On the Strong Completeness of Åqvist's Dyadic Deontic Logic G

Xavier Parent

54 avenue de l'Elisa, 83100 Toulon, France
Xavier.Parent@univ-provence.fr

Abstract. Åqvist's dyadic deontic logic **G**, which aims at providing an axiomatic characterization of Hansson's seminal system DSDL3 for conditional obligation, is shown to be strongly complete with respect to its intended modelling.

Keywords: Conditional obligation, preference-based semantics, strong completeness, DSDL3.

1 Introduction

The present study is mainly concerned with so-called preferential semantics for conditional obligation. These rely on a binary relation, which ranks all possible worlds in terms of comparative goodness or betterness. Structures of this sort seem to have made their first explicit appearance in print with the paper of Hansson [1]. There they are used to give a semantic analysis of contrary-to-duty (or secondary) obligations, which tell us what comes into force when some other (primary) obligations are violated. A number of researchers have followed Hansson's suggestion, providing a more comprehensive investigation of the treatment of contrary-to-duty obligations within a preference-based approach. It is not the purpose of this paper to evaluate such a treatment. The interested reader should consult the relevant literature (see, e.g., [2,3,4,5,6,7,8]).

In what follows, I shall focus on another long-standing problem, that of axiomatizing the logic of conditional obligation as outlined by Hansson in the aforementioned pioneering paper.[1] An important step towards resolving such an issue has been taken by Spohn [9]. There the focus is on the class of models corresponding to the system known as DSDL3, which Hansson wished to be regarded as his 'official' one. An axiomatic characterization of the logic is given, and proved semantically complete with respect to the model theory, in the sense that every formula of this calculus is shown to be provable if and only if it is valid. Metatheorems of this sort are frequently called *weak completeness theorems*− the object of the present paper is to extend Spohn's result to obtain a *strong* completeness theorem for dyadic deontic logic; i.e., I will show that a formula A of

[1] The systems proposed by Hansson (he confidently calls them 'dyadic standard systems of deontic' - DSDL) are purely semantical.

R. van der Meyden and L. van der Torre (Eds.): DEON 2008, LNAI 5076, pp. 189–202, 2008.

this calculus can be deduced from a (possibly infinite) set Γ of formulae if and only if Γ entails A. Reference will be made to Åqvist's axiomatic system **G** (see, e.g., [10,11]). It is essentially a reformulation of the Hansson-Spohn calculus in terms of modal logic. Unless I am mistaken, the strong completeness problem for **G** has not been settled yet. It will here be answered in the affirmative. Moreover, it will be shown that (as conjectured by Åqvist himself) **G** remains complete if the assumption of a linear or total ordering among possible worlds is dropped. It is the assumption that any pair of possible worlds are mutually comparable under the betterness relation: either one is better than the other, or they are of equal value. The fact that such an assumption does no work was already known by Spohn, at least for his reconstruction of DSDL3.

The plan of this paper is as follows. In section 2, I present Åqvist's dyadic deontic system **G**, and its associated semantics. Two classes of models will be discussed, one of them corresponding to Hansson's system DSDL3. In section 3, I introduce the notion of a canonical structure, and prove a number of lemmata, which in section 4 will suffice to establish the desired completeness of the system with respect to the two classes of models.

2 Syntax, Semantics and Proof Theory

The language of **G** has, in addition to a set Prop of propositional variables and the usual Boolean sentential connectives, the following characteristic primitive logical connectives : the alethic modal operators $\Box$ (for necessity) and $\Diamond$ (for possibility) ; and the two dyadic deontic operators $\bigcirc(-/-)$ and $P(-/-)$, which may be read as *'It ought to be that ..., given that ...'* and *'It is permitted that ..., given that ...'*, respectively. The set $\mathcal{L}$ of well-formed formulae (wffs) is defined in the usual way. There are no restrictions as to iterations of dyadic deontic operators and modal ones.

The system comes with a possible worlds semantics *à la* Kripke. I begin with the idea of an H-model ('H' is mnemonic for Hansson), by which I understand a structure

$$\mathcal{M} = (W, \succeq, V)$$

in which

(i) $W \neq \emptyset$ (W is a set of 'possible worlds')
(ii) $\succeq \subseteq W \times W$ (Intuitively, $\succeq$ is a betterness or comparative goodness relation; '$x \succeq y$' can be read as 'world x is at least as good as world y'.)
(iii) V : Prop $\rightarrow \mathcal{P}(W)$ (V is an assignment, which associates a set of possible worlds to each propositional letter p).

I write $\mathcal{M} \models_x A$ to mean that *sentence A is true at world x in $\mathcal{M}$*. Such a notion is defined in the usual way except that, for $x, y \in W$,

$$\mathcal{M} \models_x \Box A \text{ iff } \forall y \ (\mathcal{M} \models_y A)$$
$$\mathcal{M} \models_x \Diamond A \text{ iff } \exists y \ (\mathcal{M} \models_y A)$$
$$\mathcal{M} \models_x \bigcirc(B/A) \text{ iff } \forall y\, (\, (\, \mathcal{M} \models_y A \,\&\, \forall z(\mathcal{M} \models_z A \Rightarrow y \succeq z)\,) \Rightarrow \mathcal{M} \models_y B\,)$$
$$\mathcal{M} \models_x P(B/A) \text{ iff } \exists y\, (\, (\, \mathcal{M} \models_y A \,\&\, \forall z(\mathcal{M} \models_z A \Rightarrow y \succeq z)\,) \,\&\, \mathcal{M} \models_y B\,)$$

