

AN ANALYSIS OF HANSSON'S DYADIC DEONTIC LOGIC¹

0. INTRODUCTION AND SUMMARY

Recently, Bengt Hansson presented a paper about dyadic deontic logic,² criticizing some purely axiomatic systems of dyadic deontic logic and proposing three purely semantical systems of dyadic deontic logic which he confidently called dyadic standard systems of deontic logic (DSDL1–3). Here I shall discuss the third by far most interesting system DSDL3 which is operating with preference relations. First, I shall describe this semantical system (Sections 1.1–1.3). Then I shall give an axiomatic system (Section 1.4) which is proved to be correct (Section 2) and complete (Section 3) with respect to Hansson's semantics. Finally, in face of these results Hansson's semantics will be discussed from a more intuitive standpoint. After emphasizing its intuitive attractiveness (Section 4.1) I will show that two objections often discussed in connection with preference relations do not apply to it (Section 4.2 and 4.3); more precisely, I will show that the connectedness condition for preference relations can be dropped and that, in a sense, it is not necessary to compare two possible worlds differing in infinitely many respects. (What exactly is meant by this, will become clear later on.) Yet there is a third objection to Hansson's semantics which points to a real intuitive inadequacy of DSDL3. A way of removing this inadequacy, which corresponds to Hansson's own intuitions as well as to familiar metaethical views, is suggested, but not technically realized (Section 4.4). In the last section (Section 4.5) I shall briefly show that DSDL3 is decidable, as expected.

1. THE FORMAL SYSTEM DSDL3

1.1 *The Basic Language.* Our starting point is the basic language (BL), as Hansson calls it,³ which we assume, without loss of generality, to be the ordinary propositional logic with propositional variables ' p ', ' q ' etc. and with the usual logical connectives ' \neg ', ' \wedge ', ' \vee ', ' \rightarrow ' and ' \leftrightarrow '. As variables for formulas

of BL we shall use ' A ', ' B ' and ' C ' with or without subscripts. Throughout this paper ' t ' and ' k ' are to stand for a certain tautology and a certain contradiction of BL respectively, say ' $p \vee \neg p$ ' and ' $p \wedge \neg p$ '.

Further, we shall need some semantical concepts for BL: ' 1 ' and ' 0 ' are used as the truth value 'true' and 'false' respectively. As variables for Boolean valuations of BL we shall use ' β ' with or without subscripts. Then we define for all formulas A of BL: $\hat{A} = \{\beta \mid \beta(A) = 1\}$. Thus, \hat{A} is the set of all Boolean valuations of BL which make A true; \hat{t} is the set of all Boolean valuations of BL and \hat{k} is the null set; we have further e.g. $\hat{A} \hat{\wedge} \hat{B} = \hat{A} \cap \hat{B}$, and the fact that A logically implies B can be simply expressed by ' $\hat{A} \subseteq \hat{B}$ '.

1.2 *The Syntax of DSDL3.* The concept of formula is then recursively defined for DSDL3 by the following rules:

- F1: If A and B are formulas of BL, then ' $O(B|A)$ ' is a formula of DSDL3.
- F2: If D and E are formulas of DSDL3, so are ' $\neg D$ ', ' $(D \wedge E)$ ', ' $(D \vee E)$ ', ' $(D \rightarrow E)$ ' and ' $(D \leftrightarrow E)$ '.
- F3: ' $P(B|A)$ ' will be used as an abbreviation for ' $\neg O(\neg B|A)$ '.

We shall employ the usual rules for omitting parentheses. As will be noticed, the syntax of DSDL3 is a rather weak one; no mixed formulas (e.g. ' $p \wedge O(q|p)$ ') and not iterations of the deontic operators are allowed. ' $O(B|A)$ ' is to be read as ' B is obligatory under circumstances A ' or, in order to bring out the purely descriptive character of the formulas of DSDL3, as "according to a not mentioned, but throughout fixed person or institution (legal, social, political or moral) X , B is obligatory under circumstances A ' or simply as 'This X wants B to be the case under circumstances A '.

1.3 *The Semantics of DSDL3.* With the aid of the given semantical notions of BL we are in a position to define the central semantical concepts for DSDL3:

A relation \leq is a connected preference relation for BL (CPR) iff the following five conditions are fulfilled:

- P1: The field of \leq is a subset of \hat{t} , i.e. $\leq \subseteq \hat{t} \times \hat{t}$,
- P2: \leq is reflexive, i.e. for all $\beta \in \hat{t}$ $\beta \leq \beta$,
- P3: \leq is transitive, i.e. for all $\beta_1, \beta_2, \beta_3 \in \hat{t}$, if $\beta_1 \leq \beta_2$ and $\beta_2 \leq \beta_3$, then $\beta_1 \leq \beta_3$,
- P4: \leq is connected, i.e. for all $\beta_1, \beta_2 \in \hat{t}$ we have $\beta_1 \leq \beta_2$ or $\beta_2 \leq \beta_1$ (or both),

P5: If \hat{A} is not empty, so is $\{\beta | \beta \in \hat{A} \wedge \wedge \beta' (\beta' \in \hat{A} \rightarrow \beta' < \beta)\}$.

For any $\hat{A} \subseteq \hat{t}$ and any CPR $<$ we define $\text{Max}_{<}(\hat{A})$ to be the set appearing in P5, i.e. the set $\{\beta | \beta \in \hat{A} \wedge \wedge \beta' (\beta' \in \hat{A} \rightarrow \beta' < \beta)\}$. This is the set of maximal members of \hat{A} with respect to the reflexive order relation $<$. P5 then excludes, so to say, that the wishes of a person or institution who has a CPR grow unceasingly.

