# Don't ever do that!
## Long-term duties in $PD_eL$

Jesse Hughes and Lambèr Royakkers

January 27, 2006

**Abstract**

This paper studies long-term norms concerning actions. In Meyer's Propositional Deontic Logic ($PD_eL$), only immediate duties can be expressed, however, often one has duties of longer durations such as: "Never do that", or "Do this someday". In this paper, we will investigate how to amend $PD_eL$ so that such long-term duties can be expressed. This leads to the interesting and suprising consequence that the long-term prohibition and obligation are not interdefinable in our semantics, while there is a duality between these two notions. As a consequence, we have provided a new analysis of the long-term obligation by introducing a new atomic proposition $I$ (indebtedness) to represent the condition that an agent has some unfulfilled obligation.

## 1 Introduction

The classical deontic logic introduced by von Wright (1951) is based on a set of "ideal" or "perfect" worlds, in which all obligations are fulfilled, and introduces formula-binding deontic operators. Meyer (1988, 1989) instead based deontic logic on dynamic logic by introducing a special violation atom $V$, indicating that in the state of concern a violation of the deontic constraints has been committed. But there is a deeper difference than this stress of violation over ideal outcomes. Namely, Meyer's $PD_eL$ (Propositional Deontic Logic) is a *dynamic* logic.

Following Anderson's proposal in (1967), Meyer introduced deontic operators to propositional dynamic logic ($PDL$) as follows: an action $\alpha$ is forbidden in $w$ if doing $\alpha$ in $w$ inevitably leads to violation. Similarly, $\alpha$ is obligatory in $w$ if doing anything *other than* $\alpha$ inevitably leads to violation. In $PD_eL$, then, duties bind actions rather than conditions: one is obligated to *do* something, rather than bring about some condition.

The benefit from this approach of reducing deontic logic to dynamic logic is twofold. Firstly, in this way we get rid of most of the nasty paradoxes that have plagued traditional deontic logic (cf. Castañeda (1981)), and secondly, we have the additional advantage that by taking this approach to deontic logic and employing it for the specification of integrity constraints for knowledge based systems we can directly integrate deontic constraints with the dynamic ones.

Nonetheless, $PD_eL$ comes with its own limitations, notably in the kinds of ought-to-do statements that can be expressed. In particular, $PD_eL$'s deontic operators express norms about *immediate* rather than *eventual* actions. Certainly, some prohibitions are narrow in scope: "Do not do that *now*." But other prohibitions restrict action more broadly: "Don't *ever* do that" (i.e., at every point in the future, do not do that). Our aim here is to investigate how to amend $PD_eL$ so that such long-term norms can be expressed.

Interestingly, our semantics for long-term obligation is not as closely related to long-term prohibition as one might expect. The essential difference comes in evaluating whether a norm has been violated. A long-term prohibition against $\alpha$ is violated if there is some time at which the agent has done $\alpha$. Thus, long-term prohibitions can be expressed in terms of reaching worlds in violation. Long-term obligations are different: an obligation to do $\alpha$ is violated just in case the agent *never* does $\alpha$. But there is no world corresponding to this condition. At each world, the agent may later do $\alpha$ and thus fulfill his obligation. In learning-theoretic terms (Kelly (1996)), violations of prohibitions are verifiable with certainty but not refutable, and dually fulfillment of obligations are verifiable with certainty but not refutable.

Thus, while there *is* a duality between long-term prohibitions and obligations, the two notions are not inter-definable in our possible world semantics. Instead, we must provide a new analysis of obligation that is considerably subtler than Meyer's definition of immediate obligation. We find that the asymmetry between our long-term normative concepts is one of the most interesting and surprising consequences of our investigations.

Our presentation begins with a summary of a somewhat simplified version of $PD_eL$, introducing Meyer's definitions of (immediate) prohibition and obligation. In Section 3, we introduce our definition of long-term contiguous prohibition, an admonition to never perform a particular sequence of actions one after the other. We also introduce long-term contiguous obligation and explain why inter-definability fails for these two concepts. In Section 4, we briefly discuss non-contiguous variations for prohibitions and obligations. These include prohibitions against doing a sequence of actions in order, but with other actions interspersed (and an analog for obligations).

We close with a few comments about future directions for dynamic deontic logic.

For reasons of space, we have omitted most of the proofs. However, we have given enough properties of the relations and concepts involved so that the missing derivations are simple and straightforward. We have included a few proofs where the reasoning is not obvious and immediate from previous discussion, but our focus here is on semantic appropriateness rather than technical developments.

# 2 The basic system $PD_eL$

We present here a somewhat simplified form of $PD_eL$. Our presentation is primarily based on Meyer (1988, 1989).

## 2.1 Actions and their interpretations

$PD_eL$ is a dynamic logic aimed at reasoning about duties and prohibitions. It differs from most deontic logics by focusing on actions rather than conditions: things one ought to do (or not do) rather than conditions one ought to bring about (or avoid). The syntax is similar to other dynamic logics, with complex actions built from a set $A$ of atomic actions and complex propositions built from a set of atomic propositions. We use $a, b, \ldots$ to range over $A$. The semantics, too, are similar to other dynamic logics: models are given by an labeled transition system on a set $\mathcal{W}$ of worlds and actions are interpreted as sets of paths in this transition system.

$PD_eL$ differs from classical $PDL$ (Harel (1984); Meyer (2000)) primarily in the set of action-constructors. In particular, $PD_eL$ includes synchronous composition (doing both $\alpha$ and $\beta$ at the same time) and negation (doing something *other* than $\alpha$). Synchronous composition adds a new degree of non-determinism to our semantics, because actions can be specified to greater or lesser degree. If one does $\alpha$ & $\beta$, then he has done $\alpha$, but the converse is not true. On this approach, even atomic actions are not fully specified: $\underline{a}$ is interpreted as a set of alternatives.

Therefore, our semantics comes with an extra step: we interpret actions as sets of sequences of *fully specified one-step actions* (what Meyer calls "synchronicity sets" or "s-sets"). Meyer took his set of fully specified one-step actions to be $\mathcal{P}^+\mathcal{P}^+A$, where $\mathcal{P}^+S$ is the set $\mathcal{P}S \setminus \{\emptyset\}$ of non-empty subsets of $S$. He maps atoms $\underline{a}$ to subsets of $\mathcal{P}^+A$ via

$$\underline{a} \mapsto \{\, S \subseteq A \;\mid\; a \in S \,\},$$

and hence interpretation of actions is a function $TA \to \mathcal{P}^+\mathcal{P}^+A$. This concrete interpretation is well-motivated, but we prefer a simpler, more flexible and abstract approach. We fix a set $X$ to be our fully specified one-step actions together with a function

$$i : A \to \mathcal{P}^+X,$$

where we recover Meyer's interpretation by choosing $X = \mathcal{P}^+A$ and using the mapping above. Our alternative is more flexible in the following sense: in Meyer (1989), each pair of atomic actions can be performed simultaneously, i.e. $[\![\underline{a} \,\&\, \underline{b}]\!] \neq \emptyset$, but this is not always reasonable. By choosing $X$ and $i$ appropriately, we allow that some pairs of atomic actions (whistling and chewing crackers, say) cannot be done at the same time.