The clauses for $\Box$ and $\diamond$ are self-explanatory. These modalities are interpreted by the relation $W \times W$, and thus correspond to the universal modalities further studied by Goranko and Passy [12] among others. In fact, these modalities are not part of Hansson's account. Informally speaking, the evaluation rule for the obligation operator says that $\bigcirc(B/A)$ is true at a world x in $\mathcal{M}$ just in case B is true at all among the *best* (according to $\succeq$) worlds satisfying A. The evaluation rule for the permission operator is obtained by replacing the universal quantifier (ranging over the set of best A-worlds) with the existential one. It is worth noticing that both evaluation rules are formulated in terms of what is sometimes called *optimal* or *last* elements. These are members of S that are at least as good as any other element of S. Formally:

$$y \in \text{opt}_{\succeq}(S) \quad \Leftrightarrow \quad y \in S \ \ \& \ \ y \succeq z \text{ for all } z \in S$$

A last or optimal element of S is, thus, an upper bound of S that is contained in S.[2]

The comparative goodness relation $\succeq$ may be constrained by suitable conditions as desired. The following two classes of models will be discussed further throughout this paper. One is the class of (Åqvist's terminology) H_3-models. In such models, the relation $\succeq$ satisfies the following restrictions:

- reflexivity:
  For all $x \in W, x \succeq x$ (δ_1)
- limitedness:
  If $[\![A]\!]^{\mathcal{M}} \neq \emptyset$ then $\{x \in [\![A]\!]^{\mathcal{M}} : (\forall y \in [\![A]\!]^{\mathcal{M}})\, x \succeq y\} \neq \emptyset,$ (δ_2)
  where $[\![A]\!]^{\mathcal{M}}$ is $\{x \in W : \mathcal{M} \models_x A\}$, the 'truth-set' of A in $\mathcal{M}$
- transitivity:
  For all $x, y, z \in W, x \succeq y$ and $y \succeq z$ entail $x \succeq z$ (δ_3)

The class of H_3-models will henceforth be denoted $\mathcal{H}_3$.

The other class of structures studied in this paper is the class of (Åqvist's terminology) strong H_3-models. This class of models corresponds to Hansson's

[2] This is a non-trivial alteration of the account initially proposed by Hansson [1, pp. 143-6]. He works with so-called *maximality* under the strict order induced by $\succeq$. For a given y in S to qualify for the set of maximal elements of S, no other z in S must be strictly better than y. Formally: $y \in \max_{\succ}(S) \Leftrightarrow (y \in S \ \ \& \ \ \nexists\, z \in S\ (z \succ y))$. Here $\succ$ denotes the 'strengthened converse complement' of $\succeq$, defined by $z \succ y$ iff $z \succeq y$ and $y \not\succeq z$. So the previous definition can be rephrased as:

$$y \in \max_{\succeq}(S) \quad \Leftrightarrow \quad y \in S \ \ \& \ \ \forall z \in S\ (\, z \succeq y \Rightarrow y \succeq z\,)$$

Clearly, $\text{opt}_{\succeq}(S) \subseteq \max_{\succeq}(S)$, but not generally the converse. In particular, the maximal set will not necessarily match the optimal set if S is only partially ordered by $\succeq$. The notions of 'optimality' and 'maximality' are more fully discussed by Sen [13].

official system DSDL3. In such models, the following additional constraint is placed on $\succeq$:

- strong connectedness (totalness, or linearity) :

$$\text{For all } x, y \in W, \text{ either } x \succeq y \text{ or } y \succeq x \qquad (\delta_4)$$

There is, then, no more need to explicitly require $\succeq$ to be reflexive. For (δ_1) follows from (δ_4). The class of strong H_3-models will be denoted by $\mathcal{H}_3{}^s$.

Care should be taken with the limitedness condition (δ_2). Its main purpose is to forbid infinite (ascending) sequences of ever more perfect worlds. (δ_2) should not be confused with the following condition, of which (δ_4) is just a special case:

- well-orderedness:

$$\text{For all } X \subseteq W \text{ if } X \neq \emptyset \text{ then } \{x \in X : (\forall y \in X)\, x \succeq y\} \neq \emptyset \qquad (\delta_2')$$

(δ_2') entails (δ_2). The converse does not hold generally, but only in special cases. One of them is worth mentioning. It is the case where the language is generated from a finite set of atomic propositions. Notoriously, any subset X of the set of all valuations is, then, definable, in the following sense: for all $X \subseteq W$ there exists a formula $A \in \mathcal{L}$ such that $X = [\![A]\!]^{\mathcal{M}}$.[3] Using this further assumption, (δ_2') − and, by the same way, (δ_4) − can easily be derived from (δ_2). The distinction between the class of H_3-models and the class of strong H_3-models vanishes.

