Annals of Mathematics and Artificial Intelligence 9(1993)1–23 

# Conditional obligation, deontic paradoxes, and the logic of agency

Paul Bartha
*Department of Philosophy, University of Pittsburgh, Pittsburgh, PA 15260, USA*

## Abstract

A variant of Belnap's stit-semantics due to Horty and von Kutschera is used to develop a semantics of obligation. A partial completeness result is stated. The semantics is then used to discuss conditional obligation as well as two paradoxes of deontic logic. The paper argues for the importance of an analysis of agency for deontic logic.

## 1. Introduction

Consider the following argument, adapted from Castañeda [6]:

(a) Alchourrón is obligated to do the following: if Bulygin sends him the draft of their latest joint paper, revise it.

(b) Bulygin has sent Alchourrón the draft of their latest joint paper.

Therefore,

(c) Alchourrón is obligated to revise the draft.

Castañeda points out that this straightforward reasoning cannot be accommodated by most existing deontic calculi.

To see why, let A stand for "Alchourrón revises the draft" and B for "Bulygin sends Alchourrón the draft". Then we can represent the argument symbolically:

$$\begin{array}{ll} \text{(a)} & O(B \supset A) \\ \text{(b)} & \underline{B\qquad\qquad} \\ \text{(c)} & O(A) \end{array} \tag{1}$$

Here, O stands for the obligation operator. The proposition

$$O(B \supset A) \supset (B \supset OA),$$

which would allow detachment of the obligation OA in (1), does not belong to the standard system of deontic logic.[1] Consequently, the argument is invalid.

The argument becomes valid if we replace $O(B \supset A)$ by $B \supset OA$ in (a). No corresponding change is required in English. The English version of the argument seems acceptable as stated. This suggests that English is less sensitive than standard deontic logics to the difference between the two forms $O(B \supset A)$ and $B \supset OA$. These two forms will be referred to as $O \supset$-statements and $\supset O$-statements, respectively.[2]

Castañeda has developed one approach that brings the reasoning of deontic logic closer to that of English. He has suggested [6] that the two forms of obligation are equivalent under one assumption. Specifically, he has proposed that the rule

$$O(B \supset A) \Leftrightarrow B \supset OA$$

is valid whenever B is a "circumstance or condition" and A is "an action deontically considered as the focus of obligatoriness". He develops a calculus which takes the "circumstance/action as focus" distinction as primitive.

This paper takes a different approach, starting from the equally important distinction between agentive and non-agentive sentences. Belnap [2] has formalized this distinction by providing a semantics for the agentive construction "$\alpha$ sees to it that A" (written as [$\alpha$ *stit:* A]). In this paper, a simple semantics of obligation is developed as an extension of Belnap's theory of agency. The basic idea is inspired by Anderson's reduction of deontic logic to alethic modal logic [1]. Sections 2 and 3 explain the semantics of *stit* and obligation, respectively. The resulting concept of obligation is compared to other systems of deontic logic. A partial completeness result is described in section 4.

The semantical system is then used to analyze conditional obligation. In particular, as sections 5 and 6 show, it provides a precise way to define a "circumstance" such that the argument (1) becomes valid under the assumption that B is a circumstance. In the remainder of the paper, the semantics is used to shed light on two paradoxes of deontic logic.

The point is not to give a final solution to any paradox or problem. Rather, we hope to show that many of the problems of deontic logic are essentially problems about agency rather than obligation. We also hope that readers will see that *stit* theory can be a useful tool in thinking about such problems.

## 2. Semantics of *dstit*

The semantics of [$\alpha$ *stit:* A], where $\alpha$ is an agent and A a sentence, are defined by Belnap in [2] and [4]. A similar notion, now known as *dstit*, was

[1] Føllesdal and Hilpinen [9] set out the axioms for the "standard system of deontic logic", known as KD, in which $O(B \supset A) \supset (B \supset OA)$ does not hold. This proposition is usually rejected in any system which is based on "deontically perfect world" semantics; see Hintikka [11].

[2] This terminology was suggested by Nuel Belnap. There is no standard way of referring to the two forms.

developed independently by von Kutschera [17] and Horty. Belnap's *stit* sematics tries to capture "the present fact that A is guaranteed by a prior choice of $\alpha$".[3] *dstit*, by contrast, attempts to represent the fact that A is guaranteed by a present choice of $\alpha$.[4] The following is a brief outline of the semantics of *dstit*.[5]

Formulas in our language are constructed from propositional variables by truth-functional connectives $\neg$ and $\wedge$, as well as modal operators $\Box$, $\mathcal{L}$, $\boldsymbol{F}$, $\boldsymbol{P}$, and $[\alpha\, dstit{:}\,]$. As usual, $\vee$, $\supset$, $\equiv$, $\top$ and $\bot$ are introduced as abbreviations. We use A, B, etc. to range over formulas.

The fundamental notions for the semantics of *dstit* are *moment*, *history*, *agent*, *choice set*, and *possible choice*. Time is modeled as a tree-like set of *moments* partially ordered by earlier/later. Time branches towards the future, but not into the past. Additionally, we assume *historical connection*: any two moments have a lower bound. This "tree" picture of time goes back to Prior, who attributes it to Ockham.[6] Formally, T is a nonempty set with a partial order $\leq$ subject to *no downward branching*, $\forall m \forall m' \forall m''(m' \leq m \wedge m'' \leq m \rightarrow m' \leq m'' \vee m'' \leq m')$, and *historical connection*, $\forall m \forall m' \exists m''(m'' \leq m \wedge m'' \leq m')$. We define $m < m'$ iff $m \leq m'$ and $m \neq m'$.

A *history* in T is a maximal chain of moments, i.e., a maximal branch of the tree. Two histories $h$ and $h'$ are *undivided at* $m$ (written $h \equiv_m h'$) iff $\exists m'(m < m' \wedge m' \in h \cap h')$. The *no backward branching* condition together with *historical connection* implies that any two distinct histories share an initial segment, divide at one moment, and remain separated from then on.[7]

$\overline{\alpha}$, a primitive, is a set of individual agents denoted by $\alpha$, $\alpha'$, $\beta$, etc. For each moment $m$, set $H_{(m)} = \{h: m \in h\}$, the set of all histories containing (passing through) $m$. A *choice set* for $\alpha$ at a moment $m$ is a partition of $H_{(m)}$; we write $\mathbf{Choice}_\alpha(m)$ for this partition. A *possible choice* for $\alpha$ is any member of this partition (i.e. a set of histories). The partition function is subject to the restriction that there can be *no choice between undivided histories*, i.e. $\forall h \forall h' \forall H(h \equiv_m h' \wedge H \in \mathbf{Choice}_\alpha(m) \rightarrow (h \in H \leftrightarrow h' \in H))$. Two histories $h$, $h'$ belonging to the same possible choice for $\alpha$ at $m$ are called *choice-equivalent for* $\alpha$ *at* $m$ or *choice*$_\alpha(m)$*-equivalent* (written $h' \equiv_m^\alpha h$); no choice that $\alpha$ is able to make at $m$ can tell them apart.

We call a *frame* any quadruple $\mathbf{F} = \langle \mathbf{T}, \leq, \overline{\alpha}, \mathbf{Choice} \rangle$ with components satisfying the above conditions. Following Prior, we evaluate truth in such a structure relative to moment–history pairs. A *model* M on F is a pair $\langle \mathbf{F}, \mathbf{V} \rangle$, where F is a frame and V is a valuation such that for each propositional variable $p$, $\mathbf{V}(p)$ is a subset of $\{(m, h)/m \in h\}$. Then for any formula A, we define the truth of A at moment–history pair $(m, h)$ in M, written $\mathbf{M} \models \mathrm{A}[m, h]$, as follows:

[3] This is Belnap's description in [2].

[4] A comparison of the two notions is not relevant here. See [2], notes 11 and 16, for discussion. The added 'd' in *dstit* stands for 'deliberatively'.

[5] This section draws on Thomason [14,15], Chellas [7], Belnap [2,4], and Xu [18].

[6] Prior's original presentation of the Ockhamist tense logic is found in [13]. Thomason [16] provides a good account.

[7] Provided suprema of all subsets of T exist in T. This assumption is not needed in what follows.