To summarize: we fix a set $A$ of atomic actions, a set $X$ of fully specified one-step actions and a function $i : A \to \mathcal{P}X$ (together with a set of atomic propositions). We build a set $TA$ of *action terms* from the elements of $A$. Each action will be interpreted as a set of sequences over $X$, yielding

$$[\![-]\!] : TA \to \mathcal{P}(X^{<\omega}).$$

The set $[\![\alpha]\!]$ represents the alternative fully specified ways of doing $\alpha$. Such $X$-sequences will define a set of paths in our $X$-labeled transition system on $\mathcal{W}$ and this yields the usual interpretation of the dynamic operator $[\alpha]$, but let us not get ahead of ourselves.

The set $TA$ of action terms is defined by

$$\beta ::= \underline{a} \mid \underline{\emptyset} \mid \epsilon \mid \textbf{any} \mid \alpha \cup \beta \mid \alpha \,\&\, \beta \mid \alpha;\beta \mid \overline{\beta}$$

The action $\underline{a}$ represents the (not fully specified) atomic action $a$, $\underline{\emptyset}$ the impossible action, $\epsilon$ the do-nothing action,[1] **any** the do-any atomic action and $\textbf{any}^*$ the do-any complex action. As mentioned, $\alpha \,\&\, \beta$ represents simultaneous performance of $\alpha$ and $\beta$ and $\overline{\beta}$ represents doing *anything but* $\beta$. As usual, $\alpha \cup \beta$ represents the non-deterministic choice between $\alpha$ and $\beta$ and $\alpha;\beta$ for the sequential composition of actions $\alpha$ and $\beta$.

Before defining the interpretation $TA \to \mathcal{P}(X^{<\omega})$, we must introduce a bit of terminology for sequences.

If $n \leq |s|$, the sequence $s \restriction n$ is the prefix of $s$ of length $n$, i.e.

$$s \restriction n = \langle s_0, s_1, \ldots, s_{n-1} \rangle.$$

If $n \geq |s|$, then $s \restriction n = s$.

We write $s * t$ for the concatenation of $s$ and $t$. We say that $s$ is a *prefix* of $t$ if there is some $n$ such that $s = t \restriction n$, equivalently there is some $r$ such that $t = s * r$, and $s$ is a *proper* prefix if the chosen $r$ is not the empty sequence $\langle\rangle$. Two sequences are *comparable* if one is a prefix of the other. If $S$ is a set of sequences, we define

$$\mathsf{cmp}(S, s) \Leftrightarrow \exists t \in S \;.\; t \text{ is comparable to } s.$$

A set $S$ of sequences is *$n$-uniform* iff every sequence $s$ in $S$ has length $n$. If $S$ is $n$-uniform for some $n$, then it is *uniform*. We will also say that $\alpha$ is ($n$-)uniform if $[\![\alpha]\!]$

[1] This symbol does not occur in other versions of dynamic logic in the literature. It is, however, comparable with the $\epsilon$ process in process algebra.

Definition of $[\![\alpha]\!]$

$$
\begin{aligned}
[\![\underline{a}]\!] &= \{\, \langle x \rangle \;\mid\; x \in i(a) \,\} \\
[\![\mathbf{any}]\!] &= \{\, \langle x \rangle \;\mid\; x \in X \,\} \\
[\![\mathbf{any}^*]\!] &= X^{<\omega} \\
[\![\underline{\emptyset}]\!] &= \emptyset \\
[\![\epsilon]\!] &= \{\langle\ \rangle\} \\
[\![\alpha \cup \beta]\!] &= [\![\alpha]\!] \cup [\![\beta]\!] \\
[\![\alpha;\beta]\!] &= \{\, r * s \;\mid\; r \in [\![\alpha]\!],\ s \in [\![\beta]\!] \,\} \\
[\![\alpha_1 \mathbin{\&} \alpha_2]\!] &= \{\, s \;\mid\; s \in [\![\alpha_i]\!] \text{ and } \exists n \,.\, s \upharpoonright n \in [\![\alpha_j]\!],\ j \neq i \,\} \\
[\![\overline{\alpha}]\!] &= \begin{cases} \{\, s * \langle x \rangle \;\mid\; \mathsf{cmp}([\![\alpha]\!], s) \wedge \neg \mathsf{cmp}([\![\alpha]\!], s * \langle x \rangle) \,\} & \text{if } [\![\alpha]\!] \neq \emptyset \\ [\![\mathbf{any}]\!] & \text{else} \end{cases}
\end{aligned}
$$

Table 1: The interpretation of actions as sets of $X$-sequences.

is and that a set $S \subseteq TA$ of actions is uniform if there is some $n$ such that each $\alpha \in S$ is $n$-uniform.

Our definition of $[\![-]\!] : TA \to \mathcal{P}(X^{<\omega})$ is found in Table 1. This definition is a slight simplification of Meyer (1989). In Meyer's system, all the sequences are infinite, but only finite initial segments are *relevant* (specified) by marking the s-sets. We do not deal with marked s-sets, since we admit finite sequences and all s-sets in the treated sequences we consider "relevant". Furthermore, we do not have the restriction of actions to be in normal form, i.e., every subexpression of the form $\alpha \cup \beta$ has the property that $\alpha \mathbin{\&} \beta =_{\mathcal{A}} \underline{\emptyset}$, and dually, every subexpression of the form $\alpha \mathbin{\&} \beta$ has the property that $\overline{\alpha} \mathbin{\&} \overline{\beta} =_{\mathcal{A}} \underline{\emptyset}$. This condition is necessary in Meyer's system, since otherwise some axioms would not be sound.[2] This is a result of his definition of the "$\cup$"-operator, which is not the set-theoretic union as in our language. It gives the union of two sets of sequences but subtracts every sequence $s$ in the union comparable with some sequence $t$ in the union and is not a proper prefix of $t$. So, $[\![\alpha \cup (\alpha;\beta)]\!] \subseteq [\![\alpha]\!]$, which is not a property in our language.

Consequently, we lose a few properties, such as the desirable property $\overline{\overline{\beta}} =_{\mathcal{A}} \beta$. However, these properties play no significant role in our development of long-term norms. Thus, we prefer to simplify $PD_eL$ and focus on the original work as far as possible.

See Figures 1 and 2 for a pictorial explanation of & and $\overline{-}$.

If every sequence in $[\![\alpha]\!]$ is also in $[\![\beta]\!]$, then consequences of doing $\beta$ are also consequences of doing $\alpha$. Because this fact is so basic to our reasoning, we introduce the partial order $\leq_{\mathcal{A}}$, defined by

$$\alpha \leq_{\mathcal{A}} \beta \text{ iff } [\![\alpha]\!] \subseteq [\![\beta]\!].$$

We write $\alpha =_{\mathcal{A}} \beta$ iff $[\![\alpha]\!] = [\![\beta]\!]$.

We give some of the basic properties regarding $\leq_{\mathcal{A}}$ and $=_{\mathcal{A}}$ in Table 2. In each case, the derivation is fairly simple. Moreover, Table 2 contains every property needed to derive the properties discussed hereafter.

[2]E.g. axiom $[\alpha \cup \beta]\phi \equiv [\alpha]\phi \wedge [\beta]\phi$: In Meyer's system it holds that $\underline{a} =_{\mathcal{A}} \underline{a} \cup \underline{a};\underline{b}$, however, $[\underline{a} \cup \underline{a};\underline{b}]\phi \equiv [\underline{a}]\phi \not\equiv [\underline{a}]\phi \wedge [\underline{a};\underline{b}]\phi$.