The notion of semantic consequence is used in its 'local' sense. A set Γ of formulae is said to be true at a state x in $\mathcal{M}$ (notation: $\mathcal{M} \models_x \Gamma$) if all members of Γ are true at x. A formula A is said to be a (local) semantic consequence of Γ over some class $\mathcal{C}$ of models (notation: $\Gamma \models_{\mathcal{C}} A$) if for all models $\mathcal{M}$ from $\mathcal{C}$, and all points x in $\mathcal{M}$, if $\mathcal{M} \models_x \Gamma$ then $\mathcal{M} \models_x A$. Finally, Γ is said to be satisfiable in $\mathcal{C}$ if there is a model $\mathcal{M}$ from $\mathcal{C}$, and a point x in $\mathcal{M}$, such that $\mathcal{M} \models_x \Gamma$. Brackets will be omitted when Γ is a singleton, i.e. a wff A will be said to be satisfiable in $\mathcal{C}$, if the set $\{A\}$ is satisfiable in $\mathcal{C}$.[4]

In Åqvist [10,11] the proof theory for **G** is defined as shown below:

$$\text{All truth functional tautologies} \qquad (\text{PL})$$

$$\text{S5-schemata for } \Box \text{ and } \Diamond \qquad (\text{S5})$$

$$P(B/A) \leftrightarrow \neg \bigcirc (\neg B/A) \qquad (\text{DfP})$$

$$\bigcirc (B \to C/A) \to (\bigcirc(B/A) \to \bigcirc(C/A)) \qquad (\text{COK})$$

$$\bigcirc (B/A) \to \Box \bigcirc (B/A) \qquad (\text{Abs})$$

$$\Box A \to \bigcirc(A/B) \qquad (\text{CON})$$

---

[3] Cf. Makinson [14, p. 62]. For an example showing that (in the infinite case) a set may not be definable by any formula, see Schlechta [15, p. 29].

[4] As mentioned, the universal modality $\Box$ is not used by Hansson. It is natural to ask whether such a modal operator can be dispensed with, by switching to the so-called global semantic consequence. Perhaps the job done by one can equally be done by the other. This is a topic for future research.

$$\Box(A \leftrightarrow B) \to (\bigcirc(C/A) \leftrightarrow \bigcirc(C/B)) \tag{Ext}$$

$$\bigcirc (A/A) \tag{Id}$$

$$\bigcirc (C/A \wedge B) \to \bigcirc(B \to C/A) \tag{C}$$

$$\Diamond A \to (\bigcirc(B/A) \to P(B/A)) \tag{D*}$$

$$(P(B/A) \wedge \bigcirc(B \to C/A)) \to \bigcirc(C/A \wedge B) \tag{S}$$

$$\text{If } \vdash A \text{ and } \vdash A \to B \text{ then } \vdash B \tag{MP}$$

$$\text{If } \vdash A \text{ then } \vdash \Box A \tag{N}$$

A few comments on the axioms involving the deontic modalities might be in order. (DfP) introduces 'P' as the dual of '$\bigcirc$' in the usual way. (COK) is the conditional analogue of the familiar distribution axiom K. (Abs) is the absoluteness axiom of Lewis [16], and reflects my deliberate choice not to make the ranking world-relative. (CON) is the deontic counterpart of the familiar necessitation rule. (Ext) permits the replacement of equivalent sentences in the antecedent of deontic conditionals. (Id), (C) and (S) are familiar from the literature on non-monotonic logic. (Id) is the deontic analogue of the identity principle. The question of whether this is a reasonable law for deontic conditionals has been much debated. A defence of (Id) can be found in Hansson [1] and Prakken and Sergot [5] − this line of defence is discussed in Parent [17, ch. 3]. (C) corresponds to the so-called 'conditionalization' principle (also referred to as 'the hard half of the deduction theorem'), which is part of Kraus and colleagues' system C for cumulative inference relations (see [18]). Axiom (S) has been introduced into the literature by Spohn [9]. The latter axiom is very reminiscent of the restricted principle of strengthening of the antecedent known as 'rational monotony', which is part of so-called system R (see [19]). This other principle says the following:

$$(P(B/A) \wedge \bigcirc(C/A)) \to \bigcirc(C/A \wedge B) \tag{RM}$$

It is straightforward to show that (S) and (RM) are deductively equivalent given the rest of the system. The deontic version of (RM) is discussed in Goble [20]. (D*) is the conditional analogue of the familiar modal axiom D.

Now the usual notions of theoremhood, deducibility and consistency become available. First, a wff A is said to be a theorem of **G** (written $\vdash_{\mathbf{G}} A$) if A belongs to the smallest subset of wffs that contains every instance of (PL)-(S), and is closed under (MP) and (N). Next, a wff A is said to be deducible in **G** from assumptions Γ (written $\Gamma \vdash_{\mathbf{G}} A$) if there are sentences B_1,..., $B_k \in \Gamma (k \geq 0)$ such that $\vdash_{\mathbf{G}} (B_1 \wedge ... \wedge B_k) \to A$. Finally, a set Γ of sentences is said to be consistent in **G** if $\bot$ is not deducible in **G** from Γ, and inconsistent otherwise. Again, I will omit brackets when Γ is a singleton.

It may be noted that deducibility is compact, in the sense that deducibility from a set of sentences always implies deducibility from a finite portion of that set. This follows at once from the fact that the number of conjuncts in the antecedent of the requisite conditional $(B_1 \wedge ... \wedge B_k) \to A$ is always finite. There is an alternative way of expressing compactness, using consistency: a set Γ of sentences is consistent iff every finite subset of Γ is consistent. The compactness property in these two (equivalent) forms will be used in the completeness proof below.

