote the subset by $[\alpha]$. Then, the pessimistic and optimistic utility functions defined above are respectively reduced to

$$E_*(\alpha) = \min_{w \in [\alpha]} \mu(w)$$

and

$$E^*(\alpha) = \max_{w \in [\alpha]} \mu(w).$$

Then the semantics of $\alpha \succ \beta$ in $LAP2$ is precisely a generalization of the criteria $E^*(\alpha) > E^*(\beta)$. A generalization of the pessimistic criteria would also be straightforward.

6. Concluding Remarks

In the preceding sections, a logic for reasoning about some mental attitudes of agents is developed. We start from the dynamic deontic logic, generalize it to a logic of action preference, and then incorporate the commitment and fulfillment actions to get the logic of commitment. We would like to emphasize again that the generalization from deontic logic to preference logic is necessary for the handling of conflicting commitments. If we still use the deontic model as the base of LC and consider the commitment $!\varphi$ as the action modifying the optimal worlds *opt* to the φ-worlds, then when a sequence of commitments are consistent, we can still keep the set *opt* nonempty. However, if we have a commitment sequence $!p; !q$ and p and q are inconsistent, then this will result in an empty set of optimal worlds, and consequently no actions are permitted after the commitment. In the preference-based model, this can be easily handled since the worlds satisfying most committed goals are the relatively optimal ones, so the agent can try to achieve as many goals as possible and leave the unfulfilled ones to the future.

6.1. Perspectives

In the logic of commitment, a commitment action only changes the preference of the agent and does not cause any state transition, whereas a fulfillment action result in both preference change and state transition, so the latter can be seen as partially complementary to the former. We can imagine a proper complementary operator to commitment, i.e. *retraction* . Let $\sharp(\varphi)$ denote the retraction of the goal φ. Then for a sequence $\sigma \in \Pi_0^*$, we can have a weak and strong retraction operators. Specifically, define $\sigma -_w \varphi$ as the result of deleting φ from σ and $\sigma -_s \varphi$ as the result of deleting all formulas $\psi \models_1 \varphi$ from σ. Then the semantics for weak retraction is to define $[\![\sharp\varphi]\!]_3(w,\sigma) = \{(w, \sigma -_w \varphi)\}$, whereas that for strong retraction is $[\![\sharp\varphi]\!]_3(w,\sigma) = \{(w, \sigma -_s \varphi)\}$. Which interpretation of the retraction actions is appropriate will depend on the application. If the retraction is meant to model the avoidance of something to happen, then the strong interpretation is more appropriate. On the other hand, if it is intended to model the canceling of a particular goal, then the weak one may be better.

In the development of LC, we only consider the single agent case, so the commitment may be considered as an internal one. If the logic is extended to the multi-agent environment, then we must consider the problem of external commitment. When an agent wants to achieve some goals but can not, she may ask other agents for help. If agent a accept the request of b and commits to achieve φ, then b may retract φ from her own commitments while a would add φ to hers. So in this case, the action of agent a committing to φ for b, $!(a,b,\varphi)$ is defined as $!(a,\varphi);\sharp(b,\varphi)$. The semantics of the action expressions $!(a,\varphi)$ and $\sharp(b,\varphi)$ is modified to fit the multiagent framework. Formally, $[\![\cdot]\!]_3 : \Pi_3 \to \mathcal{P}((W \times (\Phi_0^*)^n) \times (W \times (\Phi_0^*)^n))$, where n is the number of agents, is defined to satisfy constraints like this: $[\![!(a,\varphi)]\!]_3(w, \sigma_1, \cdots, \sigma_a, \cdots, \sigma_n) = \{(w, \sigma_1, \ldots, \sigma_a \cdot \varphi, \ldots, \sigma_n)\}$.

Another aspect we ignore in this paper is the informational attitudes of agents. For example, in the definition of decision choice action $\alpha \oplus \beta$, we use the preference wff $\alpha \succ \beta$ directly. However, in the practice, it is possible that the agent does not know which action is better for her. So if we introduce the modal operator B into our language, we can redefine the choice action as

$$\alpha \oplus \beta = B(\alpha \succ \beta) \to \alpha/(B(\beta \succ \alpha) \to \beta/(\alpha \cup \beta)).$$

The standard semantics of belief or epistemic modal operators((Fagin et al., 1996)) can be modularly added to ours. Nevertheless, in addition to the influence of agent's belief on her action, some actions modifying her belief, such as **ask**, **tell**, etc., can also be added to the language. However, the semantics of such actions involves the construction of common knowledge and may be rather complicated. Some pioneering work has been done in this topic(van Linder et al., 1994; van Linder et al., 1995a; van Linder et al., 1995c; Gerbrandy and Groeneveld, 1997), however, further study of the interaction of these mental attitudes is still needed. In general, we are convinced that belief, preference (including obligation), and action (including commitment and speech acts) (BPA) are three of the most important mental attitudes for description of agents. We believe that the BPA architecture may constitute a basis for the logical analysis of multi-agent systems.