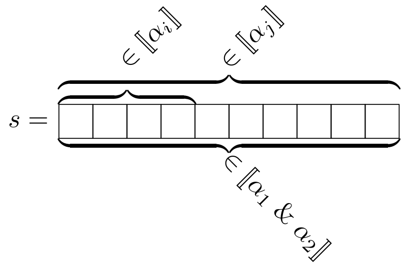


Figure 1: $s$ is in $[\![\alpha_1 \mathbin{\&} \alpha_2]\!]$ if $s$ is in one of the $\alpha_i$'s and a prefix of $s$ is in the other.

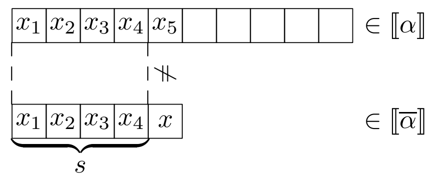


Figure 2: $s * \langle x \rangle \in [\![\overline{\alpha}]\!]$ if $s$ is comparable to something in $[\![\alpha]\!]$, but $s * \langle x \rangle$ is not.

| **Properties of $\leq_\mathcal{A}$** | |
|---|---|
| $\overline{\alpha \cup \beta} =_\mathcal{A} \overline{\alpha} \mathbin{\&} \overline{\beta}$ | $\alpha \mathbin{\&} \beta \leq_\mathcal{A} \alpha$ if $\{\alpha, \beta\}$ is uniform |
| $\overline{\alpha \mathbin{\&} \beta} =_\mathcal{A} \overline{\alpha} \cup \overline{\beta}$ if $[\![\alpha \mathbin{\&} \beta]\!] \neq \emptyset$ | $\alpha \mathbin{\&} \beta \leq_\mathcal{A} \alpha; \mathbf{any}^*$ |
| $\overline{\alpha} \leq_\mathcal{A} \overline{\alpha; \beta}$ | $\alpha \mathbin{\&} (\beta \cup \gamma) =_\mathcal{A} (\alpha \mathbin{\&} \beta) \cup (\alpha \mathbin{\&} \gamma)$ |
| $\alpha \leq_\mathcal{A} \mathbf{any}^*$ | $\alpha \cup (\beta \mathbin{\&} \gamma) \leq_\mathcal{A} (\alpha \cup \beta) \mathbin{\&} (\alpha \cup \gamma)$ |
| $\alpha \leq_\mathcal{A} \alpha; \mathbf{any}^*$ | $\overline{\mathbf{any}} =_\mathcal{A} \overline{\mathbf{any}^*} =_\mathcal{A} \emptyset$ |
| $\alpha \leq_\mathcal{A} \mathbf{any}^*; \alpha$ | $\overline{\alpha} =_\mathcal{A} \overline{\alpha; \mathbf{any}} =_\mathcal{A} \overline{\alpha; \mathbf{any}^*}$ |
| $\underline{a} \leq_\mathcal{A} \mathbf{any}$ | $\overline{\mathbf{any}^*; \alpha} =_\mathcal{A} \underline{\emptyset}$ if $[\![\alpha]\!] \neq \emptyset$ |
| $\alpha; \underline{\emptyset} =_\mathcal{A} \underline{\emptyset}; \alpha =_\mathcal{A} \underline{\emptyset}$ | $\overline{\underline{\emptyset}} =_\mathcal{A} \mathbf{any}$ |
| $\epsilon; \beta =_\mathcal{A} \beta; \epsilon =_\mathcal{A} \beta$ | |
| $\left.\begin{array}{l} \alpha \cup \gamma \leq_\mathcal{A} \beta \cup \delta \\ \alpha; \gamma \leq_\mathcal{A} \beta; \delta \end{array}\right\}$ if $\alpha \leq_\mathcal{A} \beta$ and $\gamma \leq_\mathcal{A} \delta$ | |

Table 2: Basic properties of $\leq_\mathcal{A}$.

## 2.2 Formulas and their interpretations

In the previous section, we interpreted action terms as sets of sequences over $X$. The final step in defining semantics for $PD_eL$ is interpreting sequence-world pairs as sets of paths in our model and using this to interpret formulas. In fact, as with other dynamic logics, we are not interested in the paths per se, but only with the final worlds in each path. This simplifies our definitions a bit.

A $PD_eL$ model consists of a set $\mathcal{W}$ of worlds together with an $X$-labeled transition system and an interpretation of atomic propositions. We define an interpretation

$$[\![-]\!] : X^{<\omega} \to (\mathcal{P}\mathcal{W})^\mathcal{W}$$

taking a sequence $\langle x_1, \ldots, x_n \rangle$ and world $w$ to the set of all $w'$'s reachable from $w$ via a path like so:

$$w \xrightarrow{x_1} w_1 \xrightarrow{x_2} \cdots \xrightarrow{x_n} w'.$$

Explicitly,

$$[\![\langle\rangle]\!](w) = \{w\},$$
$$[\![s * \langle x\rangle]\!](w) = \{\, w' \mid \exists w'' \in [\![s]\!](w) \text{ and } w'' \xrightarrow{x} w' \,\}.$$

This induces an interpretation $[\![-]\!] : TA \to (\mathcal{PW})^{\mathcal{W}}$ defined by

$$[\![\alpha]\!](w) = \{\, w' \mid \exists s \in [\![\alpha]\!] \;.\; w' \in [\![s]\!](w) \,\}.$$

Clearly, we are overloading the notation $[\![-]\!]$ here, but we hope that our meaning is clear from context. When we write $[\![\alpha]\!]$, we mean a set of $X$-sequences and when we write $[\![\alpha]\!](w)$, we mean a set of worlds.

Assertions in $PD_eL$ are either the atomic proposition, logical compositions $\neg\phi$, $\phi_1 \vee \phi_2$, $\phi_1 \wedge \phi_2$, $\phi_1 \to \phi_2$, $\phi_1 \equiv \phi_2$, or expressions of the form $[\alpha]\phi$ with intended meaning that $\phi$ holds after the performance of action $\alpha$. The semantics of the formula $[\alpha]\phi$ is defined by

$$w \models [\alpha]\phi \text{ iff } \forall w' \in [\![\alpha]\!](w) \;.\; w' \models \phi.$$

**Inference rules**

$$\frac{\phi}{[\alpha]\phi}(\text{N}) \qquad\qquad \frac{\phi \to \psi \quad \phi}{\psi}(\text{MP})$$

**Axioms**

| | |
|---|---|
| every propositional tautology | $[\beta]\phi \to [\alpha]\phi$ if $\alpha \leq_{\mathcal{A}} \beta$ |
| $[\beta](\phi_1 \to \phi_2) \to ([\beta]\phi_1 \to [\beta]\phi_2)$ | $[\alpha \cup \beta]\phi \equiv [\alpha]\phi \wedge [\beta]\phi$ |
| $[\alpha;\beta]\phi \equiv [\alpha]([\beta]\phi)$ | $[\alpha]\phi \vee [\beta]\phi \to [\alpha \,\&\, \beta]\phi$ if $\{\alpha, \beta\}$ is uniform |
| $[\emptyset]\phi$ | $[\mathbf{any}]\phi \to [\underline{a}]\phi$ |
| $[\epsilon]\phi \equiv \phi$ | $[\mathbf{any}^*]\phi \to \phi \wedge [\mathbf{any}][\mathbf{any}^*]\phi$ |

**Deontic definitions**

| | |
|---|---|
| $f(\alpha) \equiv [\alpha]V$ | $o(\alpha) \equiv [\overline{\alpha}]V$ |

**$PD_eL$ theorems**

| | |
|---|---|
| $f(\beta) \to f(\alpha)$ if $\alpha \leq_{\mathcal{A}} \beta$ | $f(\alpha \cup \beta) \equiv f(\alpha) \wedge f(\beta)$ |
| $f(\alpha \,\&\, \beta) \wedge o(\alpha) \equiv f(\beta) \wedge o(\alpha)$ | $o(\alpha) \vee o(\beta) \to o(\alpha \cup \beta)$ |
| $o(\alpha \cup \beta) \wedge f(\alpha) \to o(\beta)$ if $\alpha$, $\beta$ are both 1-uniform | $o(\alpha \,\&\, \beta) \equiv o(\alpha) \wedge o(\beta)$ if $[\![\alpha \,\&\, \beta]\!] \neq \emptyset$ |

$$\left.\begin{array}{l} f(\alpha;\beta) \equiv [\alpha]f(\beta) \\ f(\alpha) \vee f(\beta) \to f(\alpha \,\&\, \beta) \end{array}\right\} \text{ if } \{\alpha, \beta\} \text{ is uniform}$$

Table 3: The theory $PD_eL$.