The soundness result, i.e. that

$$\Gamma \vdash_{\mathbf{G}} A \quad \Rightarrow \quad \Gamma \models_{\mathcal{C}} A \text{ (where } \mathcal{C} \in \{\mathcal{H}_3, \mathcal{H}_3^s\})$$

follows immediately from the definitions involved. Observe that the semantic validity of the Spohn sentence (S) – alias (RM) – depends on (δ_3) alone. This is in contrast to the situation in non-monotonic logics, where the validity of the principle of rational monotony is tightly connected to the assumption that the preference relation is a total order.

The adequacy result, i.e. the converse implication

$$\Gamma \models_{\mathcal{C}} A \quad \Rightarrow \quad \Gamma \vdash_{\mathbf{G}} A$$

takes a little bit of work. It can be established by adapting the standard modal technique of constructing a canonical model (see, for instance, Chellas [21] or Blackburn et al. [22]). The points of the canonical model are maximal consistent sets of sentences. In the present semantical context, the main difficulty is to define the comparative goodness relation in such a way that the semantic truth-conditions for formulae starting with a deontic operator coincide with the set-membership relation between formulae and maximal consistent sets. Åqvist [10,23,11] has developed the technique of so-called *systematic frame constants* as a solution to the latter difficulty. Such a technique provides a means of encoding the betterness relation into the syntax, whereby enabling us to talk and reason about the goodness of the maximal consistent sets in the canonical model. The idea behind the proposed construction (which has roots in Lewis [16]) involves extending the language with a family of propositional constants, $\{Q_i\}_{1 \leq i < \omega}$ (the so-called "systematic frame constants"), which are indexed by the set of positive integers. These are used to attach a "rank" (or "level of perfection") to every maximal consistent sets. Intuitively, Q_1 refers to an ideal situation, Q_2 refers to a sub-ideal one, Q_3 refers to a sub-sub-ideal one, and so forth. The completeness of **G** is established indirectly, by taking a detour through the system $\mathbf{G}_q^\star$ that results from the addition of suitable axiom schemata. Some govern the behavior of all the Q_i, and others their interplay with the normative modalities. Further detail about how the latter system is used to establish the completeness of **G** can be found in Åqvist [10, p. 184-91]. The basic idea is to define a canonical model for **G** using maximal consistency in $\mathbf{G}_q^\star$ as the criterion for worldhood.

The following two observations have motivated my attempt to prove the completeness of **G** by other means. First, on Åqvist's own admission, the desired completeness remains conjectural, because the proposed argument rests on an unestablished lemma. Next, it has been argued by Hansen [24, p. 130] that Åqvist's conjectured proof fails with respect to strong completeness. To make his point, Hansen considers the case of an 'infinitely bad' set, call it Γ_0. It is made up of

- countably many propositional letters p_i $(1 \leq i < \omega)$
- the primary obligation $\bigcirc\neg p_1$, taken as a shorthand for $\bigcirc(\neg p_1/\top)$ (where $\top$ is any tautology), and
- the sequence of ever more specific contrary-to-duty obligations

$$\bigcirc(\neg p_{i+1}/p_1 \wedge p_2 \wedge ... \wedge p_i) \text{ for all } i \text{ such that } 1 \leq i < \omega$$

Hansen points out that Γ_0 is syntactically inconsistent in $\mathbf{G}^\star_q$. The systematic frame constants are indexed by the set of positive integers. Therefore, no systematic frame constants can consistently be added to Γ_0, and thus no rank (or level of ideality) can be assigned to such a set. The reason why should be obvious to the reader. Γ_0 cannot be ideal (i.e. $\Gamma_0 \cup \{Q_1\} \vdash_{\mathbf{G}^\star_q} \bot$), because Γ_0 violates the primary norm $\bigcirc\neg p_1$. Neither can Γ_0 be sub-ideal (i.e. $\Gamma_0 \cup \{Q_2\} \vdash_{\mathbf{G}^\star_q} \bot$), since Γ_0 violates the contrary-to-duty obligation $\bigcirc(\neg p_2/p_1)$. Neither can Γ_0 be sub-sub-ideal (i.e. $\Gamma_0 \cup \{Q_3\} \vdash_{\mathbf{G}^\star_q} \bot$), since it also violates the contrary-to-contrary-to-duty obligation $\bigcirc(\neg p_3/p_1 \wedge p_2)$. And so on indefinitely.

The purpose of the next section is to define a canonical model for **G** directly, without making reference to $\mathbf{G}^\star_q$ or to any other such system. The only notion of consistency I shall use is consistency in **G**. The worlds will be ordered from the standpoint of a given world, by just comparing the extent to which they comply with the obligations contained there.