$\mathbf{M} \models p[m, h]$ *iff* $(m, h) \in \mathbf{V}(p)$, for propositional variable $p$;

$\mathbf{M} \models \neg \mathrm{A}[m, h]$ *iff* $\mathbf{M} \not\models \mathrm{A}[m, h]$ (not $\mathbf{M} \models \mathrm{A}[m, h]$);

$\mathbf{M} \models \mathrm{A} \wedge \mathrm{B}[m, h]$ *iff* $\mathbf{M} \models \mathrm{A}[m, h]$ and $\mathbf{M} \models \mathrm{B}[m, h]$;

$\mathbf{M} \models \boldsymbol{P}\mathrm{A}[m, h]$ *iff* $\exists m' < m$ $(\mathbf{M} \models \mathrm{A}[m', h])$;

$\mathbf{M} \models \boldsymbol{F}\mathrm{A}[m, h]$ *iff* $\exists m' > m$ $(\mathbf{M} \models \mathrm{A}[m', h])$;

$\mathbf{M} \models \Box \mathrm{A}[m, h]$ *iff* $\mathbf{M} \models \mathrm{A}[m', h']$ for all $m'$, and all $h' \in H_{(m')}$;

$\mathbf{M} \models \mathcal{L}\mathrm{A}[m, h]$ *iff* $\mathbf{M} \models \mathrm{A}[m, h']$ for all $h' \in m$.

$\boldsymbol{P}$ and $\boldsymbol{F}$ are the operators for *past* and *future*. $\Box$A is read as *necessarily* A. $\mathcal{L}$A is read as A *is settled* (*at a particular moment*). The scope of $\Box$ is all histories and all moments, whereas the scope of $\mathcal{L}$ is only the histories through one moment.[8)]

Finally, two conditions must be satisfied in order to have $\mathbf{M} \models [\alpha\ dstit:\ \mathrm{A}]\,[m, h]$:

(1) *Positive condition.* $\mathbf{M} \models \mathrm{A}[m, h']$ for all $h'$ with $h' \equiv_m^{\alpha} h$. (By making the possible choice containing history $h$, $\alpha$ guarantees that A is true – since A holds on all histories consistent with $\alpha$'s choice.)

(2) *Negative condition.* $\mathbf{M} \models \neg \mathrm{A}[m, h'']$ for some $h''$ with $m \in h''$; i.e. it is not the case that $\mathbf{M} \models \mathcal{L}\mathrm{A}[m, h]$. The moment–history pair $(m, h'')$ is called a *counter*. (Thus, $\alpha$ has a *real* choice about A, since it is not the case that A is settled true regardless of what $\alpha$ does.)

The basic picture is shown in fig. 1.

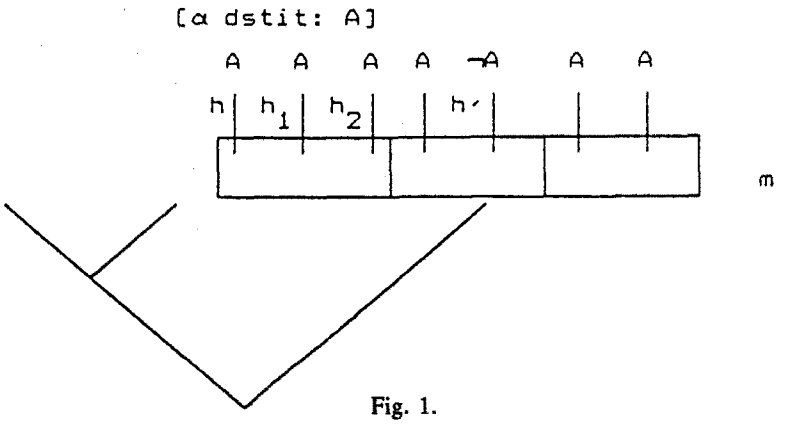


Fig. 1.

[8)] This conception of settledness is due to Prior. Both Prior and Thomason often use $\Box$ to represent settledness (see, for example, [16]). Here, we reserve $\Box$ for quantification over all moments, which we wish to distinguish explicitly from settledness at one moment.

In fig. 1, the moment $m$ is a node in the tree which has been blown up to reveal the choice structure. Each of the three boxes represents a possible choice of $\alpha$ at $m$, i.e. a set of histories through $m$, such that every such history goes through exactly one box. Sometimes, a possible choice will be referred to as a choice box. An important convention is that different histories shown coming out of the same choice box need not necessarily split at $m$. They may split later. Note that $[\alpha\ dstit:\ \mathrm{A}]$ fails at $(m, h)$ precisely when either A is false at $(m, h')$ for some $h' \equiv^{\alpha}_{m} h$, or A is settled true at $m$.

We define the *validity of A for* $\mathbf{M}$, $\mathbf{M} \models \mathrm{A}$, as $\mathbf{M} \models \mathrm{A}[m, h]$ for all $m \in \mathbf{T}$ and all $h \in H_{(m)}$; evidently, $\mathbf{M} \models \mathrm{A}$ *iff* $\mathbf{M} \models \Box\mathrm{A}[m, h]$ for some $(m, h)$. A *is valid for* $\mathbf{F}$, $\mathbf{F} \models \mathrm{A}$, *iff* $\mathbf{M} \models \mathrm{A}$ for all models $\mathbf{M}$ on $\mathbf{F}$. Finally, we write $\models \mathrm{A}$ *iff* $\mathbf{M} \models \mathrm{A}$ for all models $\mathbf{M}$.

To facilitate the presentation of a partial completeness result (section 4), it is useful to introduce Chellas' modal operator $\Delta\alpha$ [7]. We could add it to our language and provide the following truth definition:

$$\mathbf{M} \models \Delta\alpha\mathrm{A}[m, h] \quad \textit{iff} \quad \mathbf{M} \models \mathrm{A}[m, h'] \text{ for all } h' \text{ with } h' \equiv^{\alpha}_{m} h. \qquad (*)$$

Evidently, $\Delta\alpha\mathrm{A}$ corresponds precisely to the positive condition for $[\alpha\ dstit:\ \mathrm{A}]$. In fact, if $\Delta\alpha\mathrm{A}[m, h]$ holds on $\mathbf{M}$, then either A is settled true at $m$, or the negative condition is satisfied and $[a\ dstit:\ \mathrm{A}]$ must be true at $(m, h)$. We make use of these observations to introduce $\Delta\alpha$ instead as an abbreviation:

$$\Delta\alpha\mathrm{A} =_{\mathrm{df}} \mathcal{L}\mathrm{A} \vee [\alpha\ dstit:\ \mathrm{A}]. \qquad (2)$$

It is easy to see that, so defined, $\Delta\alpha$ agrees with the semantic condition (*). It is also easily verified that any one of $\Delta\alpha$, $\mathcal{L}$, and $[\alpha\ dstit:\ ]$ could be defined in terms of the other two operators, since we have as valid formulas:

**T1** $[\alpha\ dstit:\ \mathrm{A}] \leftrightarrow \Delta\alpha\mathrm{A} \wedge \neg\mathcal{L}\mathrm{A}$

and

**T2** $\mathcal{L}\mathrm{A} \leftrightarrow \Delta\alpha\mathrm{A} \wedge \neg[\alpha\ dstit:\ \mathrm{A}]$.

It should also be noted that the three operators $\Box$, $\mathcal{L}$, and $\Delta\alpha$, by their truth definitions, are just like the modal operator in **S5** modal logic.

## 3. Semantics of obligation

For convenience in this and following sections, we will abbreviate $[\alpha\ dstit:\ \mathrm{A}]$ as $[\alpha\colon \mathrm{A}]$.

### 3.1. DEFINITION OF OBLIGATION OPERATOR

Belnap and Perloff [5] have claimed that deontic logic should be treated as an extension of a modal logic of agency. Their "restricted complement thesis"

requires that deontic constructions take agentive sentences as complements: in a sentence Op, p must be (or be equivalent to) a *stit* sentence. A justification for this claim is that practical obligations, "ought-to-do"'s, should be connected to a specific action by a specific agent. Regardless of whether the restricted complement thesis is correct, we feel that "ought-to-do"'s are a good place to start. This section develops a semantics for deliberative obligations – obligations binding on an agent at the moment when he makes a choice.[9]

Anderson [1] suggested the following reduction of deontic to alethic modal logic. Let $\mathcal{S}$ be a constant proposition which we call a *sanction*. We exploit the connection between obligation and sanction by defining

$$\mathrm{O}p \Leftrightarrow \Box(\neg p \supset \mathcal{S}).$$

The $\Box$ ensures that the implication is strict. Here, p is an arbitrary proposition.