We summarize the rules and axioms for our simplified $PD_eL$ in the top half of Table 3. This theory is sound but not complete (see Meyer (1988, 1989)).

Thus far, we have defined a variant of $PDL$, with no particular relevance for reasoning about prohibitions or obligations. The reduction of deontic operators to dynamic ones uses Anderson's violation atom $V$ (1967) to represent deontic violations. This yields deontic operators $f$ and $o$, representing that an action is prohibited/obligatory, resp., presented in Table 3 along with some prominent theorems.

Note that these definitions of prohibition and obligation are about *immediate* actions. Let us focus on prohibition for a moment. A world $w$ satisfies $f(\alpha)$ just in case *in* $w$, the result of doing $\alpha$ is a world in violation. But if our agent performs some other action first, say $\beta$, then he may no longer be in $w$ and so the fact that $w \models f(\alpha)$ is not relevant for him. In other words, $f(\alpha)$ expresses that an agent is prohibited from doing $\alpha$ *now*, not that he is prohibited from *ever* doing $\alpha$.

We close this section with some comments about one unfortunate consequence of this approach. It seems reasonable that, if $\alpha$ is forbidden, so is any action beginning with $\alpha$, i.e. $f(\alpha) \to f(\alpha;\beta)$. But this property does not hold in general. Indeed, it is easy to see that

$$\vdash (f(\epsilon) \to f(\epsilon; \mathbf{any}^*)) \equiv (V \to [\mathbf{any}^*]V).$$

Thus, if one wants $f(\alpha) \to f(\alpha;\beta)$ to hold in general, he must either give up the defining axiom for $[\alpha;\beta]$ or require $V \to [\mathbf{any}^*]V$. This is a very strong and usually undesirable condition which we briefly discuss in Section 3.1.

It may be argued that, in the end, a dynamic deontic logic indeed wants $f(\alpha) \to f(\alpha;\beta)$ and $\neg(V \to [\mathbf{any}^*]V)$. The natural way to satisfy this is to change the semantics to interpret $[\![\alpha]\!](w)$ as a set of paths and define: $w \models [\alpha]V$ just in case for each path

$$w \xrightarrow{x_1} w_1 \xrightarrow{x_2} \cdots \xrightarrow{x_n} w_n$$

in the interpretation of $\alpha$, there is a world $w_1 \models V$. Such an interpretation violates not only $[\alpha;\beta]\phi \equiv [\alpha][\beta]\phi$, but also the axiom K: $[\beta](\phi_1 \to \phi_2) \to ([\beta]\phi_1 \to [\beta]\phi_2)$. We postpone this alternative for later research.

# 3 The long-term contiguous system

We turn our attention now to long-term norms. The definitions of $f$ and $o$ from the previous section are intended to express immediate duties, but often one has duties of longer duration, such as: "Never do that", or, "Do this someday." In this section, we introduce the machinery to express such long-term norms.

## 3.1 The long-term, contiguous prohibition

As previously mentioned, the formula $f(\alpha)$ expresses one important kind of prohibition, namely, that the agent is prohibited from doing $\alpha$ in *this* world. But many prohibitions are stronger than this. They express that one is *never* allowed to perform a particular act[3], such as "Never point a loaded gun at an innocent man and pull the trigger." But we mean a particular interpretation of this "never". We do not mean "in every world, one should not point a loaded gun. . . ," but rather "in every world *reachable from the current world*, one should not point a loaded gun. . . "

Such prohibitions are easy to express in the logic at hand. One is *never* allowed to do $\alpha$ just in case, for any action $\beta$, doing $\beta$ followed by $\alpha$ results in violation. In other words, in world $w$, one is forever prohibited in doing $\alpha$ iff for all $\beta$,

$$\forall w' \in [\![\beta;\alpha]\!](w) \ . \ w' \models V.$$

[3]In practice, such prohibitions are likely to include a conditional, such as, "Always obey local traffic laws unless in an emergency." Admittedly, these conditional prohibitions are not expressible as $F(\alpha)$ in the sense given here, because we have omitted the test action constructor $\phi?$. But we did so only for simplicity's sake. The test constructor presents no particular difficulty for our logic.

But this is true just in case for every $w' \in [\![\mathbf{any}^*;\alpha]\!](w)$, we have $w' \models V$, i.e. just in case $w \models f(\mathbf{any}^*;\alpha)$. Thus, we define $F(\alpha) \equiv f(\mathbf{any}^*;\alpha)$.

We summarize our definition of $F$ and give a derived rule of inference and several theorems in Table 4. The derivations are straightforward.

| **Defining axiom** | **Rule of inference** |
|---|---|
| $F(\alpha) \equiv f(\mathbf{any}^*;\alpha)$ | $\dfrac{f(\alpha) \rightarrow f(\beta)}{F(\alpha) \rightarrow F(\beta)}$ |

**$PD_eL$ theorems for $F$**

| | |
|---|---|
| $F(\beta) \rightarrow F(\alpha)$ if $\alpha \leq_{\mathcal{A}} \mathbf{any}^*;\beta$ | $F(\alpha) \rightarrow f(\alpha)$ |
| $F(\beta) \rightarrow F(\alpha)$ if $\alpha \leq_{\mathcal{A}} \beta$ | $F(\alpha;\beta) \equiv [\mathbf{any}^*;\alpha]f(\beta)$ |
| | $F(\alpha;\mathbf{any}^*;\beta) \equiv [\mathbf{any}^*;\alpha]F(\beta)$ |
| $F(\alpha) \rightarrow F(\beta;\alpha)$ | $F(\alpha) \wedge F(\beta) \equiv F(\alpha \cup \beta)$ |
| $F(\alpha) \equiv F(\mathbf{any}^*;\alpha)$ | $F(\alpha) \vee F(\beta) \rightarrow F(\alpha \,\&\, \beta)$ if $\{\alpha,\beta\}$ is uniform |

Table 4: The definition of $F$ and several consequences.

## 3.2 The long-term, contiguous obligation

As we have seen, long-term prohibition is easy to express in $PD_eL$ and in fact requires only one small change to Meyer's 1989 presentation: the addition of the action $\mathbf{any}^*$. But the situation for long-term obligation is considerably more subtle.

Suppose one has promised to repay a loan, without any deadline regarding repayment. When can we conclude that he has failed to fulfill his duty? What behavior satisfies his obligation? One fulfills his obligation provided there is *some* point at which he actually repays the loan. Until then, he has an outstanding obligation.