3 A Canonical Model for G

The following derived rule and theorems are listed for future reference:

$$\text{If } \vdash B \rightarrow C \text{ then } \vdash \bigcirc(B/A) \rightarrow \bigcirc(C/A) \quad \text{(RCOM)}$$

$$\bigcirc(B_1/A) \wedge ... \wedge \bigcirc(B_n/A) \rightarrow \bigcirc(B_1 \wedge ... \wedge B_n/A)(n \geq 2) \quad \text{(AND)}$$

$$\Diamond A \rightarrow \neg \bigcirc (\bot/A) \quad \text{(COD)}$$

$$\bigcirc (C/A \vee B) \rightarrow (\bigcirc(C/A) \vee \bigcirc(C/B)) \quad \text{(DR)}$$

The abbreviations RCOM and COD are taken from Chellas [21]. The proofs of (RCOM), (AND) and (COD) are straightforward, and are omitted. (DR) is the deontic version of the principle usually referred to as 'disjunctive rationality' in the non-monotonic literature. The proof of (DR) requires a little more work. For the details, the reader is asked to consult, e.g., Makinson [25, p. 94]. The derivation presented there appeals to the following additional law, known as 'cautious monotony':

$$(\bigcirc(B/A) \wedge \bigcirc(C/A)) \rightarrow \bigcirc(C/A \wedge B) \quad \text{(CM)}$$

It is perhaps easier to verify that the logic contains (CM) by breaking the argument into cases. If we have $\Diamond A$, then (CM) follows from (RM), since ($D^\star$) allows us to weaken $\bigcirc(B/A)$ into $P(B/A)$. If we do not have $\Diamond A$, then (CM) follows from (Ext), because $\neg\Diamond A$ implies $\Box(A \leftrightarrow (A \wedge B))$.

Definition 1. *Let $W^\star$ be the set of all maximal consistent sets of sentences (MCSs). Let w be a fixed element of $W^\star$. The canonical model generated by w can be defined as the triplet*

$$\mathcal{M}^w = (W, \succeq, V)$$

where:

(i) $W = \{x \in W^\star : \text{ for each } A, \text{ if } \Box A \in w \text{ then } A \in x\}$
(ii) $x \succeq y$ *if and only if either*
  (a) there is no consistent A such that $\{B : \bigcirc(B/A) \in w\} \subseteq y$ (the vacuous case) or
  (b) there is a sentence $A \in x \cap y$ such that $\{B : \bigcirc(B/A) \in w\} \subseteq x$
(iii) V *= the valuation function such that for all p in Prop:*

$$V(p) = \{x \in W : p \in x\}$$

Condition (i) says that W is just the restriction of $W^\star$ to the set of MCSs containing all the wffs A for which $\Box A$ is in the 'generating' world w. This is needed to deal with the alethic modalities. Lemma 1 below clarifies the import of (ii). Intuitively, such a lemma says that the *best* (according to $\succeq$) MCSs among those containing A are precisely those containing all the wffs B for which $\bigcirc(B/A)$ is in the 'generating' world w.

Lemma 1. *If $\succeq$ is defined as in clause (ii)* supra*, then the following two conditions are equivalent (for any x and y in W):*

(I) $A \in x$ *and* $x \succeq y$ *for all y that contains the sentence A*
(II) $\{B : \bigcirc(B/A) \in w\} \subseteq x$

Proof. From the definition of $\succeq$ one sees that (II) entails (I) (given axiom Id). For the converse direction suppose (I) holds, and let B be such that $\bigcirc(B/A) \in w$. We need to show that $B \in x$. Consider the set $\Gamma = \{C : \bigcirc(C/A) \in w\}$. We make the following claims:

Claim 1. Γ is consistent, and can be extended to a maximal consistent set, call it Γ^+.

Verification. The second claim follows from the first (modulo Lindenbaum's lemma). To prove the first claim, suppose Γ is not consistent. By compactness, this means that there is some finite subset $\{C_1, ..., C_n\}$ of Γ such that $\vdash_{\mathbf{G}} (C_1 \wedge ... \wedge C_n) \rightarrow \bot$. By (AND) and (RCOM), $\bigcirc(\bot/A) \in w$. By (COD) and (S5), $\Box\neg A \in w$ so that $\neg A \in x$. Since x is consistent, $A \notin x$, contrary to assumption. We, thus, conclude that Γ is consistent after all.

Claim 2. Γ^+ belongs to W.

Verification. This follows from the fact that, in the presence of (CON), we have

$$\{C : \Box C \in w\} \subseteq \{C : \bigcirc(C/A) \in w\} \subseteq \Gamma^+$$

Claim 3. Γ^+ contains the sentence A.

Verification. Follows from (Id).

We can now apply hypothesis (I) to conclude that $x \succeq \Gamma^+$. By construction, $\{C : \bigcirc(C/A) \in w\} \subseteq \Gamma^+$. Therefore, $x \succeq \Gamma^+$ means that there exists a sentence $D \in x \cap \Gamma^+$ such that

$$\{E : \bigcirc(E/D) \in w\} \subseteq x \tag{1}$$

But we can see that $P(D/A) \in w$. If not, then (DfP) would yield $\neg D \in \Gamma^+$, and thus Γ^+ would be inconsistent. On the other hand, $\bigcirc(B/A) \in w$ entails $\bigcirc(D \to B/A) \in w$. By (S), (Ext) and (C) we conclude $\bigcirc(A \to B/D) \in w$. We can, then, apply (1) to get $A \to B \in x$ and, then, conclude. □

With this established, the rest is easy. First, we lift the 'truth = membership' equation to arbitrary formulae:

Theorem 1 (Truth Lemma). *Let w be a fixed maximal consistent set of sentences, and let $\mathcal{M}^w$ be the canonical model generated by w. Then, for any formula A and x in W,*

$$\mathcal{M}^w \models_x A \text{ iff } A \in x$$

Proof. The proof is by induction on the complexity of A, as measured by the number of logical operators occurring in it. The base case follows from the definition of V in the canonical model. The boolean cases are handled in the usual way, and so are the modal cases. In the modal cases, it might be helpful first to show that (by virtue of the S5 schemata) the relation $R \subseteq W^\star \times W^\star$ defined by putting xRy whenever $\{C : \Box C \in x\} \subseteq y$ is an equivalence relation on the set $W^\star$ of all maximal consistent sets. The fact that R is symmetric, i.e.

$$\{C : \Box C \in x\} \subseteq y \Rightarrow \{D : \Box D \in y\} \subseteq x \tag{2}$$

will be used in the proof of the deontic cases, to which I now turn. I shall focus on the case where A is $\bigcirc(C/B)$. The following is to be established:

$$\mathcal{M}^w \models_x \bigcirc(C/B) \text{ iff } \bigcirc(C/B) \in x$$

For the right-to-left direction, assume $\bigcirc(C/B) \in x$ and let $y \in [\![B]\!]^{\mathcal{M}^w}$ be such that $y \succeq z$ for all z in $[\![B]\!]^{\mathcal{M}^w}$. By the inductive hypothesis, $B \in y$, and $y \succeq z$ for any $z \in W$ such that $B \in z$. Using lemma 1, we get

$$\{B' : \bigcirc(B'/B) \in w\} \subseteq y \tag{3}$$

Now, in the presence of (CON), $\Box \bigcirc (C/B) \in x$ can validly be inferred from $\bigcirc(C/B) \in x$. Using (2), we then get $\bigcirc(C/B) \in w$. From this together with (3), we obtain $C \in y$. By the inductive hypothesis, C is true at y. This shows that $\bigcirc(C/B)$ is true at x as wishes.

For the left-to-right direction, assume that $\bigcirc(C/B)$ is true at x. Using the truth-clause for $\bigcirc$, the inductive hypothesis and lemma 1, and invoking the definition of W, we first get

$$\forall y \in W^\star : (\{E : \Box E \in w\} \subseteq y \;\&\; \{D : \bigcirc(D/B) \in w\} \subseteq y) \Rightarrow C \in y$$

This itself simplifies into (see claim 2 in the proof of lemma 1 above)

$$\forall y \in W^{\star} : \{D : \bigcirc(D/B) \in w\} \subseteq y \Rightarrow C \in y \tag{4}$$

(4) says that C belongs to every maximal consistent extension of

$$\{D : \bigcirc(D/B) \in w\}$$

By the second corollary to Lindenbaum's lemma, C is derivable from that set, i.e.,

$$\vdash_{\mathbf{G}} (D_1 \wedge ... \wedge D_n) \rightarrow C$$

for sentences $D_1, ..., D_n (n \geq 0)$ such that

$$\bigcirc(D_1/B), ..., \bigcirc(D_n/B) \in w$$

Without loss of generality, we can assume that the number of D_i is finite, given compactness. So, using (AND), we first obtain

$$\bigcirc(D_1 \wedge ... \wedge D_n/B) \in w$$

Using (RCOM), we get

$$\bigcirc(C/B) \in w$$

By (CON),

$$\Box \bigcirc (C/B) \in w$$

The definition of W, then, yields the desired conclusion $\bigcirc(C/B) \in x$.

The proof that the theorem holds when A is $P(C/B)$ is similar in structure. Details are omitted. □

We can now check that the comparative goodness relation $\succeq$ of the canonical model has the required properties:

Lemma 2 (Verification Lemma). *If $\succeq$ is taken as in definition 1, then $\succeq$ is limited (δ_2), transitive (δ_3) and strongly connected (δ_4).*

Proof. Limitedness is easily checked. Assume A is true at some x in W. By theorem 1, $A \in x$. Re-running the proof for the '(II) $\Rightarrow$ (I)' direction of lemma 1, claims 1 to 3, we get that W contains at least one y such that $\{A\} \subseteq \{B : \bigcirc(B/A) \in w\} \subseteq y$. Again, by theorem 1, A is true at y. Consider any z at which A is true. By theorem 1, A is in z, and hence in $y \cap z$. By definition 1 (ii), $y \succeq z$ as expected.

Strong connectedness can be proved by *reductio ad absurdum.* Assume $x \not\succeq y$ and $y \not\succeq x$. The former entails that there is a consistent A such that $\{B : \bigcirc(B/A) \in w\} \subseteq y$, whilst the latter implies that there is a consistent C such that $\{B : \bigcirc(B/C) \in w\} \subseteq x$. In virtue of (Id), $A \in y$ and $C \in x$ so that $A \vee C \in y \cap x$. Using (DR), one might then conclude that either

$$\{B : \bigcirc(B/A \vee C) \in w\} \subseteq \{B : \bigcirc(B/A) \in w\} \subseteq y$$

or

$$\{B : \bigcirc(B/A \vee C) \in w\} \subseteq \{B : \bigcirc(B/C) \in w\} \subseteq x$$

Either way we are done.