We will use something similar to define the obligation operator. We will restrict our attention to the case where p is a *stit* sentence. Three questions arise:

(1) What is the appropriate interpretation of the propositional constant $\mathcal{S}$ added to our language?

(2) What should the scope of the modal operator be?

(3) How should we deal with the negation of p where p is a *stit* sentence?

(1) The intended meaning of $\mathcal{S}$ is unambitious – something such as "there is wrongdoing", or "there is a violation of the rules". The connection between obligation and $\mathcal{S}$ should be unproblematic. Even though the paper sometimes speaks of $\mathcal{S}$ as a sanction, we are not entitled to interpret it as punishment or censure, which has no logical connection to obligation.

So as to avoid confusing the obligations of different agents, $\mathcal{S}$ should be indexed by agent. Then we can interpret $\mathcal{S}_\alpha$ as "$\alpha$ does something wrong". We suppress the subscript, however, since throughout the paper we will only be concerned with one agent's obligations at a time.[10]

(2) The scope of the modal operator will be all histories through a given moment. We will replace $\Box$ with $\mathcal{L}$. There are two reasons for this. First, "failing to see to it that A" may be a case of wrongdoing at some moments, but not at others. We do not want to limit ourselves to obligations which remain constant for all time. $\Box$ is unsuitable. Second, the truth of $[\alpha\colon \mathrm{A}]$ involves consideration of all histories through a moment (by the negative condition), so it is not unreasonable to suppose that the truth of $\mathrm{O}[\alpha\colon \mathrm{A}]$ does as well.[11]

[9] See Thomason [14] for the contrast between deliberative and judgemental obligation.

[10] We might also add a second index $\Gamma$ for the *authority* (individuals or perhaps institutions) whose rules are violated. Again, this added complexity is not required at present.

[11] In section 9, we will discuss some difficulties that arise from the choice of $\mathcal{L}$. Another alternative is to use the Chellas operator $\Delta\alpha$. It turns out that this is a bad choice, since the resulting definition of obligation makes the sentence $[\alpha\colon \mathrm{A}] \supset \mathrm{O}[\alpha\colon \mathrm{A}]$ valid in all models.

(3) With regard to the negation of $[\alpha\colon A]$, there are three evident possibilities, leading to three alternative definitions of $O[\alpha\colon A]$:

$$O[\alpha\colon A] \Leftrightarrow \mathcal{L}(\neg[\alpha\colon A] \supset \mathcal{S}), \tag{3a}$$

$$O[\alpha\colon A] \Leftrightarrow \mathcal{L}([\alpha\colon \neg[\alpha\colon A]] \supset \mathcal{S}), \tag{3b}$$

$$O[\alpha\colon A] \Leftrightarrow \mathcal{L}([\alpha\colon \neg A] \supset \mathcal{S}). \tag{3c}$$

(3b) and (3c) are harsher definitions than (3a), in the sense that if $O[\alpha\colon A]$ holds under (3a), then the obligation also holds for both (3b) and (3c). In fact, (3b) and (3c) must be rejected as too harsh. On either definition, it turns out that $O[\alpha\colon A]$ holds for all $\alpha$ if A is any tautology or contradiction. In what follows, we work with (3a): $\alpha$ is obligated to see to it that A just in case it is settled that if she does not see to it that A, then there is wrongdoing. The basic picture for $O[\alpha\colon A]$, then, is shown in fig. 2.

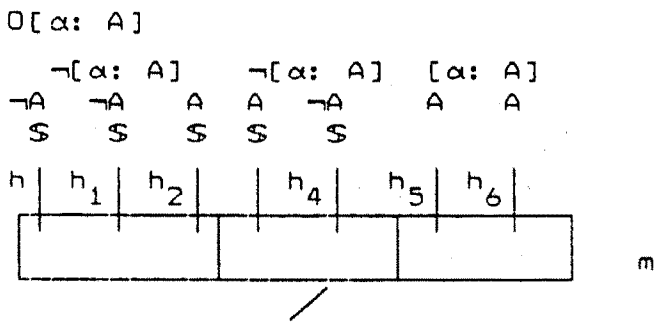


Fig. 2.

### 3.2. COMPARISON WITH OTHER SYSTEMS

It is interesting to compare this system, which we shall call **SA**,[12] to other systems of deontic logic – in particular, to the standard system **KD** and to Anderson's own work. It turns out that many of the axioms and rules in these other systems can be reformulated as valid principles in **SA**. Yet we have made no assumptions about $\mathcal{S}$, or the obligation operator, beyond the definition (3a). To the extent that validities in **SA** are reasonable, the paper's claim that a logic of obligation can be constructed as an extension to a logic of agency is strengthened.

First, there are the standard equivalences between permission, forbidding and obligation that we find in these systems:

[12] It combines STIT theory with Andersonian devices, or the Sanction with a logic of Agency.

$$Pp \equiv \neg O \neg p,$$

$$Fp \equiv O \neg p.$$

As Belnap and Perloff have argued [5], when we restrict the complements of the deontic operators to be *stit* sentences, the most reasonable definitions are

$$P[\alpha\colon A] \equiv \neg O[\alpha\colon \neg[\alpha\colon A]] \quad \text{(permitted = not obligated to refrain from doing)}, \tag{4}$$

$$F[\alpha\colon A] \equiv O[\alpha\colon \neg[\alpha\colon A]] \quad \text{(forbidden = obligated to refrain from doing)}. \tag{5}$$

Note that by iterating the modality, we avoid the problem of "negated actions". Since we also have, symmetrically,

$$O[\alpha\colon A] \equiv \neg P[\alpha\colon \neg[\alpha\colon A]] \quad \text{(obligated = not permitted to refrain)}, \tag{6}$$

(4) and (6) imply, by transitivity, that

$$O[\alpha\colon A] \equiv O[\alpha\colon \neg[\alpha\colon \neg[\alpha\colon A]]].$$

In fact, it can be verified directly that $[\alpha\colon A] \equiv [\alpha\colon \neg[\alpha\colon \neg[\alpha\colon A]]]$ is valid. It says that refraining from refraining from seeing to it that A is equivalent to seeing to it that A.[13]

The standard system **KD** contains the tautologies of propositional calculus, the rule modus ponens, the above equivalences between permission, forbidding and obligation, and three additional axioms and rules:[14]

(KD1) $O(p \supset q) \supset (Op \supset Oq)$ the K-axiom or principle of deontic detachment,

(KD2) $Op \supset Pp$ the D-axiom (obligatory implies permitted),

(KD3) $p \vdash Op$ O-necessitation.

The analogue of (KD1) for agentive sentences is

$$O[\alpha\colon p \supset q] \supset (O[\alpha\colon p] \supset O[\alpha\colon q]), \tag{7}$$

which is a valid principle in **SA**. It is a direct consequence of a fact about agency, namely, that

$$[\alpha\colon p \supset q] \supset ([\alpha\colon p] \supset [\alpha\colon q]). \tag{8}$$

[13] See [5] and [2] for more discussion of refraining from refraining. The equivalence between seeing to it that and refraining from refraining, unproblematic for *dstit*, is a more delicate matter for Belnap's *stit*.

[14] The version cited here is based on [12].

To see (8), suppose $[\alpha\colon p \supset q]$ and $[\alpha\colon p]$ hold at $(m, h)$ on model **M**. Then $p \supset q$ and p, and hence q, hold at all $(m, h')$ for $h' \equiv^{\alpha}_{m} h$ (see fig. 3). Further, $\mathbf{M} \models [\alpha\colon p \supset q][m, h]$ requires a counter $(m, h'')$ (see section 2) where $p \supset q$ is false, and thus where q is false. The positive and negative conditions for $[\alpha\colon q]$ at $(m, h)$ are both satisfied.