One is naturally tempted to define $O$ in terms of $F$, just as $o$ was defined in terms of $f$. Unfortunately this yields unreasonable results: suppose $O(\underline{a}) \equiv F(\overline{\underline{a}})$. The latter is equivalent $[\mathbf{any}^*]f(\overline{\underline{a}})$ and hence to $[\mathbf{any}^*]o(\underline{a})$. But $[\mathbf{any}^*]o(\underline{a}) \rightarrow [\mathbf{any}^*;\underline{a}]o(\underline{a})$ and so this obligation can never be discharged! After doing $\underline{a}$, the agent must do it again and again and .... Clearly something is wrong here.

So what properties should $O$ capture? When we assert that $\alpha$ is a long-term obligation, we mean that we cannot discharge our obligation *without* first doing $\alpha$. This does not mean that we must do $\alpha$ *now*. It also does not mean that if we do $\alpha$, our obligation will be discharged: if we are indebted Paul, we have to acquire the necessary funds to repay him. But acquiring the funds is not enough to discharge our debt, of course: we must also repay Paul! Long term obligations are about necessary actions, rather than sufficient actions.

Thus, $O(\alpha)$ should be interpreted as: our obligations will not be discharged *unless* we do some act in $[\![\mathbf{any}^*;\alpha;\mathbf{any}^*]\!]$. That is, whatever fully specified $s$ we perform, if $s \notin [\![\mathbf{any}^*;\alpha;\mathbf{any}^*]\!]$, then the outcome will be a world in which we still have an unfulfilled duty. Consequently, we introduce a new action, $\widehat{\alpha}$ and interpret it as:

$$[\![\widehat{\alpha}]\!] = X^{<\omega} \setminus [\![\mathbf{any}^*;\alpha;\mathbf{any}^*]\!].$$

We also introduce a defining axiom scheme for $\widehat{\alpha}$ in Table 5 along with a few useful properties. We provide proofs for the axiom scheme and the property $\widehat{\alpha} \leq_{\mathcal{A}} \widehat{\alpha \,\&\, \beta}$, since these proofs are not as obvious as the others.

| | | |
|---|---|---|
| **Defining axiom** | | |
| $([\widehat{\alpha}]\varphi \wedge [\beta]\neg\varphi) \rightarrow ([\mathbf{any}^*; \alpha; \mathbf{any}^*]\psi \rightarrow [\beta]\psi)$ | | |
| **Properties for $\widehat{\alpha}$** | | |
| If $\alpha \leq_{\mathcal{A}} \beta$ then $\widehat{\beta} \leq_{\mathcal{A}} \widehat{\alpha}$ | $\widehat{\alpha} \leq_{\mathcal{A}} \widehat{\alpha \,\&\, \beta}$ | $\widehat{\beta} \leq_{\mathcal{A}} \widehat{\alpha \,\&\, \beta}$ |
| $\widehat{\alpha \cup \beta} \leq_{\mathcal{A}} \widehat{\alpha} \,\&\, \widehat{\beta}$ | $\widehat{\alpha} \leq_{\mathcal{A}} \widehat{\alpha; \beta}$ | $\widehat{\beta} \leq_{\mathcal{A}} \widehat{\alpha; \beta}$ |
| | $\widehat{\alpha} =_{\mathcal{A}} \widehat{\mathbf{any}^*; \alpha}$ | $\widehat{\alpha} =_{\mathcal{A}} \widehat{\alpha; \mathbf{any}^*}$ |

Table 5: Properties of the $\widehat{\alpha}$ constructor.

*Proof of defining axiom.* We aim to show that, for every world $w$, pair of actions $\alpha$, $\beta$ and pair of formulas $\phi$, $\psi$,

$$w \models ([\widehat{\alpha}]\varphi \wedge [\beta]\neg\varphi) \rightarrow ([\mathbf{any}^*; \alpha; \mathbf{any}^*]\psi \rightarrow [\beta]\psi).$$

Suppose that $w \models [\widehat{\alpha}]\varphi \wedge [\beta]\neg\varphi$. First, we will establish that, for every $s \in [\![\widehat{\alpha}]\!] \cap [\![\beta]\!]$, $[\![s]\!](w) = \emptyset$. Let such $s$ be given. Because $w \models [\widehat{\alpha}]\phi$, we see that $[\![s]\!](w) \subseteq [\![\varphi]\!]$. But since $w \models [\beta]\neg\phi$, we also see that $[\![s]\!](w) \subseteq [\![\neg\varphi]\!]$. Hence $[\![s]\!](w)$ is empty.

Now suppose that $w \models [\mathbf{any}^*; \alpha; \mathbf{any}^*]\psi$ and we will complete the proof by showing $w \models [\beta]\psi$. Let $s \in [\![\beta]\!]$ be given and we must show $[\![s]\!](w) \subseteq [\![\psi]\!]$. If $s \in [\![\mathbf{any}^*; \alpha; \mathbf{any}^*]\!]$ then $[\![s]\!](w) \subseteq [\![\psi]\!]$ by assumption. Otherwise, $[\![s]\!](w) = \emptyset$ and so is trivially contained in $[\![\psi]\!]$.

□

*Proof of $\widehat{\alpha} \leq_{\mathcal{A}} \widehat{\alpha \,\&\, \beta}$.* We will prove the claim by showing that

$$[\![\mathbf{any}^*; (\alpha \,\&\, \beta); \mathbf{any}^*]\!] \subseteq [\![\mathbf{any}^*; \alpha; \mathbf{any}^*]\!].$$

Let $s$ be an element of the set $[\![\mathbf{any}^*; (\alpha \,\&\, \beta); \mathbf{any}^*]\!]$. Then there are sequences $s_1$, $s_2$ and $s_3$ such that $s_2 \in [\![\alpha \,\&\, \beta]\!]$ and $s = s_1 * s_2 * s_3$.

By definition of $[\![\alpha \,\&\, \beta]\!]$, there is some $n$ such that $s_2 \restriction n \in [\![\alpha]\!]$. Thus, we can find $t_1$ and $t_2$ such that $s_2 = t_1 * t_2$ and $t_1 \in [\![\alpha]\!]$ (namely, we take $t_1 = s_2 \restriction n$). Then $s = s_1 * t_1 * (t_2 * s_3)$ and hence $s \in [\![\mathbf{any}^*; \alpha; \mathbf{any}^*]\!]$, as desired. □

Finally, there is one more characteristic difference between long-term obligation and prohibition. When we are obligated to pay a debt, say, or perform a promised act, then we have an unsatisfied duty. This is not the same as being in violation nor is there any obvious way of expressing this condition in terms of our violation predicate $V$. Rather, we should introduce a new atomic proposition $I$ (for *indebtedness*) to represent the condition that an agent has some unfulfilled obligation and define $O$ in terms of $I$. We investigate some possible relations between $I$ and $V$ below.

Thus $O(\alpha)$ represents that $\alpha$ is a necessary means to $\neg I$, i.e. that a $\neg I$ world will not be reached unless we do some sequence in $[\![\alpha]\!]$. Therefore, we propose to define $O(\alpha)$ by $[\widehat{\alpha}]I$.