Finally, in this paper, we mainly focus on the semantics for the logic of commitments. The proof-theoretical aspect of the logic remains unexamined yet. The development of an axiomatic system and some proof methods and the further restriction of the language to a tractable fragment are all topics deserving further study.

Acknowledgement

The research is supported by the NSC of Taiwan under grant no. 88-2213-E-001-001. We thank the anonymous referees for providing valuable comments which result in the improvement of the presentation of the paper.

References

Anderson, A.: 1958, "'A reduction of deontic logic to alethic modal logic"'. *Mind* **67**, 100–103.

Åqvist, L.: 1984, "'Deontic logic"'. In: D. Gabbay and F. Guenthner (eds.): *Handbook of Philosophical Logic, Vol II: Extensions of Classical Logic*. D. Reidel Publishing Company, pp. 605–714.

Bell, J. and Z. Huang: 1997, "'Dynamic goal hierarchies"'. In: L. Cavedon, A. Rao, and W.Wobcke (eds.): *Intelligent Agent Systems : Theoretical and Practical Issues: Workshop at PRICAI'96*. pp. 88–103.

Brown, M.: 1996, "'Doing as we ought: Towards a logic of simply dischargable obligations"'. In: M. Brown and J. Carmo (eds.): *Deontic Logic, Agency, and Normative Systems: The 3rd International Workshop on Deontic Logic in Computer Science*. pp. 47–65.

Cavedon, L., A. Rao, and G. Tidhar: 1997, "'Social and individual commitment"'. In: L. Cavedon, A. Rao, and W.Wobcke (eds.): *Intelligent Agent Systems : Theoretical and Practical Issues: Workshop at PRICAI'96*. pp. 152–163.

Dignum, F., J.-J. Meyer, R. Wieringa, and R. Kuiper: 1996, "'A modal approach to intentions, commitments and obligations: intention plus commitment yields obligation"'. In: M. Brown and J. Carmo (eds.): *Deontic Logic, Agency, and Normative Systems: The 3rd International Workshop on Deontic Logic in Computer Science*. pp. 80–97.

Dignum, F. and B. van Linder: 1996, "'Modelling social agents: Communication as action"'. In: J. Muller, M. Wooldridge, and N. Jennings (eds.): *Intelligent agents III : Agent Theories, Architectures, and Languages : ECAI'96 Workshop (ATAL)*. pp. 205–218.

Dubois, D., D. L. Berre, H. Prade, and R. Sabbadin: 1998, "'Logical representation and computation of optimal decisions in a qualitative setting"'. In: *Proc. of the 15th National Conference on Artificial Intelligence*.

Dubois, D., H. Fargier, and H. Prade: 1997a, "'Decision making under ordinal preferences and uncertainty"'. In: *Proc. of the 13th Annual Conference on Uncertainty in Artificial Intelligence*. pp. 571–164.

Dubois, D. and H. Prade: 1995, "'Possibility theory as a basis for qualitative decision theory"'. In: *Proc. of the 14th International Joint Conference on Artificial Intelligence*. pp. 1924–1930.

Dubois, D. and H. Prade: 1997, "'Constraint satisfaction and decision under uncertainty based on qualitative possibility theory"'. In: *Proc. of the 6th IEEE International Conference on Fuzzy Systems*. pp. 23–30.

Dubois, D. and H. Prade: 1998, "'Posiibilistic logic in decision"'. In: *Proc. of the ECAI'98 Workshop: Decision Theory Meets Artificial Intelligence–Qualitative and Quantitative Approaches*. pp. 11–21.

Dubois, D., H. Prade, and R. Sabbadin: 1997b, "'A possibilistic logic machinery for qualitative decision"'. In: *Proc. of AAAI Spring Symposium on Qualitative Preferences in Deliberation and Practical Reasoning*. pp. 47–54.

Dubois, D., H. Prade, and R. Sabbadin: 1997c, "'Decision under qualitative uncertainty with Sugeno integrals—An axiomatic approach "'. In: *Proc. of the 7th International Fyzzy Systems Association World Congress*. pp. 441–446.