Now we define the concept of valuation for DSDL3: Φ is the valuation of DSDL3 belonging to the CPR $<$ iff the following two conditions are fulfilled:

- V1: Φ is a Boolean valuation of the set of all formulas of DSDL3,
 V2: $\Phi(O(B|A)) = 1$ iff $\text{Max}_{<}(\hat{A}) \subseteq \hat{B}$.

The reader will easily verify that there is exactly one valuation of DSDL3 belonging to $<$, for every CPR $<$; so the use of the definite article is justified. In the following the valuation of DSDL3 belonging to the CPR $<$ is denoted by $\Phi_{<}$. Intuitively V2 says that according to the CPR $< B$ is obligatory under circumstances A iff B excludes no possible world (i.e. Boolean valuation of BL) in \hat{A} which is most preferable or most ideal within \hat{A} according to $<$. Equivalently, B is permitted under circumstances A according to $<$ iff there is at least one possible world which is most preferable or most ideal within \hat{A} according to $<$ and in which B is true.

Now, a formula D of DSDL3 is satisfiable iff there is at least one CPR $<$ such that $\Phi_{<}(D) = 1$. A formula D of DSDL3 is valid iff $\Phi_{<}(D) = 1$ for all CPR $<$.

1.4 *The Calculus of DSDL3*. Hansson offers only this semantics and some valid and invalid formulas.⁴ But one does not yet see clearly which are the valid formulas and which the invalid ones. For this reason I have proved the following calculus to be correct and complete with respect to the given semantics:

First we have the axioms of the standard system of ordinary deontic logic carried over to the dyadic case:

- A0: $O(A|A)$
 A1: $\neg O(k|A)$, if $\hat{A} \neq \hat{k}$
 A2: $O(B \wedge C|A) \leftrightarrow O(B|A) \wedge O(C|A)$
 A3: $O(B|A) \leftrightarrow O(B'|A')$, if $\hat{A} = \hat{A}'$ and $\hat{B} = \hat{B}'$.

The only specific axiom of dyadic deontic logic is the axiom

$$\text{A4: } P(B|A) \rightarrow (O(C|A \wedge B) \leftrightarrow O(B \rightarrow C|A))$$

which after a bit of reflection the reader will surely find intuitively acceptable.

Then we need of course

$$\text{A5: } \quad \text{a correct and complete set of propositional axioms and rules.}$$

The calculus of DSDL3 then consists of the axioms and rules A0–A5. The concepts of proof, provability or theoremhood, refutability, inference etc. may then be defined for DSDL3 in the usual fashion.

2. THE CORRECTNESS OF THE CALCULUS OF DSDL3

Here we have to show: For each theorem D of DSDL3 we have $\Phi_{\prec}(D) = 1$ for all CPR \prec .

We will prove this by induction on the length of the proof for D . Thus we first assume that D is one of the axioms of the calculus of DSDL3:

(0) Let D be of the form $O(A|A)$. Since $\text{Max}_{\prec}(\hat{A}) \subseteq \hat{A}$ for all CPR \prec per definitionem, we have $\Phi_{\prec}(D) = 1$ for all CPR \prec .

(1) Now let D be of the form $\neg O(k|A)$, where $\hat{A} \neq \hat{k}$. Then $\Phi_{\prec}(D) = 1$ holds for all CPR \prec , since for each CPR \prec we have $\text{Max}_{\prec}(\hat{A}) \neq \hat{k}$ because of P5.

(2) Now let D be of the form of A2. Because of $\widehat{B \wedge C} = \hat{B} \cap \hat{C}$ we have $\text{Max}_{\prec}(\hat{A}) \subseteq \widehat{B \wedge C}$ iff $\text{Max}_{\prec}(\hat{A}) \subseteq \hat{B}$ and $\text{Max}_{\prec}(\hat{A}) \subseteq \hat{C}$, for all CPR \prec ; thus $\Phi_{\prec}(D) = 1$ for all CPR \prec .

(3) If D is of the form of A3, then the assertion is trivially true on account of the assumptions in A3 about the form of D .

(4) Now let D be of the form of A4. Then we have $\Phi_{\prec}(D) = 1$ for all CPR \prec which is seen in the following way:

If $\Phi_{\prec}(P(B|A)) = 0$, there is nothing to prove. So let us assume that $\Phi_{\prec}(P(B|A)) = 1$, i.e. that

$$\text{(a) } \quad \text{Max}_{\prec}(\hat{A}) \cap \hat{B} \neq \emptyset.$$

From this we have to prove that $\text{Max}_{\prec}(\hat{A} \cap \hat{B}) \subseteq \hat{C}$ iff $\text{Max}_{\prec}(\hat{A}) \subseteq \widehat{B \cup C}$. Since $\text{Max}_{\prec}(\hat{A}) \subseteq \widehat{B \cup C}$ iff $\text{Max}_{\prec}(\hat{A}) \cap \hat{B} \subseteq \hat{C}$, this is proved when we

have proved that

$$(b) \quad \text{Max}_{<}(\hat{A} \cap \hat{B}) = \text{Max}_{<}(\hat{A}) \cap \hat{B}, \quad \text{i.e. that}$$

$$(c) \quad \{\beta | \beta \in \hat{A} \cap \hat{B} \wedge \wedge \beta' (\beta' \in \hat{A} \cap \hat{B} \rightarrow \beta' < \beta)\} =$$

$$\{\beta | \beta \in \hat{A} \cap \hat{B} \wedge \wedge \beta' (\beta' \in \hat{A} \rightarrow \beta' < \beta)\}, \quad \text{i.e. that,}$$

since the right side of (c) is obviously a subset of the left side of (c),

$$(d) \quad \{\beta | \beta \in \hat{A} \cap \hat{B} \wedge \wedge \beta' (\beta' \in \hat{A} \cap \hat{B} \rightarrow \beta' < \beta)\} \setminus \{\beta | \beta \in \hat{A} \cap \hat{B} \\ \wedge \wedge \beta' (\beta' \in \hat{A} \rightarrow \beta' < \beta)\} = \emptyset, \text{ i.e. that}$$

$$(e) \quad \{\beta | \beta \in \hat{A} \cap \hat{B} \wedge \wedge \beta' (\beta' \in \hat{A} \cap \hat{B} \rightarrow \beta' < \beta)$$

$$\wedge \vee \beta' (\beta' \in \hat{A} \wedge \neg \beta' < \beta)\} = \emptyset.$$

