A Logic for Reasoning about Action, Preference, and Commitment

Churn-Jung Liau[1]

Abstract. In this paper, we study the logical semantics of actions, preference, and commitment. The semantics 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. While the preference will influence the agent's choice of actions, commitment is a kind of actions that will change the agent's preference. The integrated semantics make it possible to express and reasoning about the mutual relationship among these mental attitudes in a common logic.
Key words: Deontic logic, dynamic logic, logic of commitment, logic of preference.

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[12]. Since then, different formal deontic logic systems have been developed. Among them, Meyer's propositional deontic logic is one of the most interesting[6]. His system is based on the reduction to dynamic logic[4], so both actions and propositions can be represented and the distinction of "ought-to-do" and "ought-to-be" can be made clearly. Recently, the applications of deontic logics to computer science and artificial intelligence has received more and more attention[7]. Since dynamic logic is the logic for reasoning about computer program, Meyer's system 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[1] 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 [5], 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.

On the other hand, though in agent-oriented programming[8], commitment plays an important role, its semantics is not formally given there. In our study, we find it is feasible to integrate commitment into the logic of preference. While ordinary actions in dynamic logic induce state transition relations on the possible world models, the commitment actions will change the agent's preference. Then, the agent will choose her action according to the preference induced from her commitment. Consequently, we can have a logic for reasoning about actions, preference, and commitment simultaneously.

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. Finally, we give the conclusion and some perspective of the works.

2 Review of Deontic Logic as Dynamic 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
   - $A \cup \{\emptyset, \mathbf{u}, \mathbf{i}\} \subseteq \Pi_1$, and
   - if $\alpha, \beta \in \Pi_1$ and $\varphi \in \Phi_1$, then $\alpha;\beta, \alpha \cup \beta, \alpha\&\beta, \overline{\alpha}, \varphi \rightarrow \alpha/\beta \in \Pi_1$.

Let $\Phi_0 = \{\varphi \in \Phi_1 \mid V$ does not occur in $\varphi\}$ denote the set of ordinary dynamic logic formulas, and $\Pi_0 = \{\alpha \in \Pi_1 \mid V$ 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

[1] Institute of Information Science, Academia Sinica, Taipei,Taiwan, E-mail: liaucj@iis.sinica.edu.tw

ECAI 98. *13th European Conference on Artificial Intelligence*
Edited by Henri Prade
Published in 1998 by John Wiley & Sons, Ltd.

composition, nondeterministic choice, and simultaneous execution of α and β respectively, whereas $\overline{\alpha}$ means the non-execution of α and $\varphi \rightarrow \alpha/\beta$ 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 program. 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. Since the formal semantics is rather involved, we will only present the informal one based on the ordinary state transition semantics of dynamic logics.

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 $[\![\cdot]\!]_1 : \Pi_1 \rightarrow \mathcal{P}(W \times W)$ 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. $[\![\overline{\varphi \rightarrow \alpha/\beta}]\!]_1 = [\![\varphi \rightarrow \overline{\alpha}/\overline{\beta}]\!]_1$

  ($[\![\cdot]\!]_1$6): (conditional actions)

$$[\![\varphi \rightarrow \alpha/\beta]\!]_1(w) = \begin{cases} [\![\alpha]\!]_1(w), & \text{if } w \models_1 \varphi, \\ [\![\beta]\!]_1(w), & \text{if } w \not\models_1 \varphi. \end{cases}$$

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 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 $\alpha \in \Pi_1$.

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

If $M = \langle W, \models_1, [\![\cdot]\!]_1, opt\rangle$ 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 $\langle W, \models_2, [\![\cdot]\!]_2, \geq\rangle$, 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. Then, we assume in any preferential model, the set of optimal worlds is nonempty. 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. $u > v$). Then $[\![\cdot]\!]_2 : \Pi_2 \rightarrow \mathcal{P}(W \times W)$ must still satisfy ($[\![\cdot]\!]_1$1)–($[\![\cdot]\!]_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 $[\![\alpha]\!]_2(w) \geq [\![\beta]\!]_2(w)$,
($\models_2$5) $w \models_2 \alpha \succ \beta$ iff $[\![\alpha]\!]_2(w) > [\![\beta]\!]_2(w)$,

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.