The proof that $\succeq$ is transitive is a bit tricky. Suppose $x \succeq y$ and $y \succeq z$. Assume $y \succeq z$ means that there exists a $B \in y \cap z$ such that

$$\{B' : \bigcirc(B'/B) \in w\} \subseteq y \qquad (*)$$

(Otherwise, $x \succeq z$ holds trivially.) Given this, $x \succeq y$ entails that there is $C \in x \cap y$ such that

$$\{C' : \bigcirc(C'/C) \in w\} \subseteq x \qquad (**)$$

Clearly, $B \vee C \in x \cap z$. The following is to be established:

$$\{D : \bigcirc(D/B \vee C) \in w\} \subseteq x \qquad (5)$$

Note that $P(C/B \vee C) \in w$. For otherwise, using (DfP), the maximality of w and (DR), we would have either $\bigcirc(\neg C/C) \in w$ or $\bigcirc(\neg C/B) \in w$. None can occur, because a direct application of (**) and (*) would yield the result that C does not belong to the union of x and y $-$ contradicting the assumption made above that C belongs to their intersection. The proof of (5) is then as follows. Assume $\bigcirc(D/B \vee C) \in w$. By (RCOM), $\bigcirc(C \rightarrow D/B \vee C) \in w$. By (S), $\bigcirc(D/(B \vee C) \wedge C) \in w$. By (Ext), $\bigcirc(D/C) \in w$. Using (**), we then get $D \in x$ as wishes. □

The above results are similar to Boutilier's [26] theorem 3.36. There the focus is on belief revision theory. Due to this shift of emphasis, my proofs are different from those presented there.

4 Completeness

I first deal with the totally ordered case. The completeness of **G** with respect to the class $\mathcal{H}_3^s$ of strong H_3-models follows easily from the following:

Theorem 2. *Every consistent set of sentences is satisfiable in* $\mathcal{H}_3^s$.

Proof. Let Γ be any consistent set of sentences. By Lindenbaum's lemma, Γ has a maximal extension, call it Γ_ω. Form the canonical structure generated by Γ_ω, i.e., the structure $\mathcal{M}^{\Gamma_\omega}$ as defined *supra*. By lemma 2, $\mathcal{M}^{\Gamma_\omega}$ belongs to $\mathcal{H}_3^s$. By theorem 1 above, we obtain in particular that for each sentence A

$$\mathcal{M}^{\Gamma_\omega} \models_{\Gamma_\omega} A \text{ iff } A \in \Gamma_\omega$$

Since $\Gamma \subseteq \Gamma_\omega$, we thus have

$$\mathcal{M}^{\Gamma_\omega} \models_{\Gamma_\omega} A \text{ for any } A \text{ in } \Gamma$$

as required. □

Theorem 3 (Completeness, total order case). *For each set of formulae Γ and formula A, the equivalence*

$$\Gamma \vdash_{\mathbf{G}} A \quad \Leftrightarrow \quad \Gamma \models_{\mathcal{H}_3^s} A$$

holds.

Proof. The left-to-right implication is just soundness, so it suffices to check out the right-to-left implication. The argument is standard. Suppose $\Gamma \models_{\mathcal{H}_3^s} A$. Then $\Gamma \cup \{\neg A\}$ is not satisfiable in $\mathcal{H}_3^s$, and hence theorem 2 gives $\Gamma \cup \{\neg A\} \vdash_{\mathbf{G}} \bot$. By simple propositional manipulations, we get $\Gamma \vdash_{\mathbf{G}} A$ as required. □

Theorem 3 is a strong completeness result. As already emphasized, the argument uses compactness many times, and thus no restrictions are placed on the cardinality of the premisse set Γ, at least in principle. I say 'in principle', because the question of whether the above result is immune from a similar objection as the one raised against the systematic frame constants account remains an open problem. As mentioned, a drawback of the latter account is that it fails to assign a rank to the 'indefinitely' bad set as described on p. 194. Take the canonical structure generated from the maximal consistent extension of such a set. It is natural to ask if the definition of $\succeq$ in the canonical model does a better job. That issue calls for further exploration, which will not be attempted in this paper.

I now turn to the partially ordered case. In Åqvist [10, p. 182] and Åqvist [11, p. 249], the question is raised whether **G** is also strongly complete with respect to the class $\mathcal{H}_3$ of models. Based on our previous results we can give a positive answer to this last question.

Corollary 1 (Completeness, partial order case). *For each set of formulae Γ and formula A, the equivalence*

$$\Gamma \vdash_{\mathbf{G}} A \quad \Leftrightarrow \quad \Gamma \models_{\mathcal{H}_3} A$$

holds.

Proof. We already have the soundness part. The proof of the adequacy claim requires only two lines:

$$\begin{aligned} \Gamma \models_{\mathcal{H}_3} A \quad &\Rightarrow \quad \Gamma \models_{\mathcal{H}_3^s} A \\ &\Rightarrow \quad \Gamma \vdash_{\mathbf{G}} A \quad \text{(by th. 3)} \end{aligned}$$

□

The above result shows that within the present set-up the strong connectedness assumption (δ_4) has no import, in the sense that the logic is unaffected by imposing this requirement or not. At first, this may seem surprising. In a way, this is not, given the following:

- H_3-models include the requirement of limitedness (δ_2), and (as mentioned) in the finite case limitedness entails connectedness (δ_4),
- As can easily be verified, the schema $A \geq B \vee B \geq A$ is always valid as long as limitedness is assumed. Here the relation $\geq$ between formulas is defined in the usual fashion, i.e. by the rule: $A \geq B$ iff $\Diamond(A \vee B) \rightarrow P(A/A \vee B)$.