Fagin, R., J. Halpern, Y. Moses, and M. Vardi: 1996, *Reasoning about Knowledge*. MIT Press.

Gerbrandy, J. and W. Groeneveld: 1997, "'Reasoning about information change"'. *Journal of Logic, Language, and Information* **6**, 147–169.

Halpern, J.: 1996a, "'Defining relative likelihood in partially-ordeded preferential structures"'. In: E. Horvitz and F. Jensen (eds.): *Proceedings of the 12th Annual Conference on Uncertainty in Artificial Intelligence*. pp. 299–306.

Halpern, J.: 1996b, 'Should knowledge entail belief?'. *Journal of Philosophical Logic* **25**, 483–494.

Harel, D.: 1984, "'Dynamic logic"'. In: D. Gabbay and F. Guenthner (eds.): *Handbook of Philosophical Logic, Vol II: Extensions of Classical Logic*. D. Reidel Publishing Company, pp. 497–604.

Hintikka, J.: 1962, *Knowledge and Belief*. Cornell University Press.

Liau, C.-J.: 1997, "'A semantics for logics of preference based on possibility theory"'. In: *Proc. of the 7th International Fyzzy Systems Association World Congress*. Prague, pp. 243–248.

Liau, C.-J.: 1998, "'A logic for reasoning about action, preference, and commitment"'. In: *Proc. of the 13th European Conference on Artificial Intelligence*. pp. 552–556.

Meyer, J.-J. C.: 1988, "'A different approach to deontic logic: Deontic logic viewed as a variant of dynamic logic"'. *Notre Dame J. of Formal Logic* **29**(1), 109–136.

Meyer, J.-J. C. and R. Wieringa: 1993, *Deontic Logic in Computer Science: Normative System Specification*. John Wiley & Sons Ltd.

Shoham, Y.: 1993, "'Agent-oriented programming"'. *Artificial Intelligence* **60**(1), 51–92.

Singh, M.: 1997a, "'Commitment among autonomous agents in information-rich environments"'. In: M. Boman and W. V. deVelde (eds.): *Multi-agent rationality : 8th European Workshop on Modelling Autonomous Agents in a Multi-Agent World*. pp. 141–155.

Singh, M.: 1997b, "'Commitment in the architecture of a limited, rational agent"'. In: L. Cavedon, A. Rao, and W.Wobcke (eds.): *Intelligent Agent Systems : Theoretical and Practical Issues: Workshop at PRICAI'96*. pp. 72–87.

van der Hoek, W., B. van Linder, and J.-J. Meyer: 1994, "'A logic of capability"'. In: A. Nerode and Y. V. Matiyasevich (eds.): *Logical Foundation of Computer Science*. pp. 366–378.

van der Torre, L. and Y. Tan: 1997, "'Prohairetic deontic logic and qualitative decision theory"'. In: *Proceedings of AAAI Spring Symposium on Qualitative Preferences in Deliberation and Practical Reasoning*.

van Linder, B., W. van der Hoek, and J.-J. Meyer: 1994, "'Tests as epistemic updates"'. In: A. Cohen (ed.): *Proc. of ECAI*. pp. 331–335.

van Linder, B., W. van der Hoek, and J.-J. Meyer: 1995a, "'Actions that make you change your mind"'. In: I. Wachsmuth, C. Rollinger, and W. Brauer (eds.): *Proc. of KI-95*. pp. 185–196.

van Linder, B., W. van der Hoek, and J.-J. Meyer: 1995b, "'Formalizing motivational attitudes of agents: On preference, goals and commitments"'. In: M. Wooldrige, J. Muller, and M. Tambe (eds.): *Intelligent Agents II: Agent Theories, Architectures, and Languages*. pp. 17–32.

van Linder, B., W. van der Hoek, and J.-J. Meyer: 1995c, "'Seeing is believing: and so are hearing and jumping"'. In: M. Gori and G. Soda (eds.): *Proc. of AI*IA-95*. pp. 402–413.

Veltman, F.: 1996, "'Defaults in update semantics"'. *Journal of Philosophical Logic* **25**, 221–261.

von Wright, G.: 1951, "'Deontic logic"'. *Mind* **60**, 1–15.

von Wright, G.: 1963, *The Logic of Preference*. Edinburgh University Press.

Wellman, M. and J. Doyle: 1991, "'Preferential semantics for goals"'. In: *Proceedings of the 9th National Conference on Artificial Intelligence*. pp. 698–703.

Zadeh, L.: 1978, "'Fuzzy sets as a basis for a theory of possibility"'. *Fuzzy Sets and Systems* **1**(1), 3–28.