Now we want to derive a contradiction from the additional assumption that (e) does not hold. So let us assume that there exist β_0 and β_1 with

$$(f) \quad \beta_0 \in \hat{A} \cap \hat{B} \quad \text{and} \quad \wedge \beta' (\beta' \in \hat{A} \cap \hat{B} \rightarrow \beta' < \beta_0) \quad \text{and}$$

$$(g) \quad \beta_1 \in \hat{A} \quad \text{and} \quad \neg \beta_1 < \beta_0.$$

Then we have:

$$(h) \quad \text{Max}_{<}(\hat{A}) \cap \hat{B} = \{\beta | \beta \in \hat{A} \cap \hat{B} \wedge \wedge \beta' (\beta' \in \hat{A} \rightarrow \beta' < \beta)\}$$

(ex definitionem)

$$= \{\beta | \beta \in \hat{A} \cap \hat{B} \wedge \wedge \beta' (\beta' \in \hat{A} \rightarrow \beta' < \beta) \wedge \beta < \beta_0\} \quad (\text{with (f)})$$

$$= \{\beta | \beta \in \hat{A} \cap \hat{B} \wedge \wedge \beta' (\beta' \in \hat{A} \rightarrow \beta' < \beta) \wedge \neg \beta_1 < \beta\} \quad (\text{with (g)}$$

$$= \emptyset, \quad \text{and P3)}$$

since $\beta_1 \in \hat{A}$. But (h) is a contradiction to our assumption (a).

(5) Finally, let D be one of the axioms of propositional logic. Then $\Phi_{<}(D) = 1$ holds for all CPR $<$, since each $\Phi_{<}$ is a Boolean valuation of DSDL3.

The induction step is trivial too: Our only inference rules are those of propositional logic. Hence, the fact that for all CPR $<$ $\Phi_{<}$ is a Boolean valuation of DSDL3 assures that if $\Phi_{<}$ assigns '1' to all premisses of any application of any inference rule, then it assigns '1' to the conclusion of that application.

This completes the proof of correctness.

3. THE COMPLETENESS OF THE CALCULUS OF DSDL3

Now we want to prove that each formula D of DSDL3 for which $\Phi_{\leq}(D) = 1$ holds by all $\text{CPR } \leq$ is provable in the calculus of DSDL3. To that end we show that for each formula D of DSDL3 which is not provable in the calculus of DSDL3 there exists a $\text{CPR } \leq$ with $\Phi_{\leq}(D) = 0$. This proof will not be as easy as the last one, in fact it is rather complicated, though it is wholly constructive and uses only elementary means despite its set-theoretic formulation. It is composed of six parts:

(1) Let D be a formula of DSDL3 which is not provable in the calculus of DSDL3. Then $\neg D$ is not refutable in the calculus of DSDL3. Now we form a disjunctive normal form of $\neg D$; one member of this disjunction must then be not refutable in the calculus of DSDL3; let \mathcal{D}_0 be the set of the conjunction members of this disjunction member.

Further, let \mathcal{A} be a set of formulas of BL containing nothing but exactly one formula logically equivalent to each formula of BL built up only of propositional variables of BL occurring in the formulas of \mathcal{D}_0 . Since these are finitely many propositional variables it is obvious that \mathcal{A} is a finite set.

Now a \mathcal{A} -formula is defined to be a formula of DSDL3 of the form $O(B|A)$ or $P(B|A)$ where $A, B \in \mathcal{A}$. Since \mathcal{A} is finite, there are only finitely many \mathcal{A} -formulas. Now, we may assume without loss of generality that \mathcal{D}_0 is a set of \mathcal{A} -formulas. (If \mathcal{D}_0 should not be a set of \mathcal{A} -formulas, then \mathcal{D}_0 is to transform to a set of \mathcal{A} -formulas logically equivalent to \mathcal{D}_0 only by F3 and A3.)

Let \mathcal{D} then be a set of \mathcal{A} -formulas with the following three properties:

- (1.1) $\mathcal{D}_0 \subseteq \mathcal{D}$,
- (1.2) either $P(B|A) \in \mathcal{D}$ or $O(\neg B|A) \in \mathcal{D}$ holds for all $A, B \in \mathcal{A}$,⁵
- (1.3) \mathcal{D} (i.e. the finite conjunction of all members of \mathcal{D}) is not refutable in the calculus of DSDL3.

Such a set may be constructed in the following way: We start from \mathcal{D}_0 . Then we form the set \mathcal{D}'_0 of all \mathcal{A} -formulas which are derivable from \mathcal{D}_0 in the calculus of DSDL3. Thus \mathcal{D}'_0 is not refutable in the calculus of DSDL3. If \mathcal{D}'_0 fulfills (1.2) our construction ends. If there are $A, B \in \mathcal{A}$ with neither $P(B|A) \in \mathcal{D}'_0$ nor $O(\neg B|A) \in \mathcal{D}'_0$, we then define $\mathcal{D}_1 = \mathcal{D}'_0 \cup \{P(B|A)\}$. \mathcal{D}_1 is likewise not refutable in the calculus of DSDL3. Then we form the set \mathcal{D}'_1 of all \mathcal{A} -formulas which are derivable from \mathcal{D}_1 in the calculus of DSDL3, and so on. Since there are only finitely many \mathcal{A} -formulas, this procedure

must stop after finitely many steps. This construction also shows that there really are sets which fulfill (1.1)–(1.3).