We give a few simple theorems regarding $O$ in Table 6. The proofs are routine.

| |
|---|
| **Defining axiom** |
| $O(\alpha) \equiv [\widehat{\alpha}]I$ |
| $PD_eL$ **theorems for** $O$ |

| | |
|---|---|
| $O(\beta) \to O(\alpha)$ if $\widehat{\alpha} \leq_{\mathcal{A}} \widehat{\beta}$ | $O(\alpha \mathbin{\&} \beta) \to O(\alpha) \wedge O(\beta)$ |
| $O(\beta) \to O(\alpha)$ if $\beta \leq_{\mathcal{A}} \alpha$ | $O(\alpha;\beta) \to O(\alpha) \wedge O(\beta)$ |
| $O(\alpha) \equiv O(\mathbf{any}^*;\alpha)$ | $O(\alpha) \vee O(\beta) \to O(\alpha \cup \beta)$ |
| $O(\alpha) \equiv O(\alpha;\mathbf{any}^*)$ | |

Table 6: The definition of $O$ and some consequences.

There is an unfortunate consequence of this definition: if an agent is in a world in which $\neg I$ is unreachable, he is obligated to do *everything*, which is absurd. Thus, one may be tempted to amend the definition so that $O(\alpha)$ is defined as

$$[\widehat{\alpha}]I \wedge \langle \mathbf{any}^* \rangle \neg I.$$

We do not use this more complicated definition here, partly for simplicity's sake and partly for consistency with the prior definition of $o$. Also, the amended definition has its own motivational problem: an agent that cannot reach $\neg I$ is never obligated to do *anything*, which seems similarly absurd.

We have presented only the most basic and useful theorems for $O$ in Table 6, but this list can be extended in many natural directions. For instance, if one is obligated to do $\underline{a}$ and also $\underline{b}$, then he is obligated to do $\underline{a}$ and later $\underline{b}$ or $\underline{b}$ and later $\underline{a}$ or both at once, i.e.

$$O(\underline{a}) \wedge O(\underline{b}) \equiv O((\underline{a};\mathbf{any}^*;\underline{b}) \cup (\underline{b};\mathbf{any}^*;\underline{a}) \cup (\underline{a} \mathbin{\&} \underline{b})).$$

This is the long-term analogue of $o(\alpha \mathbin{\&} \beta) \equiv o(\alpha) \wedge o(\beta)$ and applies to any pair of 1-uniform actions.

Another intuitive example: one expects that, if an agent is obliged to eventually do $\alpha;\beta$, then after doing $\alpha$, he will still be obliged to do $\beta$. In fact, this is not quite the case, if $\alpha$ is sufficiently complex, but a similar claim *does* hold. For this, let us introduce a new action, $\ldots\alpha$.

Let $[\![\ldots\alpha]\!]$ be the set of sequences $s$ such that (a) $s$ ends in an $\alpha$-sequence, i.e. $s = t * t'$ where $t' \in [\![\alpha]\!]$ and (b) no proper prefix of $s$ ends in an $\alpha$-sequence. In other words, $[\![\ldots\alpha]\!] = [\![\mathbf{any}^*;\alpha]\!] \setminus [\![\mathbf{any}^*;\alpha;\mathbf{any};\mathbf{any}^*]\!]$. Then one can easily show:

$$\models O(\alpha;\beta) \to O(\alpha) \wedge [\ldots\alpha]O(\beta).$$

In other words, the agent obligated to do $\alpha;\beta$ is still obligated to do $\beta$ at the instant he has first completed $\alpha$. For atomic actions $\underline{a}$, it is easy to see that $\ldots\underline{a} = \widehat{\underline{a}};\underline{a}$, and thus

$$\models O(\underline{a};\beta) \to O(\underline{a}) \wedge [\widehat{\underline{a}};\underline{a}]O(\beta).$$

This formula is analogous to the formula $o(\alpha;\beta) \to o(\alpha) \wedge [\alpha]o(\beta)$, which is valid if $[\![\beta]\!] \neq \emptyset$.

Because $O$ is defined in terms of a new indebtedness proposition, we have lost the strong connection between prohibition and obligation. In fact, we think this is natural: long term obligations are not a simple conjugate of prohibition. They impose looser restrictions on behavior and cannot be captured in terms of $V$. Nonetheless, it would be natural to suppose some connection between $I$ and $V$ and we briefly consider a few proposals here.

Loosely, a moral agent aims to reach a world in which his obligation is relieved: he aims to realize $\neg I$. This is not quite accurate, however, since new obligations may be created before paying old ones. Borrowing from Peter to pay Paul relieves the obligation to Paul at the expense of creating a new obligation and thus remaining in a state of indebtedness. Nonetheless, this strategy is not obviously immoral (provided that Peter is eventually repaid, perhaps by borrowing from Paul), regardless of its practical merits. A moral agent may fulfill each obligation without ever reaching a $\neg I$-world!

On the other hand, perhaps one should avoid situations in which he can *never* fulfill his outstanding obligations without first acquiring new ones. In such situations, the agent has exceeded his ability to meet his duties. Thus, one may wish to relate $I$ and $V$ by requiring that our models satisfy the *axiom of eventual repayment*:

$$[\mathbf{any}^*]I \rightarrow V \qquad (\mathbf{ER})$$

Alternatively, we may wish to restrict the ways in which an agent discharges his obligations. One should not fulfill obligations by doing prohibited acts: it may be okay to borrow from Peter to pay Paul, but robbing Peter is out of bounds. The natural way to restrict such disreputable strategies is the converse of *eventual repayment*, which we call the *axiom of forbidden means*:

$$V \rightarrow [\mathbf{any}^*]I \qquad (\mathbf{FM})$$

Thus, if one *ever* reaches a world in violation, he is thereafter in an $I$-world. With this axiom, one can prove

$$O(\alpha \cup \beta) \wedge F(\alpha) \rightarrow O(\beta).$$

(The proof depends on a proof of the abstruse property $\widehat{\alpha \cup \beta} \cup (\mathbf{any}^*; \alpha; \mathbf{any}^*) =_{\mathcal{A}} \widehat{\beta} \cup (\mathbf{any}^*; \alpha; \mathbf{any}^*)$, but is otherwise straightforward.)

One may adopt *both* of the above axioms. In this case, $[\mathbf{any}^*]I$ is equivalent to $V$. But this means that, once in a $V$ world, every path leads to a $[\mathbf{any}^*]I$ world, and hence to another $V$ world. Consequently, such models satisfy the *axiom of unforgiveness*:

$$V \equiv [\mathbf{any}^*]V. \qquad (\mathbf{UF})$$

Once an agent is in violation, he remains there. This is an unforgiving model of deontic logic!

# 4 The non-contiguous system

In the previous section, we explored long-term prohibitions and obligations of certain kinds of actions, namely *contiguous* actions. The prohibition $F(\alpha; \beta)$ expresses that one is never allowed to do $\alpha$ *immediately* followed by $\beta$, but it does not restrict one from doing $\alpha$, then something else and then $\beta$.

It seems reasonable that most prohibitions do involve such contiguous actions. One is not allowed to aim the gun at an innocent and pull the trigger, but he can aim the gun at an innocent[4], then point it at the ground and pull the trigger. The effect of aiming the gun can be undone before pulling the trigger.

[4] Let us briefly ignore very sensible rules regarding gun handling and safety and the very disturbing effect caused by staring down the wrong end of a gun barrel.

But in rare situations, the effects of doing $\alpha$ cannot be undone and thus one should *never* do $\beta$ thereafter. Suppose that a big yellow button arms a bomb and that, once armed, it cannot be disarmed. Suppose also that a big red button detonates the bomb if it is armed. Then one should never press the yellow button followed eventually by pressing the red button. We denote this kind of prohibition by $F^*$.

Admittedly, such strong prohibitions tend to be as artificial as our bomb example, but we claim that long-term non-contiguous *obligations* are fairly common. We will discuss these in Section 4.1, but let us first examine prohibitions.