Proposition 1 *The following wffs are valid in $LAP2$.*

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

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

Some features of the semnatics deserve further explanation. First, since the logic is motivated from PD_eL, the preference relation on the worlds reflects the deontic ideality of them. Nevertheless, this does not exclude the possible existence of other preference relations. For example, according to the semantics, the wff $(\beta \succ \alpha) \supset F\alpha$ is valid. This is reasonable since the antecedent of the conditional means that the execution of α will result in deontically subideal worlds. However, if we consider an alternative binary connective $\succ'$ interpreted in the same way as above by another preference relation $\geq'$ except that the clause ($\models_2 0$') is not imposed on $>'$, then $(\beta \succ' \alpha) \supset F\alpha$ is no longer valid. Intuitively, a driver can prefer violating traffic law. However, in this case, his preference is not in the same direction with the deontic ideality imposed by the law.

Second, the definition of $X \geq Y$ for $X, Y \subseteq W$ from the preference relation between worlds is in a somewhat weak form, so the correct reading of $\alpha \succ \beta$ should be "there is some way by which the execution of α would be better than β". For example, the third wff of Proposition 1. seems counterintuitive at the first glance since an agent may prefer the joint execution of α_1 and α_2 (e.g. drinking white wine and eating fish) over β(e.g. drinking red wine) without preferring α_i ($i = 1, 2$) alone over β. However, according to our reading, if the agent prefer drinking white wine and eating fish over drinking red wine, then she must prefer drinking white wine in some way (i.e. in the way of eating fish simultaneously) over drinking red wine. Of course, alternative definitions of $X \geq Y$ are possible which will result in different valid wffs in the system. For example, a stronger definition is that $X \geq Y$ if for all $u \in X$ and $v \in Y$, $u \geq v$ or the definition proposed recently in [3] can be used. These alternative definitions would be explored further in the future.

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

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

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. Interestingly, $LAP2$ is a conservative extension of PD_eL.

Theorem 1 *If $S \cup \{\varphi\} \subseteq \Phi_1$, then $S \models_1 \varphi$ iff $S \models_2 \varphi$.*

This result shows that $LAP2$ is indeed more expressive than PD_eL in deontic reasoning.

In the preferential models, the relation $\geq$ is a preorder. A preorder is also called a ranked order 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[13], 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 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.

A *commitment model* is a sextuple $\langle W, \models_1, [\![\cdot]\!]_1, \models_3, [\![\cdot]\!]_3, \geq\rangle$, 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. $[\![\cdot]\!]_1 : \Pi_1 \rightarrow \mathcal{P}(W \times W)$ satisfies ($[\![\cdot]\!]_1 1$)–($[\![\cdot]\!]_1 6$),
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 $w, \sigma \models_3 \psi$,
   ($\models_3 2$) $w, \sigma \models_3 \neg\varphi \Leftrightarrow w, \sigma \not\models_3 \varphi$,
   ($\models_3 3$) $w, \sigma \models_3 [\alpha]\varphi \Leftrightarrow \forall (u, \tau) \in [\![\alpha]\!]_3(w, \sigma), (u, \tau \models_3 \varphi)$, where $[\![\alpha]\!]_3(w, \sigma) = \{(u, \tau) \mid (((w, \sigma), (u, \tau)) \in [\![\alpha]\!]_3\}$,
   ($\models_3 4$) $w, \sigma \models_3 \alpha \succeq \beta$ iff $P_W([\![\alpha]\!]_3(w, \sigma)) \geq_\sigma P_W([\![\beta]\!]_3(w, \sigma))$, and
   ($\models_3 5$) $w, \sigma \models_3 \alpha \succ \beta$ iff $P_W([\![\alpha]\!]_3(w, \sigma)) >_\sigma P_W([\![\beta]\!]_3(w, \sigma))$, and
5. $[\![\cdot]\!]_3 : \Pi_3 \rightarrow \mathcal{P}((W \times \Phi_0^*) \times (W \times \Phi_0^*))$ must satisfy

(coherence): if $\alpha \in \Pi_0$, then $[\![\alpha]\!]_3(w,\sigma) = \{(u,\sigma) \mid u \in [\![\alpha]\!]_1(w)\}$,
(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).

The definition of $S \models_3 \varphi$ is as above for all $S\cup\{\varphi\} \subseteq \Phi_3$. 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 $[\![\cdot]\!]_3$. Note that $\models_3$ and $[\![\cdot]\!]_3$ only agree with $\models_1$ and $[\![\cdot]\!]_1$ 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 $[\![\cdot]\!]_3$ directly.