(2) Now, the completeness of the calculus of DSDL3 is proved when we find a CPR \prec such that $\Phi_{\prec}(E) = 1$ for all $E \in \mathfrak{D}$. For then we have $\Phi_{\prec}(E) = 1$ for all $E \in \mathfrak{D}_0$ and hence $\Phi_{\prec}(\neg D) = 1$ and $\Phi_{\prec}(D) = 0$.

By defining

$$(2.1) \quad \tilde{A} = \cap \{ \hat{B} \mid O(B|A) \in \mathfrak{D} \} \text{ for all } A \in \mathfrak{A}$$

we reduce our task to constructing a CPR \prec such that

$$(2.2) \quad \text{Max}_{\prec}(\hat{A}) = \tilde{A} \text{ for all } A \in \mathfrak{A},$$

as can be seen in the following way:

The formulas in \mathfrak{D} are of the form $O(B|A)$ or of the form $P(B|A)$. If $O(B|A) \in \mathfrak{A}$, then $\tilde{A} \subseteq \hat{B}$ by (2.1) and $\text{Max}_{\prec}(\hat{A}) \subseteq \hat{B}$ by (2.2), hence $\Phi_{\prec}(O(B|A)) = 1$. If $P(B|A) \in \mathfrak{D}$, then $O(\neg B|A) \notin \mathfrak{D}$ by (1.2); thus we have $\tilde{A} \not\subseteq \neg \hat{B}$, since otherwise $O(\neg B|A)$ would be derivable from $\{O(C|A) \mid O(C|A) \in \mathfrak{D}\}$ by A2 and A3 and hence be in \mathfrak{D} , which contradicts (1.3). Thus $\tilde{A} \cap \hat{B} \neq \emptyset$ holds, which implies $\text{Max}_{\prec}(\hat{A}) \cap \hat{B} \neq \emptyset$ by (2.2), which means that $\Phi_{\prec}(P(B|A)) = 1$.

(3) Before this construction can be realised we have to prove some theorems:

$$(3.1) \quad P(A_2|A_1), \dots, P(A_n|A_{n-1}) \text{ and } P(A_1|A_n) \text{ imply } P(A_n|A_1) \text{ in the calculus of DSDL3.}$$

This we prove by deriving a contradiction from the formulas in the first column of the diagram below. How this derivation works is shown by the double arrows, whereby the deontic axioms essentially used are specified, and B is used as an abbreviation of $A_1 \vee \dots \vee A_n$:

$$\begin{array}{l} O(\neg A_n|A_1) \xrightarrow{A_4} O(\neg A_1|B) \vee O(A_1 \rightarrow \neg A_n|B) \Rightarrow O(A_n \rightarrow \neg A_1|B) \\ \Downarrow \\ P(A_1|A_n) \xrightarrow{A_4} O(\neg A_n|B) \vee \neg O(A_n \rightarrow \neg A_1|B) \Rightarrow O(\neg A_n|B) \\ \Downarrow \\ P(A_n|A_{n-1}) \xrightarrow{A_4} O(\neg A_{n-1}|B) \vee \neg O(A_{n-1} \rightarrow \neg A_n|B) \Rightarrow O(\neg A_{n-1}|B) \\ \Downarrow \\ \vdots \\ \Downarrow \\ P(A_2|A_1) \xrightarrow{A_4} O(\neg A_1|B) \vee \neg O(A_1 \rightarrow \neg A_2|B) \Rightarrow O(\neg A_1|B) \end{array}$$

The members of the last column, excluding the first, yield together

$O(\neg A_1 \wedge \dots \wedge \neg A_n | A_1 \vee \dots \vee A_n)$ by A2. This contradicts A1 by A0 and A2 provided that at least one $\hat{A}_i \neq \hat{k}$. Otherwise there is of course nothing to prove.

A trivial corollary of (3.1) is:

$$(3.2) \quad P(A_2 | A_1), \dots, P(A_n | A_{n-1}), P(A_1 | A_n) \text{ and } P(A_j | A_i) \text{ imply} \\ P(A_i | A_j) \text{ in the calculus of DSDL3 } (i, j = 1, \dots, n).$$

Remark: One obtains a helpful graphic representation of this and other statements by representing the relevant formulas of BL (in this case the A_i , in other cases the \mathfrak{R} -formulas) by points and marking among these points the deontic operators regarded as relations between the obligatory or permitted or forbidden (respectively) and the circumstances.

(4) Now we need one further theorem:

$$(4.1) \quad \text{Let } P(A_2 | A_1) \in \mathfrak{D}, \dots, P(A_n | A_{n-1}) \in \mathfrak{D} \text{ and } P(A_1 | A_n) \in \mathfrak{D}. \\ \text{Then } (\tilde{A}_1 \cup \dots \cup \tilde{A}_n) \cap \hat{A}_k = \tilde{A}_k \text{ } (k = 1, \dots, n).$$

Proof: Trivially $\tilde{A}_k \subseteq (\tilde{A}_1 \cup \dots \cup \tilde{A}_n) \cap \hat{A}_k$ holds. For proving the converse we show first:

$$(4.2) \quad \text{Under the assumptions of (4.1) } \tilde{A}_i \subseteq \widehat{\neg A_j} \cup \tilde{A}_j \text{ for all} \\ i, j = 1, \dots, n.$$

If $O(\neg A_j | A_i) \in \mathfrak{D}$, then (4.2) is trivial. If not, then $P(A_j | A_i) \in \mathfrak{D}$ according to (1.2). This implies $P(A_i | A_j) \in \mathfrak{D}$ by the assumptions, (3.2), (1.2) and (1.3). These two facts yield by A4, (1.2) and (1.3):

$$O(A_j \rightarrow C | A_i) \in \mathfrak{D} \text{ iff } O(C | A_i \wedge A_j) \in \mathfrak{D} \text{ for all } C, \text{ and} \\ O(C | A_i \wedge A_j) \in \mathfrak{D} \text{ iff } O(A_i \rightarrow C | A_j) \in \mathfrak{D} \text{ for all } C.$$

The first implies

$$\tilde{A}_i \subseteq \widehat{\neg A_j} \cup \widetilde{A_i \wedge A_j},$$

the second implies

$$\widetilde{A_i \wedge A_j} \subseteq \tilde{A}_j.$$

These together yield (4.2).