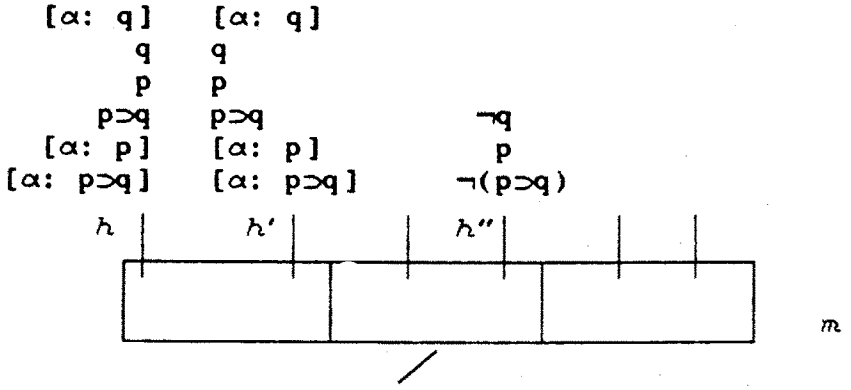


Fig. 3.

From (8), (7) follows easily. Suppose $O[\alpha\colon p \supset q]$ and $O[\alpha\colon p]$ hold at $(m, h)$. If $\neg[\alpha\colon q]$ holds at any $(m, h')$, then either $\neg[\alpha\colon p]$ or $\neg[\alpha\colon p \supset q]$ holds there, by (8). However, by definition (3a), $\mathcal{S}$ must then hold at $(m, h')$. This proves $\mathcal{L}(\neg[\alpha\colon q] \supset \mathcal{S})$, i.e. $O[\alpha\colon q]$.

Corresponding to (KD2), we have

$$O[\alpha\colon p] \supset P[\alpha\colon p]. \tag{9}$$

This seems a perfectly reasonable principle. It I am obligated to drive under the speed limit, then I am permitted to do so. In fact, (9) is valid in **SA** under the mild assumption that the sanction $\mathcal{S}$ is not settled true. Let us call a moment at which $\mathcal{S}$ is settled true a *no-good-choice* moment. In the *no-good-choice* case, it is settled that everything $\alpha$ does or does not do leads to the sanction. In this pathological situation, everything is obligatory, everything is forbidden, and nothing is permitted for $\alpha$, as a consequence of definitions (3a), (4) and (5). Clearly, (9) is false in the *no-good-choice* case.

The Andersonian system [1] adopted the axiom that the sanction should be contingent. We do not believe we can accept the corresponding axiom, $\neg\mathcal{L}\mathcal{S}$. We are not entitled to assume that *no-good-choice* moments can be ruled out *a priori*. Genuinely conflicting obligations seem at least possible, especially if the authority issuing obligations is unreasonable.

In any case, the valid analogue of (KD2), stated without proof, is

$$(O[\alpha\colon p] \wedge \neg\mathcal{L}\mathcal{S}) \supset P[\alpha\colon p]. \tag{10}$$

There is nothing in SA that corresponds to the rule of O-necessitation, (KD3), a rule that can also be derived in Anderson's system. The rule

$$[\alpha\colon \mathrm{A}] \vdash \mathrm{O}[\alpha\colon \mathrm{A}]$$

would be vacuous, since $[\alpha\colon \mathrm{A}]$ is never a logical validity (because of the counter). Furthermore, it follows from definition (3a) that $\mathrm{O}[\alpha\colon \top]$ is always false, except in the *no-good-choice* case. Perhaps it will come as a relief to learn that we are under no obligation to see to it that $2 + 2 = 4$.

## 4. Completeness

For this section only, we reduce our language (and our definition of frame, model, truth in a model, etc.) by eliminating the operators $\boldsymbol{P}$ (past) and $\boldsymbol{F}$ (future), and by restricting $\overline{\alpha}$ to contain only one agent, $\alpha$. The language still contains the constant $\mathcal{S}$. A completeness result is stated without proof.

Recalling abbreviation (2), $\Delta\alpha\mathrm{A} =_{\mathrm{df}} \mathcal{L}\mathrm{A} \vee [\alpha\ dstit:\ \mathrm{A}]$, we take as axioms for a system, $\mathrm{SA}_0$, all instances of truth-functional tautologies as well as the following schemata:[15]

**A1** $\Box(\mathrm{A} \supset \mathrm{B}) \supset (\Box\mathrm{A} \supset \Box\mathrm{B})$,

**A2** $\Box\mathrm{A} \supset \mathrm{A}$,

**A3** $\neg\Box\neg\mathrm{A} \supset \Box\neg\Box\neg\mathrm{A}$,

**A4** $\mathcal{L}(\mathrm{A} \supset \mathrm{B}) \supset (\mathcal{L}\mathrm{A} \supset \mathcal{L}\mathrm{B})$,

**A5** $\mathcal{L}\mathrm{A} \supset \mathrm{A}$,

**A6** $\neg\mathcal{L}\neg\mathrm{A} \supset \mathcal{L}\neg\mathcal{L}\neg\mathrm{A}$,

**A7** $\Delta\alpha(\mathrm{A} \supset \mathrm{B}) \supset (\Delta\alpha\mathrm{A} \supset \Delta\alpha\mathrm{B})$,

**A8** $\Delta\alpha\mathrm{A} \supset \mathrm{A}$,

**A9** $\neg\Delta\alpha\neg\mathrm{A} \supset \Delta\alpha\neg\Delta\alpha\neg\mathrm{A}$,

**A10** $\Box\mathrm{A} \supset \mathcal{L}\mathrm{A}$,

**A11** $\mathcal{L}\mathrm{A} \supset \neg[\alpha\ dstit:\ \mathrm{A}]$.

As rules of inference, we take *modus ponens* and the rule of necessitation

**RN** $\mathrm{A} \vdash \Box\mathrm{A}$.

It is easy to see from definition (2) and **A11** that **T1** and **T2** hold (see section 2), and it is also clear that the following rules are derivable:

[15] The axiomatization is based on that of Xu [18].

**R1** $A \vdash \mathcal{L}A$,

**R2** $A \vdash \Delta\alpha A$.

Axioms **A1–A9** reflect the fact that $\Box$, $\mathcal{L}$, and $\Delta\alpha$ are like the S5 modality. **A10** and **A11** state the relationships between these different modalities.

$\mathbf{SA}_0$ contains no axioms about the O operator specifically, although the $\mathcal{S}$ does occur in substitution instances of its axioms. Using the definition (3a) of O, results such as (7) and (10) can be proven as theorems of $\mathbf{SA}_0$. This is a consequence of the completeness property stated below. Without $\mathcal{S}$, the system is simply an axiomatization of *dstit*.

4.1. SOUNDNESS THEOREM

For every formula $A$, $\vdash A$ in $\mathbf{SA}_0$ only if $\mathbf{M} \models A$ for every model $\mathbf{M}$.

*Proof*

By induction on formulas. □

4.2. COMPLETENESS THEOREM

For every formula $A$, $\vdash A$ in $\mathbf{SA}_0$ if $\mathbf{M} \models A$ for every model $\mathbf{M}$.

*Proof*

By Xu [18].[16] The proof is similar to completeness proofs for S5. □

## 5. Conditional obligation

Using the semantics developed in sections 2 and 3, a precise condition can be given under which "detachment" of obligation is acceptable, so that the argument (1) goes through. The argument (1) is reformulated as follows:

$$\begin{array}{ll} \text{(a)} & O[\alpha\colon (B \supset [\alpha\colon A])] \\ \text{(b)} & \underline{B\qquad\qquad\qquad} \\ \text{(c)} & O[\alpha\colon A] \end{array} \tag{11}$$

In English:

(a) Alchourrón is obligated to see to it that if Bulygin sends him the draft of their paper, he sees to it that he (Alchourrón) revises it.

(b) Bulygin sends him the draft.

[16] Xu's result is stated only for the operators $\Delta\alpha$ and $\mathcal{L}$, but is easily generalized to include $\Box$ as well.

Therefore,

(c) Alchourrón is obligated to see to it that he revises the draft.

Version (11) is obtained from (1) in two stages. First, we replace A by $[\alpha\colon A]$ in (a) and (c), since "revising the draft" is agentive. Second, we place the conditional $B \supset [\alpha\colon A]$ inside the *dstit* sentence $[\alpha\colon (B \supset [\alpha\colon A])]$, since the obligation in (a) is that Alchourrón should see to it that the conditional is true.