We must introduce a few relations on sequences and actions in order to express non-contiguous prohibitions. The first relation, $s \sqsubseteq t$, expresses that $s$ is a subsequence of $t$ and that the last element of $s$ is the last element of $t$. Explicitly, $r \sqsubseteq s$ iff there is a $f : |r| \to |s|$ satisfying the following:

- $f$ is strictly increasing;
- $f(|r| - 1) = |s| - 1$;
- for every $i < |r|$, we have $s(f(i)) = r(i)$.

In other words, $r \sqsubseteq s$ iff $r$ is a subsequence of $s$ such that the last element of $r$ is also the last element of $s$. We say in this case that $r$ is a *tail-fixed subsequence* of $s$. See Figure 3 for an illustration.

**Properties of $\sqsubseteq$**

1. $\sqsubseteq$ is a partial order.
2. If $s_1 \sqsubseteq t_1$ and $s_2 \sqsubseteq t_2$ then $s_1 * s_2 \sqsubseteq t_1 * t_2$.
3. If $s_1 * s_2 \sqsubseteq t$, then there are $t_1$, $t_2$ such that $t = t_1 * t_2$, $s_1 \sqsubseteq t_1$ and $s_2 \sqsubseteq t_2$.
4. If $s \sqsubseteq t$ then $s \sqsubseteq r * t$ for any $r$.
5. If $s_1 * s_2 \sqsubseteq t$ then $s_2 \sqsubseteq t$.
6. $\langle x \rangle \sqsubseteq \langle x_1, \ldots, x_n \rangle$ iff $x = x_n$.

**Definition of $\widetilde{\alpha}$**

$$[\![\widetilde{\alpha}]\!] = \{\, s \in X^{<\omega} \;\mid\; \exists r \in [\![\alpha]\!] \,.\, r \sqsubseteq s \,\}$$

**Properties of $\widetilde{\alpha}$**

| | |
|---|---|
| $\alpha \leq_{\mathcal{A}} \widetilde{\alpha}$ | $\widetilde{\alpha;\beta} =_{\mathcal{A}} \widetilde{\alpha};\widetilde{\beta}$ |
| If $\alpha \leq_{\mathcal{A}} \beta$ then $\widetilde{\alpha} \leq_{\mathcal{A}} \widetilde{\beta}$ | $\widetilde{\alpha \cup \beta} =_{\mathcal{A}} \widetilde{\alpha} \cup \widetilde{\beta}$ |
| $\widetilde{\alpha;\beta} \leq_{\mathcal{A}} \widetilde{\beta}$ | $\widetilde{\underline{a}} =_{\mathcal{A}} \mathbf{any}^*;\underline{a}$ |
| $\widetilde{\widetilde{\beta}} =_{\mathcal{A}} \widetilde{\beta}$ | $\mathbf{any}^* =_{\mathcal{A}} \widetilde{\mathbf{any}^*}$ |

Table 7: Properties of $\sqsubseteq$ and $\widetilde{\alpha}$.

We can express the intended meaning of long-term non-contiguous prohibition in terms of $\sqsubseteq$. Suppose that $F^*(\beta)$ and $s \in [\![\beta]\!]$. Then any fully specified $t$ satisfying $s \sqsubseteq t$ will lead to violation. In order to express this prohibition on the level of action terms,

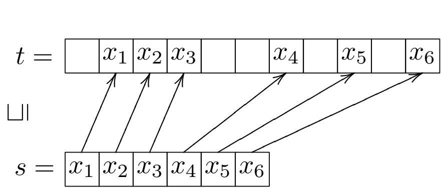


Figure 3: Illustration of $s \sqsubseteq t$.

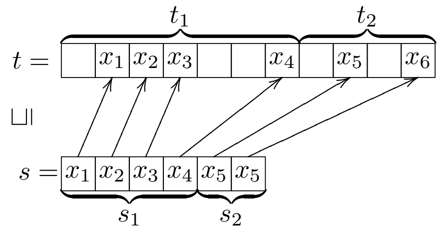


Figure 4: Illustration for $\sqsubseteq$ properties (2) and (3).

we introduce an action constructor $\widetilde{\beta}$ defined by

$$[\![\widetilde{\beta}]\!] = \{\, s \in X^{<\omega} \;\mid\; \exists r \in [\![\beta]\!] \;.\; r \sqsubseteq s \,\}.$$

Hence, $\alpha \leq_{\mathcal{A}} \widetilde{\beta}$ iff each $s$ in $[\![\alpha]\!]$ has some $r$ in $[\![\beta]\!]$ as a tail-fixed subsequence. Thus, if $w \models F^*(\beta)$ and $\alpha \leq_{\mathcal{A}} \widetilde{\beta}$, then $w \models f(\alpha)$.

In case $\alpha \leq_{\mathcal{A}} \widetilde{\beta}$, we say that $\alpha$ *involves* $\beta$. In this case, however one does $\alpha$ (whichever fully specified sequence is chosen), one is doing $\beta$ "along the way", i.e. doing some $t \in [\![\beta]\!]$ as a tail-fixed subsequence. For example, $\alpha$ involves $\underline{a}$ just in case each $s \in [\![\alpha]\!]$ has as last element some element of $i(a)$.

We summarize properties of $\sqsubseteq$ and $\widetilde{\alpha}$ in Table 7.

## 4.1 Long-term non-contiguous prohibition and obligation

| | |
|---|---|
| **Defining axioms** | |
| $F^*(\alpha) \equiv [\widetilde{\alpha}]V$ | $O^*(\alpha) = [\widehat{\widetilde{\alpha}}]I$ |
| $PD_eL$ **theorems for** $F^*$ | |
| $[\widetilde{\beta}]\phi \to [\alpha]\phi$ if $\alpha \leq_{\mathcal{A}} \widetilde{\beta}$ | $F^*(\alpha) \to F(\alpha)$ |
| $F^*(\beta) \to F^*(\alpha)$ if $\alpha \leq_{\mathcal{A}} \widetilde{\beta}$ | $F(\underline{a}) \equiv F^*(\underline{a})$ |
| $F^*(\beta) \to F^*(\alpha)$ if $\alpha \leq_{\mathcal{A}} \beta$ | $F^*(\alpha; \underline{a}) \equiv [\widetilde{\alpha}]F(\underline{a})$ |
| $F^*(\underline{a}; \alpha) \equiv [\mathbf{any}^*; \underline{a}]F^*(\alpha)$ | $F^*(\alpha; \beta) \equiv [\widetilde{\alpha}]F^*(\beta)$ |
| $F^*(\alpha \cup \beta) \equiv F^*(\alpha) \wedge F^*(\beta)$ | $F^*(\alpha) \vee F^*(\beta) \to F^*(\alpha \mathbin{\&} \beta)$ |
| $F^*(\beta) \to F^*(\alpha; \beta)$ | |
| $PD_eL$ **theorems for** $O^*$ | |
| $O^*(\alpha) \to O^*(\beta)$ if $\widetilde{\alpha} \leq_{\mathcal{A}} \widetilde{\beta}$ | $O(\alpha) \to O^*(\alpha)$ |
| $O^*(\alpha) \to O^*(\beta)$ if $\alpha \leq_{\mathcal{A}} \beta$ | $O^*(\alpha \mathbin{\&} \beta) \to O^*(\alpha) \wedge O^*(\beta)$ |
| $O^*(\alpha; \beta) \to O^*(\alpha) \wedge O^*(\beta)$ | $O^*(\alpha) \vee O^*(\beta) \to O^*(\alpha \cup \beta)$ |
| $O^*(\alpha) \equiv O^*(\mathbf{any}^*; \alpha)$ | $O^*(\underline{a}) \equiv O(\underline{a})$ |
| $O^*(\alpha) \equiv O^*(\alpha; \mathbf{any}^*)$ | |

Table 8: The long-term non-contiguous prohibition operator $F^*$.