According to (4.2) we have therefore:

$$\tilde{A}_i \subseteq (\widehat{\neg A_1} \cup \tilde{A}_1) \cap \dots \cap (\widehat{\neg A_n} \cup \tilde{A}_n) \text{ for all } i = 1, \dots, n,$$

and hence

$$\tilde{A}_1 \cup \dots \cup \tilde{A}_n \subseteq (\widehat{\neg A_1} \cup \tilde{A}_1) \cap \dots \cap (\widehat{\neg A_n} \cup \tilde{A}_n).$$

In view of the fact that $(\bigcap \hat{A}_k \cup \tilde{A}_k) \cap \hat{A}_k = \tilde{A}_k$, this yields:

$$(\tilde{A}_1 \cup \dots \cup \tilde{A}_n) \cap \hat{A}_k \subseteq (\bigcap \hat{A}_1 \cup \tilde{A}_1) \cap \dots \cap (\bigcap \hat{A}_n \cup \tilde{A}_n) \cap \hat{A}_k \subseteq \tilde{A}_k.$$

(5) Now we are in a position to construct a suitable CPR \prec , i.e. a CPR \prec fulfilling (2.2). To that end we define the following relation in

$$(5.1) \quad A_1 \approx A_2 \text{ iff } P(A_2|A_1) \in \mathfrak{D} \text{ or there are } B_1, \dots, B_m \ (m \geq 1) \text{ such that } P(B_1|A_1) \in \mathfrak{D}, P(B_2|B_1) \in \mathfrak{D}, \dots, P(B_m|B_{m-1}) \in \mathfrak{D} \text{ and } P(A_2|B_m) \in \mathfrak{D}, \text{ and conversely } P(A_1|A_2) \in \mathfrak{D} \text{ or there are } C_1, \dots, C_n \ (n \geq 1) \text{ such that } P(C_1|A_2) \in \mathfrak{D}, P(C_2|C_1) \in \mathfrak{D}, \dots, P(C_n|C_{n-1}) \in \mathfrak{D} \text{ and } P(A_1|C_n) \in \mathfrak{D}.$$

Since $P(A|A)$ is a theorem of DSDL3, we have $P(A|A) \in \mathfrak{D}$ for $A \in \mathfrak{A}$ by (1.2) and (1.3), i.e. \approx is reflexive. Trivially, \approx is symmetrical. By looking at (5.1), perhaps with aid of the suggested graphic representation, it is easily seen that \approx is also transitive. Therefore \approx is an equivalence relation in \mathfrak{A} . Let \mathfrak{A} be divided by \approx into the r equivalence classes $\mathfrak{C}_1, \dots, \mathfrak{C}_r$. Within $\{\mathfrak{C}_1, \dots, \mathfrak{C}_r\}$ we define another relation:

$$(5.2) \quad \mathfrak{C}_j \text{ before } \mathfrak{C}_i \text{ iff } i \neq j \text{ and there are } A \in \mathfrak{C}_i \text{ and } B \in \mathfrak{C}_j \text{ such that } P(B|A) \in \mathfrak{D}.$$

Now we assume that the indication of the \mathfrak{C}_i ($i = 1, \dots, r$) is such that the following holds:

$$(5.3) \quad \text{not } \mathfrak{C}_j \text{ before } \mathfrak{C}_i \text{ for all } i, j = 1, \dots, r \text{ with } i < j.$$

Such an indication is always possible, since by (5.1) and (5.2) it is excluded that there are i_1, \dots, i_k such that \mathfrak{C}_{i_1} before \mathfrak{C}_{i_2} and \dots and $\mathfrak{C}_{i_{k-1}}$ before \mathfrak{C}_{i_k} and \mathfrak{C}_{i_k} before \mathfrak{C}_{i_1} .

At last we define certain sets

$$(5.4) \quad \Omega_k = \cup \{ \tilde{A} | A \in \mathfrak{C}_k \} \ (k = 1, \dots, r) \quad \text{and} \quad \Omega = \hat{i} \bigcup_{k=1}^r \Omega_k \text{ and}$$

a certain relation

$$(5.5) \quad \beta_1 \prec \beta_2 \text{ if and only if there are } i \text{ and } j \text{ with } i \geq j \text{ such that } \beta_1 \in \Omega_i \text{ and } \beta_2 \in \Omega_j, \text{ or if } \beta_1 \in \Omega.$$

This is the desired CPR.