Argument (11) is not valid without an added assumption:

$$\mathcal{L}(\neg[\alpha\colon \neg B]). \tag{C}$$

This says that it is settled that $\alpha$ cannot see to it that B is false. Condition (C) is one way to formalize the assumption that B is a circumstance,[17] for it captures the idea that $\alpha$ cannot prevent B from being true. Assuming (C) makes the argument (11) go through. On the other hand, if (C) is false, the argument fails in general.

*Proof of (11), assuming (C)*

We assume $O[\alpha\colon (B \supset [\alpha\colon A])]$, B, and $\mathcal{L}(\neg[\alpha\colon \neg B])$ are true at $(m, h)$ in model M. By the first assumption, if $h'$ is any history through $m$, then by definition (3a),

$$\mathbf{M} \models (\neg[\alpha\colon (B \supset [\alpha\colon A])] \supset \mathcal{S})[m, h']. \tag{12}$$

We want to show that for any $h'$ through $m$,

$$\mathbf{M} \models (\neg[\alpha\colon A] \supset \mathcal{S})[m, h'], \tag{13}$$

i.e. that $O[\alpha\colon A]$ is true at $(m, h)$.

The crucial thing to notice is that from B and $\mathcal{L}(\neg[\alpha\colon \neg B])$ at $(m, h)$, it follows that for each history $h'$ through $m$, there is at least one choice-equivalent history $h''$ such that B is true at $(m, h'')$. Less formally, out of each choice box at $m$ comes at least one history $h''$ on which B is true (see fig. 4). For if this were

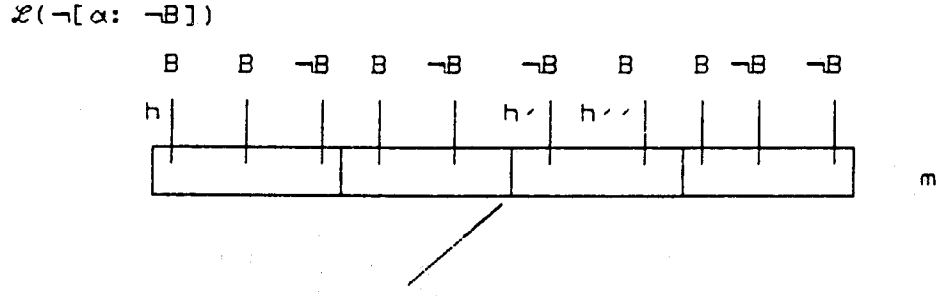


Fig. 4.

[17] Castañeda's idea of a circumstance is different, since in his system circumstances sometimes are within the agent's control.

not so, then $[\alpha: \neg B]$ would be true at $(m, h')$, with the counter at $(m, h)$, where B is true. This would violate the fact that $\neg[\alpha: \neg B]$ is settled true.

Now if $\neg[\alpha: A]$ were true at $(m, h')$, it would also be true at $(m, h'')$. Then the conditional

$$(B \supset [\alpha: A])$$

would be false at $(m, h'')$, so that $[\alpha: B \supset [\alpha: A]]$ would be false at $(m, h')$. By (12), $\mathcal{S}$ would follow at $(m, h')$, and we have shown (13). □

Thus,

$$O[\alpha: (B \supset [\alpha: A])] \supset (B \supset O[\alpha: A])$$

is valid if B is a circumstance in the sense of (C).

## 6. O⊃-statements versus ⊃O-statements

The above result is a possible explanation of the fact that English is largely indifferent to the distinction between O⊃-statements and ⊃O-statements. For the purposes of detaching the obligation $O[\alpha: A]$, the two forms $B \supset O[\alpha: A]$ and $O[\alpha: B \supset [\alpha: A]]$ are equivalent, provided only the antecedent B is a circumstance as defined by (C). The following examples show that many common O⊃-statements do satisfy the "circumstance condition".

(a) *"It is your duty to apologize, if you have behaved badly at the party".*

If we want to represent this obligation as an O⊃-statement, we suggest that it has the following form:

$$O[\alpha: \boldsymbol{P}[\alpha: B] \supset [\alpha: A]]. \tag{14}$$

In English, "You are obligated to see to it that if you saw to it that you behaved badly, you see to it that you apologize".

Now notice that the antecedent $\boldsymbol{P}[\alpha: B]$ satisfies the "circumstance condition", $\mathcal{L}(\neg[\alpha: \neg \boldsymbol{P}[\alpha: B]])$. Agent $\alpha$ cannot see to the falsity of a statement about his past seeing-to's. The reason is that either $\mathcal{L}(\boldsymbol{P}[\alpha: B])$ or $\mathcal{L}(\neg \boldsymbol{P}[\alpha: B])$ must hold at any moment–history pair. In the first case, the positive condition for $[\alpha: \neg \boldsymbol{P}[\alpha: B]]$ fails, and in the second case, the negative condition fails.

The proof of the "either-or" condition is as follows. Suppose that $\boldsymbol{P}[\alpha: B]$ is true at $(m, h)$. Then for some $m' < m$, $[\alpha: B]$ is true at $(m', h)$; consequently, $[\alpha: B]$ is true at $(m', h')$ for every $h'$ with $h' \equiv^{\alpha}_{m'} h$. Since every history $h'$ through $m$ passes through $m'$ and satisfies $h' \equiv^{\alpha}_{m'} h$,[18] it follows that $\boldsymbol{P}[\alpha: B]$ is true at $(m, h')$

[18] This is a consequence of the *no choice between undivided histories* condition discussed in section 2.

for all $h'$ through $m$. Thus, $\boldsymbol{P}[\alpha\colon \mathrm{B}]$ at $(m, h)$ implies $\mathcal{L}(\boldsymbol{P}[\alpha\colon \mathrm{B}])$. So either $\boldsymbol{P}[\alpha\colon \mathrm{B}]$ holds at all histories or at none, which is precisely the "either-or" condition.[19]

Since $\boldsymbol{P}[\alpha\colon \mathrm{B}]$ is a circumstance, whenever it is true we can detach the obligation $\mathrm{O}[\alpha\colon \mathrm{A}]$. It seems to us that many conditional obligations have the form of (14), even when the antecedent seems to be present tensed.

Consider the obligation: "It is your duty to apologize if you behave badly at the party". What is the tense of "behaving badly" relative to "apologizing"? It must be future, present (contemporaneous), or past. Taking the tense as either future or present is not a reasonable interpretation of the duty, since any apology given before or at the moment of behaving badly will hardly be convincing. This leaves us with the same obligation as before:

> "It is your duty to see to it that you (see to it that you) apologize, if you have (seen to it that you) behaved badly at the party".

The form of this obligation is then the same as (14) above.

Many conditional obligations whose antecedents are definite actions (aorists in Ancient Greek) have the form of (14). An important exception will be discussed in section 9.

(b) *"It is your duty to admonish Bob if he behaves badly (has behaved badly)".*

This could be expressed as

$$\mathrm{O}[\alpha\colon \boldsymbol{P}[\beta\colon \mathrm{B}] \supset [\alpha\colon \mathrm{A}]]. \tag{15}$$

Translated back into English, this reads "It is your duty to see to it that if Bob saw to it that he behaved badly, you see to it that you admonish him". The only difference between this and example (a) is that a different agent is involved. The same argument as above shows that $\boldsymbol{P}[\beta\colon \mathrm{B}]$ is a circumstance. Consequently, it does not really matter if we express the obligation as an $\mathrm{O}\supset$-statement or as a $\supset\mathrm{O}$-statement.