We have already tipped our hand regarding the long-term non-contiguous prohibition $F^*$. When we say that $\beta$ is forbidden in this sense, we mean that each fully specified sequence $s$ containing some $t \in [\![\beta]\!]$ as a tail-fixed subsequence leads to violation. In other words, we define $F^*(\beta) \equiv f(\widetilde{\beta})$, equivalently $F^*(\beta) \equiv [\widetilde{\beta}]V$.

With this definition, one can easily derive $F^*(\beta) \to f(\alpha)$ whenever $\alpha \leq_{\mathcal{A}} \widetilde{\beta}$. In fact we can derive a stronger consequence in this case. If $\alpha \leq_{\mathcal{A}} \widetilde{\beta}$ then $\widetilde{\alpha} \leq_{\mathcal{A}} \widetilde{\widetilde{\beta}} =_{\mathcal{A}} \widetilde{\beta}$, so $\vdash F^*(\beta) \to F^*(\alpha)$.

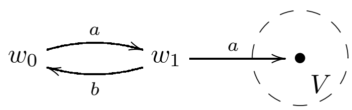


Figure 5: A counterexample: $w_1$ satisfies $f(\underline{a}) \wedge \neg F(\underline{a})$ and $w_0$ satisfies $F(\underline{a};\underline{a}) \wedge \neg F^*(\underline{a};\underline{a})$.

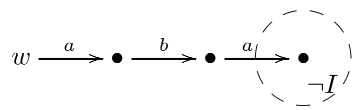


Figure 6: A counterexample: The world $w$ satisfies $O^*(\underline{a};\underline{a})$ but not $O(\underline{a};\underline{a})$.

This and other properties of $F^*$ are presented in Table 8. Again, the proofs are straightforward from the properties stated previously. We can also see now that the three prohibition operators are comparable, with $F^*$ the strongest and $f$ the weakest, since $F^*(\alpha) \to F(\alpha)$ and $F(\alpha) \to f(\alpha)$.

The converse implications are not valid, with a counterexample given in Figure 5 (where $X = A$ and $i : A \to \mathcal{P}^+A$ is the singleton map $a \mapsto \{a\}$). World $w_1$ satisfies $f(\underline{a})$ but not $F(\underline{a})$ since $[\![\langle b, a\rangle]\!](w_1) \not\subseteq [\![V]\!]$. World $w_0$ satisfies $F(\underline{a};\underline{a})$ but not $F^*(\underline{a};\underline{a})$, since $[\![\langle a, b, a\rangle]\!] \not\subseteq [\![V]\!]$.

As we have admitted, long-term non-contiguous prohibitions may be fairly rare, since they involve actions with effects that cannot be undone. However, it seems that non-contiguous *obligations* are fairly common. Suppose that Peter owes Paul five dollars, but does not have five dollars. Then he is obliged to first acquire five dollars (or more) and then repay it. But he does not have to repay the money immediately after acquiring it. Rather, he is free to do other things in between. Of course, if he loses the money in between acquiring and repayment, then he cannot discharge his obligation — but we are interested here in necessity rather than sufficiency, and it is necessary that he first acquires and some time later repays.

This suggests the definition $O^*(\alpha) \equiv O(\widetilde{\alpha})$, equivalently $O^*(\alpha) \equiv [\widehat{\widetilde{\alpha}}]I$. Properties for $O^*$ can be found in Table 8. The consequence relation between $O$ and $O^*$ is dual to that between $F$ and $F^*$, namely $O(\alpha) \to O^*(\alpha)$. Again, the converse does not hold, as indicated in Figure 6.

Unfortunately, there is no simple relation between $O$ and $o$. It is clearly not the case that $o(\alpha) \not\to O(\alpha)$, but this is not too surprising, since the motivation for $o$ (avoiding violation) is different than for $O$ (eventually reaching $\neg I$). It is not hard to show that, for 1-uniform $\alpha$, the axiom (**FM**) proves $(o(\alpha) \wedge I) \to O(\alpha)$, but a tighter relationship eludes us.

# 5 Concluding remarks

Meyer's work on $PD_eL$ has contributed a formal logic for certain kinds of obligation and prohibition, namely, the *immediate* kind. One of the nice features of his approach is that the two normative concepts are inter-definable: obligation is the same as prohibition from refraining. We aimed to extend his work to include duties of wider scope, duties to never do $\alpha$ or to eventually do $\alpha$. As we have seen, however, the natural duality between obligation and prohibition has become obscured by our possible world semantics. Obligations are violated only in the limit, and this is not expressible in terms of worlds reached along the way.

Our work is an extension of an existing framework for deontic logic to include new normative expressions. But we also believe it suggests a new direction for dynamic deontic logic. We would like to recover the duality between prohibitions and obligations that seems so natural in the immediate case. To do so, one needs to evaluate actions in terms of infinite $X$-sequences and $\mathcal{W}$-paths rather than the worlds encountered at the end of finite paths. In this conceptual setting, it makes sense to discuss failure to meet obligations (i.e., *never* doing what is required) and adherence to long-term prohibitions (i.e., *never* doing what is forbidden). Moreover, we believe that the topological approach of learning theory gives a natural framework for investigating these infinite paths. We hope to return to this topic in future work.

We also believe that some of our considerations provide argument for a hybrid of dynamic and propositional deontic logic. In Section 4.1, we discussed the obligation to obtain funds in order to repay one's debt. But why does a debt impose an obligation to obtain money? Because having money is a necessary precondition for repaying the debt and obtaining money is a means to realize this precondition. It is natural to discuss both ought-to-do and ought-to-be in explaining derivative obligations like the obligation to obtain money. We would like a single framework that includes dynamic operators for both actions and conditions and that allows for reasoning about derived obligations and prohibitions. This would allow for new constructions like, "while $\phi$, do $\alpha$." We expect that existing work on agent planning would be relevant for this project.

## References


Anderson, A.R. 1967. Some nasty problems in the formalization of ethics. *Noûs* 1: 345–360.

Castañeda, H.-N. 1981. The paradoxes of deontic logic. In *New studies in deontic logic*, ed. R. Hilpinen, 37–85. Dordrecht: Reidel.

Harel, David. 1984. Dynamic logic. In *Handbook of philosophical logic*, ed. D. Gabbay and F. Guenthner, vol. II, 497–604. D. Reidel Publishing Company.

Kelly, Kevin. 1996. *The logic of reliable inquiry.* Logic and Computation in Philosophy, Oxford University Press.

Meyer, J.-J. Ch. 1988. A different approach to deontic logic: Deontic logic viewed as a variant of dynamic logic. *Notre Dame Journal of Formal Logic* 29:106–136.

———. 1989. Using programming concepts in deontic reasoning. In *Semantics and contextual expression*, ed. R. Bartsch, J.F.A.K. van Benthem, and P. van Emde Boas, 117–145. Dordrecht/Riverton: FORIS publications.

———. 2000. Dynamic logic for reasoning about actions and agents. In *Logic-based artificial intelligence*, 281–311. Norwell, MA, USA: Kluwer Academic Publishers.

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