(6) The only thing we still have to show is that the relation \prec defined in (5.5) really is a CPR and fulfills (2.2). This may be done in the following way:

By (5.3), (5.2) and (1.2) we have:

$O(\neg B|A) \in \mathcal{D}$ for all $A \in \mathcal{C}_i$ and $B \in \mathcal{C}_j$, where $i < j$, hence $\tilde{A} \subseteq \neg B$ for all $A \in \mathcal{C}_i$ and $B \in \mathcal{C}_j$, where $i < j$, hence $\cup\{\tilde{A} | A \in \mathcal{C}_i\} \subseteq \neg B$ for all $B \in \mathcal{C}_j$ and all $i < j$, hence

$$(6.1) \quad \hat{B} \cap \Omega_i = \emptyset \text{ for all } B \in \mathcal{C}_j \text{ and all } i < j.$$

Further – and now our efforts are rewarded – we can use (4.1) and conclude, on account of the definition of the \mathcal{C}_j and (5.1), that

$$(6.2) \quad \hat{B} \cap \cup\{\tilde{A} | A \in \mathcal{C}_j\} = \tilde{B} \text{ for all } B \in \mathcal{C}_j, \text{ i.e. that}$$

$$\hat{B} \cap \Omega_j = \tilde{B} \text{ for all } B \in \mathcal{C}_j.$$

Finally, (6.1) yields by $\tilde{B} \subseteq \hat{B}$:

$\tilde{B} \cap \Omega_i = \emptyset$ for all $B \in \mathcal{C}_j$ and all $i < j$, hence $\Omega_j \cap \Omega_i = \emptyset$ for all $i < j$, hence

$$(6.3) \quad \Omega_i \cap \Omega_j = \emptyset \text{ for all } i, j = 1, \dots, r \text{ with } i \neq j.$$

By (6.3) and the definition of \prec in (5.5) it is trivial that \prec is a CPR.

Metaphorically, Ω_1 consists of the most preferable or most ideal possible worlds according to \prec , Ω_2 consists of the second most preferable ones, \dots , Ω_r consists of the second least preferable ones and Ω consists of the least preferable ones, if it is not empty. By (6.1) and (6.2) it is equally trivial that (2.2) holds, i.e. that $\text{Max}_{\prec}(\hat{B}) = \tilde{B}$ for all $B \in \mathcal{A}$ for the CPR \prec defined in (5.5).

This finishes the proof of completeness and thus of adequacy of the calculus of DSDL3.⁶

4. DISCUSSION AND CONCLUSIONS

4.1. Now, correctness and completeness results are usually of some formal interest. But apart from that, did Hansson's semantics really deserve this much effort? I think so. There is no doubt that it represents considerable progress for dyadic deontic logic; for in this field to date intuitions have been converted into axioms in a relatively uncontrolled manner, and where this can lead has been shown impressively by Hansson himself.⁷ But Hansson's semantics does for dyadic deontic logic even more, I think, than Kripkean semantics like Hintikka's⁸ do for (iterated) monadic deontic logic. The reason is, I believe, the following:

Semantical procedures are meant to back our intuitions in certain fields. And surely, in deontic logic, monadic and dyadic, our intuitions are in need

of that support. How well do the semantics in question serve this purpose? If we understand Hansson's semantics, we know what it means that A is obligatory under circumstances B according to some CPR. If we understand Hintikka's semantics for monadic deontic logic, we know what it means that A is obligatory according to some model system. Now these things are at most of indirect interest. Interpreting the obligation operator descriptively, as we do, we are in the first place interested in the notion of A being obligatory (under circumstances B) according to some (not mentioned) person or institution. Hence the semantics serve their purpose only when the CPRs or the model systems can somehow be associated with that person or institution. With CPRs this presents no intuitive difficulties, but with model systems it does. Indeed, in both cases we are thrown back upon our intuitions. But this is no surprise. Somewhere the intuitions must enter the picture. But the ways in which they enter are totally different for Hansson's semantics and Hintikka's. With Hansson's semantics we move from the intuitively unreliable field of the (conditional) obligatory to the more reliable field of preference relations, whereas Hintikka's semantics leads one into a terrain which is in any case no more reliable. And because this is so, Hansson's semantics may be readily discussed on intuitive grounds, as will be done now.

4.2. A familiar objection is concerned with the connectedness condition. As just mentioned, we are not interested in the concept of being obligatory according to a CPR, but rather in the concept of being obligatory according to some person or institution which has a CPR. Therefore, one may argue against Hansson's semantics that it seems inadequate insofar as it demands of that person or institution that it compares any two possible worlds with respect to their preferability (connectedness of CPRs); and it is rather difficult to fulfill this condition. It seems more adequate to define ordinary preference relations (OPR) as fulfilling exactly the conditions P1, P2, P3 and P5 and then proceed in the same way. But, as I will show now, it makes no difference which semantics we accept; the same formulas are valid for both.

First, by examining the correctness proof it is easily seen that all the axioms A0–A5 are valid in the semantics with OPR; at no place there did we need the connectedness condition P4. Consequently the set of theorems of DSDL3 and thus, by the completeness result, the set of valid formulas

with respect to the semantics with CPR is a subset of the valid formulas with respect to the semantics with OPR. The converse, of course, also holds simply because the set of all CPR is a subset of the set of all OPR, so that a formula which is true under all valuations of DSDL3 belonging to a OPR is, a fortiori, true under all valuations of DSDL3 belonging to a CPR.

4.3. Yet in another respect, demanding a CPR of a person or institution seems to be asking too much, because one may have to compare two possible worlds which differ in infinitely many respects, i.e. two Boolean valuations of BL which assign different truth values to infinitely many propositional variables. But we are able to diffuse even this argument, in a certain way. To that end let us define: \leq is a preference relation for BL representable by a finite set of formulas of BL (FPR) iff there exist finitely many formulas A_1, \dots, A_n of BL with $\hat{A}_i \cap \hat{A}_j = \emptyset$ for $i \neq j$ and $\bigcup_{j=1}^n \hat{A}_j = \hat{t}$ such that $\beta_1 \leq \beta_2$ holds iff there are i and j with $i \geq j$ such that $\beta_1 \in \hat{A}_i$ and $\beta_2 \in \hat{A}_j$. Again, as at the end of the completeness proof it is clear that each FPR is a CPR. Now, if we compare this definition with (5.4), (6.3) and definition (5.5) in Section 3, and if we consider that the set \mathcal{D} and all sets \mathcal{C}_i in Section 3 were finite and that consequently formulas A_1, \dots, A_r and A can be found for $\Omega_1, \dots, \Omega_r$ and Ω from (5.4) such that $\hat{A}_k = \Omega_k$ ($k = 1, \dots, r$) and $\hat{A} = \Omega$, then the completeness proof shows that each satisfiable finite set of formulas of DSDL3 is satisfiable even by a valuation belonging to a FPR. That is, we may formulate Hansson's semantics with FPRs instead of CPRs without changing the logic, i.e. the set of valid formulas.