Expression (15) is also a plausible way to formalize Alchourrón's conditional obligation to "revise the draft". In (11), we symbolized that obligation as

$$\mathrm{O}[\alpha\colon (\mathrm{B} \supset [\alpha\colon \mathrm{A}])],$$

[19] Nuel Belnap has made the following observation. On Prior's Ockhamist semantics, it is not generally true that the past is settled, i.e. that for an arbitrary sentence Q, $\mathcal{L}(\boldsymbol{P}(\mathrm{Q}))$ or $\mathcal{L}(\neg\boldsymbol{P}(\mathrm{Q}))$ must hold at each moment–history pair. For instance, it may be that neither $\boldsymbol{P}(\boldsymbol{F}(\mathrm{Q}))$ nor $\neg\boldsymbol{P}(\boldsymbol{F}(\mathrm{Q}))$ is settled true. Further, $[\alpha\colon \mathrm{Q}]$ is never settled true because of the counter. It is only the combination of the past operator $\boldsymbol{P}$ with the Horty *dstit* that leads to a form that is bound to be settled true or settled false. In this special case, the Ockhamist semantics agrees with most other accounts of the past, according to which it is always true that either $\boldsymbol{P}(\mathrm{Q})$ or $\neg\boldsymbol{P}(\mathrm{Q})$ is settled.

where $\alpha$ is Alchourrón, B is "Bulygin sends Alchourrón the draft", and A is "Alchourrón revises the draft". We then had to make the extra assumption that B was a circumstance. Since Bulygin's sending the draft is a definite action which takes place prior to the revising, Alchourrón's obligation is better formalized as

$$O[\alpha\colon (P[\beta\colon B] \supset [\alpha\colon A])].$$

"Alchourrón is obligated to see to it that if Bulygin has seen to it that he sent Alchourrón the draft, he sees to it that he revises the draft". The circumstance condition is now redundant, since $P[\beta\colon B]$ automatically satisfies it.

(c) *Present-tensed circumstances*

We might be tempted by these two examples to suppose that circumstantiality is somehow bound up with the past tense. The supposition would be false. In fact, most present-tensed statements are circumstantial. As a typical example, consider the sentence "it is raining". If you have promised to bring an umbrella if it rains, then your obligation can be represented in either of the following forms:

$$O[\alpha\colon (R \supset [\alpha\colon U])], \tag{16a}$$

$$R \supset O[\alpha\colon U]. \tag{16b}$$

Clearly, "it is raining" is circumstantial: $\mathcal{L}(\neg[\alpha\colon \neg R])$ or, in English, it is settled that you cannot see to it that it is not raining. You may chant or dance, but in the end it is up to nature to stop the rain. In fact, for a given moment, it is probably fair to assume either $\mathcal{L}(R)$ or $\mathcal{L}(\neg R)$ holds.

Since "it is raining" is a circumstance, we can detach from (16a) the obligation to bring an umbrella, $O[\alpha\colon U]$, if it actually does rain. Once again, the system SA is indifferent to whether we represent the obligation as (16a) or (16b), just as English is indifferent between the corresponding English sentences.

"It is raining" is a present-tensed non-agentive sentence. Can present-tensed agentive sentences also be circumstantial? It is an interesting fact about *dstit* that the answer depends entirely on the agent. It turns out that the following are true:

$$\mathcal{L}(\neg[\alpha\colon \neg[\beta\colon B]]), \tag{17a}$$

$$\neg\mathcal{L}(\neg[\alpha\colon \neg[\alpha\colon A]]), \text{ except when } \mathcal{L}(\neg[\alpha\colon A]). \tag{17b}$$

The first, (17a), says that other agents' present doings are always circumstantial for agent $\alpha$.[20] Expression (17b) says that $\alpha$'s own doings are never circumstantial for

[20] The proof of this assertion, which involves extending the semantics of *dstit* to multiple agents, is beyond the scope of this paper. It relies on an assumption that simultaneous choices by different agents are independent. It should be noted that the situation is different for Belnap's *stit*. There, it is possible for one agent to see to the falsity of another agent's doing something.

$\alpha$, except when it is settled that $\alpha$ does not see to something. This becomes important in the discussion of contrary-to-duty obligations (section 8). Momentarily confining our attention to the case of different agents, we have the result that SA is indifferent between

$$O[\alpha\colon ([\beta\colon B] \supset [\alpha\colon A])]$$

and

$$[\beta\colon B] \supset O[\alpha\colon A]$$

as ways of putting the obligation to see to it that you ($\alpha$) show up to the meeting if your boss ($\beta$) shows up to the meeting. In either case, we can detach $O[\alpha\colon A]$ provided $[\beta\colon B]$ holds.

We could also provide examples of future-tensed sentences satisfying the circumstance condition, such as "It will rain". The point is that circumstantiality is not derivative of temporal ordering. It depends only on what it is possible for $\alpha$ to see to at a given moment.

## 7. The Good Samaritan

The paradox of the Good Samaritan relies on the following principle:[21]

> If $\alpha$ *performs* A entails *a performs* B, then $\alpha$ *is obligated to do* A entails $\alpha$ *is obligated to do* B. (18)

The paradox now proceeds:

(a) Arthur is obligated to perform the act, call it C, of bandaging the man he will murder a week from now.

(b) But Arthur's doing C entails his doing the act of murdering a man a week hence.

So, by (a), (b) and (18),

(c) Arthur is obligated to murder a man a week hence.

We want to reject (c), but (18), (a) and (b) all seem acceptable.

People usually attempt to resolve the paradox through analysis of tense, agency and the sense of entailment in (18) and (b). We can bring all these considerations to bear in a precise way by using the *dstit* semantics. Let M stand for "Arthur murders a man". Then "Arthur murders a man a week hence" can be approximately translated as $F[\alpha\colon M]$. Therefore, we can represent (b) as $[\alpha\colon C] \supset F[\alpha\colon M]$. ("Arthur sees to it that C" entails that in the future, Arthur sees to it that he murders a man.) As a first stab (!) at (18), we can try

[21] The paradox is cited here as it is given by Castañeda [6], with minor changes.

$$([\alpha\colon \mathrm{A}] \supset [\alpha\colon \mathrm{B}]) \supset (\mathrm{O}[\alpha\colon \mathrm{A}] \supset \mathrm{O}[\alpha\colon \mathrm{B}]). \qquad (18')$$

Then (18′) does not apply to (b), since $F[\alpha\colon \mathrm{M}]$ is not an agentive sentence, but a future agentive sentence. Hence, (c) does not follow.

Furthermore, we should note that (18′) is, quite properly, invalid, as fig. 5 illustrates.

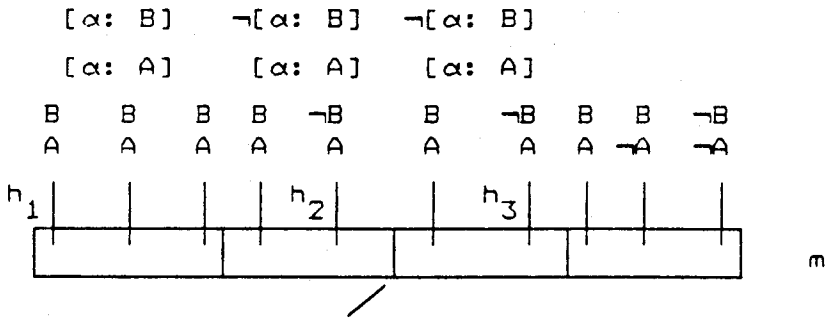


Fig. 5.

Figure 5 shows a situation in which $[\alpha\colon \mathrm{A}] \supset [\alpha\colon \mathrm{B}]$ at $h_1$, but not at $h_2$ or $h_3$. In such a situation, $\mathrm{O}[\alpha\colon \mathrm{A}] \supset \mathrm{O}[\alpha\colon \mathrm{B}]$ fails at $h_1$. Indeed, if it is possible to see to it that A (bandaging a man) without seeing to it that B (killing a man), the conditional $\mathrm{O}[\alpha\colon \mathrm{A}] \supset \mathrm{O}[\alpha\colon \mathrm{B}]$ should be false.