It can be seen that for all $\alpha \in \Pi_2$, if $((w,\sigma),(u,\tau)) \in [\![\alpha]\!]_3$, 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.

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

Proposition 2 *Let $\models_3$ denote either $\models_3^{fair}$ or $\models_3^{fcfs}$, 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. $\models_3^{fcfs} \alpha \succ \beta \supset [!\varphi]\alpha \succ \beta$

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 $M_0 = \langle W, \models_1, [\![\cdot]\!]_1, \models_3, [\![\cdot]\!]_3, \geq_0\rangle$, where $\geq_0 = W \times W$. Such models are also called *neutral* models. Let $\mathcal{M}_0$ denote the class of all neutral models.

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 "·" 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. $\sigma_{s\cdot@\varphi} = \sigma_s\backslash\varphi$.

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 $\beta \in \Pi_2$, $w,\sigma \models_3^{fair} [\overline{\beta}]\neg\varphi_\sigma \supset O\beta$,*
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 $\alpha,\beta \in \Pi_2$, $w,\sigma \models_3^{fair} \langle\alpha\rangle\chi(S_1,\sigma) \wedge [\beta]\chi(S_2,\sigma) \supset \alpha \succ \beta$*

Proposition 3 *If M_0, w, and s are defined as above, then*

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

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. 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 [9]). 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. Given a $\sigma = \varphi_1\cdot\varphi_2\cdots\varphi_n \in \Phi_0^*$, 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 4 *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 3 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 Concluding Remarks

In the preceding sections, a logic for reasoning about some mental attitudes of agents is developed. We start from the deontic dynamic 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.

5.1 Related works

The idea of considering actions updating the preference relation comes originally from the update semantics for default logic[11]. In the semantics, each subset of worlds (i.e. a proposition) is associated with a preference relation and a default of the form $\varphi \leadsto \psi$, 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 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 [8]. However, in [8], the commitment and obligation are not distinguished. Here, we model both of them in a common logic, and the first part of Proposition 3 in fact provide a formal relationship between them. That is, for a neutral agent, the unfulfilled commitment decides her obligations.

Some motivational attitudes, including preference, goal, and commitment, have been also formalized in [10]. However, in [10], 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 [10] are quite different. Moreover, the action terms is more restrictive in [10] since negated and simultaneous actions are not allowed there.

5.2 Perspectives

In the logic of commitment, a commitment action only change the preference of the agent and do 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 imange 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: $[\![!(a,\varphi)]\!]_3(w, \sigma_1, \cdots, \sigma_a, \cdots, \sigma_n) = \{(w, \sigma_1, \ldots, \sigma_a \cdot \varphi, \ldots, \sigma_n)\}$. These features would be incorporated in the full paper.

ACKNOWLEDGEMENTS

The research is partially supported by the NSC of Taiwan(ROC) under grant no. 87-2213-E-001-010. Two anonymous referees provide valuable comments. Though due to space and time limit, some of their valuable suggestions could not be discussed in detail here, it is expected that all of these would be incorporated in the full paper.

REFERENCES

[1] A.R. Anderson, "'A reduction of deontic logic to alethic modal logic"', *Mind*, **67**, 100–103, (1958).

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

[3] J.Y. Halpern, "'Defining relative likelihood in partially-ordered preferential structures"', *J. of Artificial Intelligence Research*, **7**, 1–24, (1997).

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

[5] C.-J. Liau, "'A semantics for logics of preference based on possibility theory"', in *Proc. of the 7th International Fyzzy Systems Association World Congress*, Prague, (June 1997).

[6] J.-J. Ch. Meyer, "'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, (1988).

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

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

[9] W. van der Hoek, B. van Linder, and J.-J. Meyer, "'A logic of capability"', in *Logical Foundation of Computer Science*, eds., A. Nerode and Yu. V. Matiyasevich, LNCS 813, pp. 366–378. Springer-Verlag, (1994).

[10] B. van Linder, W. van der Hoek, and J.-J.Ch. Meyer, "'Formalizing motivational attitudes of agents: On preference, goals and commitments"', in *Intelligent Agents II: Agent Theories, Architectures, and Languages*, eds., M. Wooldrige, J.P. Muller, and M. Tambe, LNAI 1037, pp. 17–32. Springer-Verlag, (1995).

[11] F. Veltman, "'Defaults in update semantics"', *Journal of Philosophical Logic*, **25**, 221–261, (1996).

[12] G.H. von Wright, "'Deontic logic"', *Mind*, **60**, 1–15, (1951).

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