With the FPRs we have not succeeded in avoiding the comparison of two Boolean valuations of BL which assign different values to infinitely many propositional variables, as suggested above. But we have attained something similar: Since each FPR is given by a finite set of formulas of BL (with certain properties) and since the formulas of this set contain only finitely many propositional variables, one has to judge the preferability of each Boolean valuation of BL only with respect to the truth values it assigns to these finitely many propositional variables. Differences in the valuation of the other propositional variables are irrelevant insofar as two Boolean valuations of BL which assign the same values to these finitely many propositional variables automatically have the same preferability.

4.4. Having refuted two possible objections we have not yet saved Hansson's semantics. As I see it, the crucial point is the validity of the formula $O(A|A)$.

Surely most people would accept ' A is obligatory under circumstances A (according to a certain person or institution)' neither as being analytically true nor as criterion of rationality, especially in the case where A is something primarily forbidden, i.e. something forbidden under tautological circumstances. This smacks a little bit like justifying the factual. On the other hand it is perfectly intelligible that according to the semantical determination of the dyadic deontic operators $O(A|A)$ is bound to be valid. So Hansson has either got different intuitions than most other people or something has gone wrong with the formalization of his intuitions.

Now being that Hansson has defended the validity of $O(A|A)$, let us look there. He says: "Obligations of the form $O(A|A)$ and $O(\neg A|A)$ will play a special role in those logics. What does it mean to say $O(\neg A|A)$? Let A be 'Smith robs Jones'. It seems rather pointless to say 'Smith ought to refrain from robbing Jones in the circumstance where he actually robs him'. If Smith *has* robbed Jones, he cannot 'undo' it. He can restore what he robbed (and this is obligatory under circumstance A in normal norm systems) but this act is not the act of refraining from robbing Jones. We may perhaps claim that the sentence in question only means that he should not have done what he did, but then there would be no reason to mention the circumstances; no matter what he actually did, he should not rob Jones."⁹ Then Hansson discusses the same point with another example and concludes thereupon "that formulas like $O(B|A)$ shall never be true if \hat{A} and \hat{B} are disjoint".¹⁰

Unfortunately Hansson's argument is inconclusive in two respects. First, Hansson only argues for not accepting $O(\neg A|A)$ as satisfiable, i.e. for accepting $P(A|A)$ as valid; he has not said anything about the validity of the stronger $O(A|A)$. Secondly – and this is much more important – he may be understood in quite another way than he seems to understand himself. His argument "it seems rather pointless ...", cited above, need not be understood as saying that it is false or impossible or irrational that $\neg A$ is obligatory under circumstances A , it can rather be understood as saying that it is vain and meaningless to say that $\neg A$ is obligatory under circumstances A . This is, as I think, the better of the two interpretations. According to it one has to regard $O(\neg A|A)$ not as logically false, but as meaningless and, consequently, $O(A|A)$ not as valid, but also as meaningless.

That is, I suggest applying Hansson's semantics, only to formulas in which the obligatory state of affairs lies in the future relative to the time the

obligation is in force; or more precisely, that $O(B|A)$ qualifies as prime formula of DSDL3 only if t_2 is later than t_1 and t_1 is later than t_0 where t_2 is the point or interval of time referred to by B , t_1 is the point or interval of time referred to by A , and t_0 is the point or interval of time at which this obligation is in force. Both $O(\neg A|A)$ and $O(A|A)$ would then not qualify as prime formulas and would be outside the scope of Hansson's semantics, the desirability of which was just acknowledged.

Now, why should this be more than a mere ad hoc ruling of the matter and how does this go together with the fact that it is obviously meaningful to say that things ought to be, or have been, different from what they are, or were? First, let me say that for Hansson this could scarcely be a mere ad hoc ruling, for it expresses exactly his initial intuition, which he formulates in this way: "The problem of conditional obligation is what happens if somebody nevertheless performs a forbidden act. Ideal worlds are excluded. But it may be the case that among the still achievable worlds some are better than others. There should then be an obligation to make the best out of the sad circumstances."¹¹ Hansson's formal translation of the last sentence is that (according to a certain person or institution) something is obligatory under certain circumstances if and only if it is true in all possible worlds most preferable within these circumstances, e.g. these possibly very sad circumstances themselves. No doubt, one might interpret the obligation operator in this way; but it would definitely be more adequate to interpret it in the way which seems also to be expressed in the last of the sentences just cited: namely that something is obligatory under certain circumstances if and only if it *continues* these circumstances as optimally as possible. Hence Hansson's intuition cannot be adequately formalized without regard to the temporal relations involved, and above I have only suggested just what these temporal relations are.

And there are also good philosophical reasons for anyone to look upon it as more than a mere ad hoc ruling. At least it is a widespread metaethical view that all ethical judgements, part of which the obligation operator is meant to explicate, are essentially concerned with guiding human action and thus are looking to the future. Dewey, for instance, says that "morality is largely concerned with controlling human nature",¹² and Stevenson is quite explicit about the point in saying "that ethical judgements look mainly to *future actions*".¹³ Similar views could be found in many other places.