When we strengthen the antecedent, we obtain a valid form of (18):

$$\mathcal{L}([\alpha\colon \mathrm{A}] \supset [\alpha\colon \mathrm{B}]) \supset (\mathrm{O}[\alpha\colon \mathrm{A}] \supset \mathrm{O}[\alpha\colon \mathrm{B}]). \qquad (18'')$$

The proof is straightforward. (Alternatively, we could put $\Box$ in place of $\mathcal{L}$.) However, (18″) still does not apply to (b). Two possibilities remain for preserving the paradox. First, we can suppose that Arthur can see to it *right now* that "a man is dead a week hence". We restate premise (b) as "Arthur's doing C entails his seeing to it (now) that a man is dead a week hence":

$$\mathcal{L}([\alpha\colon \mathrm{C}] \supset [\alpha\colon F(\mathrm{D})]), \qquad (b')$$

where D stands for "a man is dead". Premise (a) is translated as $\mathrm{O}[\alpha\colon \mathrm{C}]$. Then $\mathrm{O}[\alpha\colon F(\mathrm{D})]$ does follow. However, to assume (b′) is to assume that Arthur cannot bandage the injured man without seeing to it that he is dead a week hence. (Perhaps the bandages are coated with poison.) It is doubtful that $\mathrm{O}[\alpha\colon \mathrm{C}]$ holds under these conditions. The paradox has lost its sting, since $\mathrm{O}[\alpha\colon \mathrm{C}]$ and $\mathrm{O}[\alpha\colon F(\mathrm{D})]$ are equally objectionable.

A more reasonable approach is to represent (b) as

$\mathcal{L}([\alpha\colon \mathrm{C}] \supset F[\alpha\colon \mathrm{M}])$ (b″)

(It is settled that seeing to it that C entails, in future, murdering a man.)

and to propose yet another version of (18):

$\mathcal{L}([\alpha\colon \mathrm{A}] \supset F[\alpha\colon \mathrm{B}]) \supset (\mathrm{O}[\alpha\colon \mathrm{A}] \supset F\mathrm{O}[\alpha\colon \mathrm{B}])$ (18‴)

(If seeing to it that A entails in future seeing to it that B, then a present obligation to see to it that A entails a future obligation to see to it that B.)

Of all the formulations of (b) and (18), these seem most natural. Together, (a), (b″) and (18‴) do imply that Arthur has a future obligation to murder the man he bandages – the paradoxical result. The argument fails, however, because (b″) is always false (except in the trivial case where $[\alpha\colon \mathrm{C}]$ is settled false, i.e. when Arthur cannot bandage the man). It cannot be settled that a present agentive sentence entails a future agentive sentence. The reason is the negative condition (genuine choice) required for the truth of the future agentive sentence. In fig. 6, the counter for $[\alpha\colon \mathrm{M}]$ at $(m', h)$ is at history $h'$. So $[\alpha\colon \mathrm{M}]$ fails at $(m', h')$. Since $h' \equiv^{\alpha}_{m} h$, $[\alpha\colon \mathrm{C}] \supset F[\alpha\colon \mathrm{M}]$ is false at $(m, h')$, proving that (b″) is false.

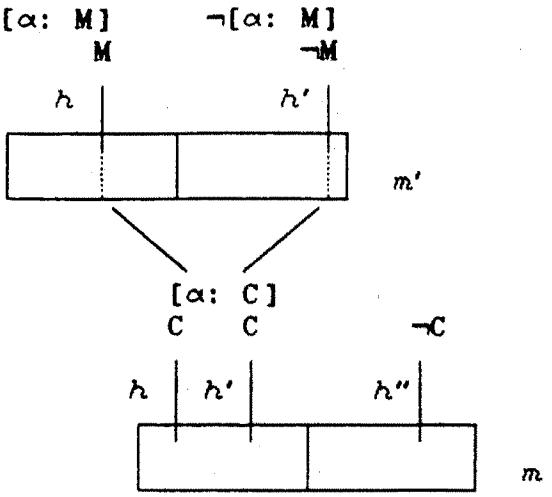


Fig. 6.

The Good Samaritan Paradox rests on ambiguities of tense, entailment and agency. The semantical system SA makes it possible to unravel these elements in a graphic fashion, so that the argument is either invalid, or one or more premises are manifestly false.

## 8. Contrary-to-duty obligations

One of the most commonly discussed problems of deontic logic is the *paradox of contrary-to-duty obligations* (*imperatives*), so named by R. Chisholm [8]. Such paradoxes arise in cases in which something forbidden is done. Consider the "gentle murder" example:

(a) A certain man is not obligated to murder his neighbour; in fact, he is obligated not to murder his neighbour.

(b) If he does murder his neighbour, he is obligated to murder gently.

(c) Murdering gently entails murdering.

(d) He murders his neighbour.

Chisholm calls the sort of obligation in (b) a *contrary-to-duty* imperative. Such an imperative says what a person ought to do if he has violated his duties. It is widely recognized that deontic logic should be able to accommodate contrary-to-duty obligations, but they pose a difficulty for the standard system. Statements (a)–(d) seem perfectly consistent, but cannot be consistently represented in the standard system.

Let 'M' stand for the sentence "$\alpha$ (the man) murders his neighbour", and 'G' for the sentence "$\alpha$ murders his neighbour gently". Then the most reasonable way to represent the four statements in standard deontic logic is as follows:

(KDa) $\neg OM;\ O\neg M,$

(KDb) $M \supset OG,$

(KDc) $G \supset M,$

(KDd) $M.$

From (KDb), (KDc) and (KDd) together with the valid schema $p \supset q \vdash Op \supset Oq$, we can infer OM, contradicting (KDa).

A possible response to this problem is to try representing the contrary-to-duty obligation (KDb) as $O(M \supset G)$. This avoids inconsistency, but (KDb) then becomes a redundant premise. For $O(M \supset G)$, and in fact $O(M \supset A)$ for any sentence A, follows from $O\neg M$. We could equally well add $O(M \supset \neg G)$. This suggests that $O(M \supset G)$ is a bad way to represent the contrary-to-duty obligation.

The semantics of SA provides a way to represent the statements as a consistent set without redundant obligations:

(SAa) $\neg O[\alpha: M];\ O[\alpha: \neg[\alpha: M]],$

(SAb) $O[\alpha: [\alpha: M] \supset [\alpha: G]],$

(SAc) $\Box([\alpha: G] \supset [\alpha: M]),$

(SAd) $[\alpha: M].$

The first statement expresses $\alpha$'s obligation to refrain from murdering as well as the fact that he has no obligation to murder. The second statement expresses the contrary-to-duty obligation. Unlike the situtation in the standard system, (SAb) does not follow from (SAa).

Furthermore, we avoid a contradiction between (SAb) and (SAa) because we cannot detach the obligation to murder gently, $O[\alpha: G]$. The reason is that $[\alpha: M]$ is not a circumstance in the sense of condition (C). Looking again at (17b), we see that it is never the case, given (SAd), that $\mathcal{L}(\neg[\alpha: \neg[\alpha: M]])$. It is *not* settled that $\alpha$ cannot refrain from murdering; the counter to (SAd) guarantees $\alpha$ the choice of refraining from murdering. The obligation to murder gently always remains just a conditional obligation, as it should.

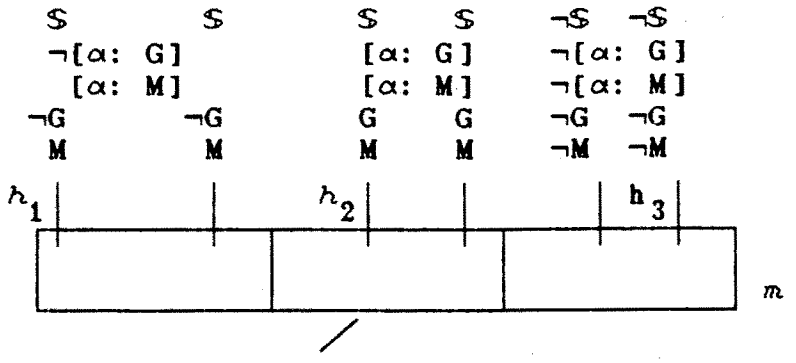


Fig. 7.

Figure 7 illustrates the situation. On history $h_1$, $\alpha$ violates both obligations (SAa) and (SAb). On history $h_2$, he violates only the obligation not to murder. Finally, on $h_3$ (where (SAd) is false) he satisfies both obligations. Clearly, $O[\alpha: G]$ does not hold, since it is possible to have $\neg[\alpha: G]$ without $\mathcal{S}$ (as at $h_3$).

Chisholm's own example [8] is the following:

(a) It ought to be that a certain man go to the assistance of his neighbours.

(b) It ought to be that if he does go he tell them he is coming.

(c) If he does not go then he ought not to tell them he is coming.

(d) He does not go.

It can be handled in the same way as the gentle murder case. Let 'A' stand for the sentence "$\alpha$ (the man) goes to the assistance of his neighbours", and 'T' for the sentence "$\alpha$ tells them he is coming". Then we symbolize the paradox as follows:

(SAa) $O[\alpha: A]$,

(SAb) $O[\alpha: [\alpha: A] \supset [\alpha: T]]$,

(SAc) $O[\alpha: \neg[\alpha: A] \supset [\alpha: \neg[\alpha: T]]]$,

(SAd) $\neg[\alpha: A]$.

Provided $[\alpha: A]$ is possible (he can assist his neighbours), $\neg[\alpha: A]$ is not a circumstance. So we may not detach the obligation to refrain from telling his neighbours he will come, and there is no contradiction. There is only a conditional, not a categorical, obligation not to tell. (If it is impossible for him to assist his neighbours, $\neg[\alpha: A]$ is a circumstance and we can detach a categorical obligation not to tell. $O[\alpha: T]$ and $O[\alpha: \neg[\alpha: T]]$ will conflict in this special case, but it can plausibly be argued that there is a genuine conflict of obligations.)

## 9. Problems with the proposed semantics of obligation

The semantics of obligation suggested here faces numerous difficulties. Two of the most serious are discussed here.

(1) The system does not permit different obligations to hold at different histories at the same moment. If an obligation hold at one moment–history pair $(m, h)$, it is settled at $m$. Yet there are situations in which an agent's choices lead to different obligations. The most obvious is making promises, as represented in fig. 8. P stands for "$\alpha$ promises to call"; C stands for "$\alpha$ calls". It seems that $O[\alpha: C]$ should hold at $(m, h_1)$, but not at $(m, h_2)$.

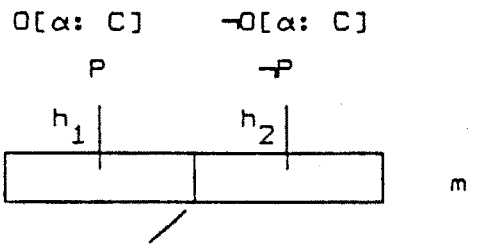


Fig. 8.

On the proposed semantics, there are two possible ways around this difficulty. One way is to argue that the obligation to call does not really come into force until a time after the promise is made, i.e. once it is settled that $\alpha$ has made the promise. What really holds at $h_1$ is the sentence $P \supset F(O[\alpha: C])$. At some later time, the obligation to call will hold over the entire moment. However, it is not satisfying to have to introduce extraneous temporal considerations to resolve the problem.

A second approach, following the treatment of contrary-to-duty obligations, it to treat the obligation to call as the conditional obligation $O[\alpha: [\alpha: P] \supset [\alpha: C]]$. Since promising is not a circumstance, the obligation to call cannot be detached.

The conditional obligation, which holds for the whole moment, is only violated on histories where the agent makes a promise to call but fails to call. This solution is also unconvincing, however, because we have to introduce a conditional structure into what seems an unconditional obligation created by the promise.

The difficulty derives from the fact that our definition of obligation quantifies over all histories through a given moment. If we replace $\mathcal{L}$ by $\Delta\alpha$ in (3a) and define

$$O[\alpha\colon A] \Leftrightarrow \Delta\alpha(\neg[\alpha\colon A] \supset \mathcal{G}),$$

then we allow for obligations which vary depending on $\alpha$'s choices. The problem with this definition is that $[\alpha\colon A] \supset O[\alpha\colon A]$ becomes valid because $[\alpha\colon A] \equiv \Delta\alpha[\alpha\colon A]$. Given $[\alpha\colon A]$, $\Delta\alpha(\neg[\alpha\colon A] \supset \mathcal{G})$ follows classically. There might be some hope if we use a non-classical logic (e.g. relevance logic), but this will have to be pursued elsewhere.

(2) There are conditional obligations of the form

$$O([\alpha\colon \boldsymbol{F}[\alpha\colon A] \supset [\alpha\colon B]]),$$

which the semantics proposed here seems unable to handle. A good example, due to Castañeda,[23] is the following:

> Mary is secretarially obligated to report to the manager by 8:45 that she won't open the office by 9 a.m., if she won't.

This fits the above form if $\alpha$ is Mary, A is "Mary does not open the office by 9 a.m.", and B is "Mary reports to the manager by 8:45 a.m." Imagine that it is now 8:45, and Mary is staying home to care for a sick child. Since the conditional obligation is in force, it seems that we should be able to detach the obligation to call the manager: $O[\alpha\colon B]$. However, we cannot do so on the given semantics. The antecedent is not a circumstance in the sense of (C), since Mary might still be able to make it to the office by 9:00.

One way around this difficulty is to generalize the notion of *stit*. When an agent $\alpha$ can see to a state of affairs A by making a series of choices, we say $\alpha$ can strategically guarantee A, written $[\alpha\ strat\colon A]$. The technical definition of a strategy is beyond the scope of this paper, but it is reasonable to expect that one could define the obligation $O[\alpha\ strat\colon A]$ in a manner which generalizes the definitions given for $O[\alpha\ dstit\colon A]$ in section 3. By distinguishing between histories which are part of a strategy and those which are not, one could hopefully provide a definition of obligation which would allow detachment in cases such as Castañeda's example.

[23] Mentioned by Castañeda in correspondence.

The simple semantics of obligation developed here needs to be improved, as the above difficulties illustrate. The point of this paper was to show that a semantics based on the logic of agency provides a useful tool for thinking about problems of conditional obligation and paradoxes of deontic logic. Considerations of tense and agency often lie at the heart of such problems, and the proposed semantics allows such considerations to be brought fully into play.

## Acknowledgement

The author is grateful to Nuel Belnap for encouragement, discussion and suggestions regarding the ideas in this paper.

## References

[1] A.R. Anderson, The formal analysis of normative systems, in: *The Logic of Decision and Action*, ed. N. Rescher (Pittsburgh, 1966).
[2] N. Belnap, Backwards and forwards in the modal logic of agency, unpublished manuscript, Department of Philosophy, University of Pittsburgh (1989). Forthcoming in Philos. Phenomen. Res. in early 1993.
[3] N. Belnap, Declaratives are not enough, Philos. Studies 59(1990)1.
[4] N. Belnap, Before refraining: concepts for agency, Erkenntnis 34(1991)137.
[5] N. Belnap and M. Perloff, Seeing to it that: A canonical form for agentives, Theoria 54(1988)175.
[6] H.-N. Castañeda, The paradoxes of deontic logic: The simplest solution to all of them in one fell swoop, in: *New Studies in Deontic Logic* (Reidel, Dordrecht, 1981).
[7] B.F. Chellas, Time and modality in the logic of agency, unpublished manuscript, Department of Philosophy, University of Calgary, Calgary, Canada (1991).
[8] R.M. Chisholm, Contrary-to-duty imperatives and deontic logic, Analysis 24(1963)33.
[9] D. Føllesdal and R. Hilpinen, Deontic logic: An introduction, in: *Deontic Logic: Introductory and Systematic Readings* (Reidel, Dordrecht, 1971).
[10] P.S. Greenspan, Conditional oughts and hypothetical imperatives, J. Philosophy 72(1975)259.
[11] J. Hintikka, Some main problems of deontic logic, in: *Deontic Logic: Introductory and Systematic Readings* (Reidel, Dordrecht, 1971).
[12] J.-J. Ch. Meyer and R.J. Wieringa, Deontic logic: A concise overview, in: *DEON'91 Proceedings*, ed. J.-J. Ch. Meyer and R.J. Wieringa (Amsterdam, The Netherlands, 1991).
[13] A. Prior, *Past, Present and Future* (Oxford, 1967).
[14] R.H. Thomason, Deontic logic and the role of freedom in moral deliberation, in: *New Studies in Deontic Logic* (Reidel, Dordrecht, 1981).
[15] R.H. Thomason, Indeterminist time and truth-value gaps, Theoria 36(1970)264.
[16] R.H. Thomason, Combinations of tense and modality, in: *Handbook of Philosophical Logic*, Vol. II (Reidel, Dordrecht, 1984).
[17] F. von Kutschera, Bewirken, Erkenntnis 24(1986)253.
[18] M. Xu, Logics of deliberative *stit*, Working Paper, Department of Philosophy, University of Pittsburgh (1992).