If this widely accepted view is really correct, then it is clear that my

suggestion above is not ad hoc, but quite natural – all the more as Hansson's semantics is apparently designed for such future-oriented obligations, as just seen. But that is not to say, and the philosophers quoted would not wish to maintain, that moralizing about the past is meaningless. It is only that this past-oriented use of ethical terms including the obligation operator is a secondary derivative use possibly following a different or more complicated logic. For instance, an important problem accruing to past-oriented obligations is that one has to consider from which of the past events one must start and which may be ignored – provided that one does not wish to ignore past events completely, simply describing a perfectly ideal past; something that could be done even with a monadic obligation operator. Future-oriented conditional obligations, on the contrary, can start from the situation given and say that if this and this happens, that or that should subsequently happen. Of course, it would be nice, if there were one logic for all sentences of the form $O(B|A)$ (and logical compounds thereof). But in looking for such a logic one must be prepared to deal with two apparently quite different uses of the obligation operator. So we may safely conclude that the suggestion above picks out just those sentences to which Hansson's semantics can be meaningfully applied.

The technical realization of this suggestion, however, is another question. This would require a combination of deontic and temporal logic thus involving many technical problems which I do not want to discuss now. This plea for temporal relations is only meant to indicate a way of making dyadic deontic logic intuitively more adequate.

4.5. Finally, I shall show that the correctness and completeness proof also yields the decidability of the set of valid formulas of DSDL3. On the one hand the set of provable formulas of DSDL3 and thus the set of valid formulas of DSDL3 is recursively enumerable, of course. On the other hand the set of satisfiable formulas of DSDL3 is also recursively enumerable, as can be seen in the following way: First, the set of all finite sets M of formulas of BL such that $\cup_{A \in M} \hat{A} = \hat{t}$ and such that for all typographically different $A, B \in M$ $\hat{A} \cap \hat{B} = \emptyset$ holds, is decidable because of the decidability of propositional logic. Hence, the set of all FPRs is recursively enumerable. Thus we gain the desired result by the fact that for each FPR \leq the set of all formulas D of DSDL3 with $\Phi_{\leq}(D) = 1$ is decidable and by the fact that each satisfiable formula can be satisfied by a FPR. But by this we do not have a convenient decision procedure.

NOTES

- ¹ This paper is part of my unpublished master thesis 'Zur deontischen Logik und Wollenslogik', Munich 1973.
- ² Hansson [3]
- ³ Hansson [3], p. 121.
- ⁴ Hansson [3], p. 143ff.
- ⁵ Here, as in the following, a statement of the form $O(C_2|C_1) \in \mathfrak{D}$ is often made without it being assured that $C_1 \in \mathfrak{A}$ and $C_2 \in \mathfrak{A}$. However, in these cases there will always be exactly one $C'_1 \in \mathfrak{A}$ and exactly one $C'_2 \in \mathfrak{A}$ logically equivalent to C_1 and C_2 respectively; that is, such a statement is then to read as $O(C'_2|C'_1) \in \mathfrak{D}$ or $P(C'_2|C'_1) \in \mathfrak{D}$.
- ⁶ Using theorem 7 from Hansson [2] it would have been possible to shorten the completeness proof at the end. For (4.2) yields: Under the assumptions of (4.1) we have $\tilde{A}_i \subseteq \neg A_j \cup \tilde{A}_j$, hence $\tilde{A}_i \cap \hat{A}_j \subseteq \tilde{A}_j$, and therefore $\tilde{A}_i \cap \hat{A}_j \cap \hat{A}_i = \tilde{A}_i \cap \hat{A}_j \subseteq \tilde{A}_j \cap \hat{A}_i$. By putting first $i = k$ and $j = k + 1$, and then $i = k + 1$ and $j = k$ in the last inclusion we obtain:
- (*) Under the assumptions of (4.1) $\tilde{A}_k \cap \hat{A}_{k+1} = \hat{A}_k \cap \tilde{A}_{k+1}$ holds for $k = 1, \dots, n - 1$.
- With (*) it is easily verified that (V, f) with $V = \{\hat{A} | A \in \mathfrak{A}\}$ and $f(\hat{A}) = \tilde{A}$ for all $\hat{A} \in V$ is a M -choice structure in the sense of definition 6 of Hansson [2], which implies by theorem 7 of Hansson [2] that there exists a CPR such as we were looking for.
- Nevertheless my proof seems preferable, since it proceeds constructively, enabling the reader to understand how the desired CPR is really constructed, whereas Hansson's theorem, because of its much greater generality, requires, as inconstructive means, the axiom of choice. Furthermore, my construction is essential for the argument put forward in Section 4.3.
- ⁷ Hansson [3], pp. 133–141.
- ⁸ Hintikka [4], pp. 67–73.
- ⁹ Hansson [3], p. 141f. I took the liberty of adjusting Hansson's notation to mine.
- ¹⁰ Hansson [3], p. 142.
- ¹¹ Hansson [3], p. 143.
- ¹² Dewey [1], p. 1.
- ¹³ Stevenson [5], p. 302.

BIBLIOGRAPHY

- [1] John Dewey, *Human Nature and Conduct*, Henry Holt, New York, 1922.
- [2] Bengt Hansson, 'Choice Structures and Preference Relations', *Synthese* 18 (1968), 443–458.
- [3] Bengt Hansson, 'An Analysis of Some Deontic Logic', in Risto Hilpinen (ed.), *Deontic Logic: Introductory and Systematic Readings*, Reidel, Dordrecht, 1971, pp. 121–147, reprinted from *Nous* 3 (1969), 373–398.
- [4] Jaakko Hintikka, 'Some Main Problems of Deontic Logic', in Risto Hilpinen (ed.), *Deontic Logic: Introductory and Systematic Readings*, Reidel, Dordrecht, 1971, pp. 59–104.
- [5] Charles L. Stevenson, *Ethics and Language*, Yale, New Haven, 1944.