There is more to connectedness than meets the eye. Such a notion is central to, e.g., questions about the possibility of deontic dilemmas, which are key questions within deontic logic today (see, e.g., [27,20]). An in-depth discussion of the role of comparability for the logic of obligation falls outside the immediate scope of this paper, and must be postponed to another opportunity.

One reviewer suggested we start with a form of the limit assumption that does not have the effects described above. Suppose one such form is available. Suppose also we redefine H_3-models by requiring they satisfy this new version of the limit assumption, rather than the old one; the evaluation rules for the deontic modalities are then re-phrased in terms of maximality rather than optimality. We would certainly get a better understanding of the role of comparability within an Hansson-type semantics, by first axiomatizing this class of models, and then investigating the effects of adding the linearity requirement. The question of whether there are in fact alternative forms which the limit assumption might take, is the main focus of my current investigations. The notion of stopperedness from [18] and [28] would not do, but perhaps there are alternative forms available. Expressed in terms of maximality, the stopperedness condition says that whenever $x \in [\![A]\!]^{\mathcal{M}}$ there is a maximal $y \in [\![A]\!]^{\mathcal{M}}$ with $y \succeq x$. Connectedness would still be involved in the framework being used, because stopperedness validates the formula $A \geq B \vee B \geq A$.

Acknowledgments. I would like to thank the anonymous referees for their helpful comments.

References

1. Hansson, B.: An analysis of some deontic logics. Noûs 4, 373–398 (1969)
2. Mott, P.: On chisholm's paradox. Journal of Philosophical Logic 2, 197–211 (1973)
3. Tomberlin, J.: Contrary-to-duty imperatives and conditional obligation. Noûs, 357–375 (1981)
4. Loewer, B., Belzer, M.: Dyadic deontic detachment. Synthese 54, 295–318 (1983)
5. Prakken, H., Sergot, M.: Dyadic deontic logic and contrary-to-duty obligation. In: [29], pp. 223–262
6. van der Torre, L., Tan, Y.H.: The many faces of defeasibility in defeasible deontic logic. In: [29], pp. 79–121

7. Carmo, J., Jones, A.: Deontic logic and contrary-to-duties. In: Gabbay, D., Guenthner, F. (eds.) Handbook of Philosophical Logic, 2nd edn., vol. 8, pp. 265–343. Kluwer Academic Publishers, Dordrecht (2002)
8. Parent, X.: Remedial interchange, contrary-to-duty obligation and commutation. Journal of Applied Non-Classical Logics 3/4, 345–375 (2003)
9. Spohn, W.: An analysis of Hansson's dyadic deontic logic. Journal of Philosophical Logic 4, 237–252 (1975)
10. Åqvist, L.: An Introduction to Deontic logic and the Theory of Normative Systems, Biliopolis, Naples (1987)
11. Åqvist, L.: Deontic logic. In: Gabbay, D., Guenthner, F. (eds.) Handbook of Philosophical Logic, 2nd edn., vol. 8, pp. 147–264. Kluwer Academic Publishers, Dordrecht (2002)
12. Goranko, V., Passy, S.: Using the universal modality: Gains and questions. Journal of Logic and Computation 2, 5–30 (1992)
13. Sen, A.: Maximization and the act of choice. Econometrica 65, 745–779 (1997)
14. Makinson, D.: Bridges from Classical to Nonmonotonic logic. King's College Publications, London (2005)
15. Schlechta, K.: Coherent Systems. Elsevier, Amsterdam (2004)
16. Lewis, D.: Counterfactuals. Blackwell, Oxford (1973)
17. Parent, X.: Non-monotonic Logics and Modes of Argumentation − The Case of Conditional Obligation. PhD thesis, University of Aix-Marseille I, France (2002)
18. Kraus, S., Lehmann, D., Magidor, M.: Nonmonotonic reasoning, preferential models and cumulative logics. Artificial Intelligence 44, 167–207 (1990)
19. Lehmann, D., Magidor, M.: What does a conditional knowledge base entail? Artificial Intelligence 55, 1–60 (1992)
20. Goble, L.: A proposal for dealing with deontic dilemmas. In: Lomuscio, A., Nute, D. (eds.) DEON 2004. LNCS (LNAI), vol. 3065, pp. 74–113. Springer, Heidelberg (2004)
21. Chellas, B.: Modal Logic. Cambridge University Press, Cambridge (1980)
22. Blackburn, P., de Rijke, M., de Venema, Y.: Modal Logic, vol. 53. Cambrigde University Press, Cambridge (2001)
23. Åqvist, L.: A completeness theorem in deontic logic with systematic frame constants. Logique & Analyse 36, 177–192 (1993)
24. Hansen, J.: On relations between Åqvist's deontic system G and van Eck's deontic temporal logic. In: Mc Namara, P., Prakken, H. (eds.) Norms, Logics and Information Systems. Frontiers in Artificial Intelligence and Applications, pp. 127–144. IOS Press, Amsterdam (1999)
25. Makinson, D.: General patterns in nonmonotonic reasoning. In: Gabbay, D., Hogger, C., Robinson, J. (eds.) Handbook of Logic in Artificial Intelligence and Logic Programming, pp. 35–110. Clarendon Press, Oxford (1994)
26. Boutilier, G.: Unifying default reasoning and belief revision in a modal framework. Artificial Intelligence 68, 33–85 (1994)
27. van der Torre, L., Tan, Y.H.: Contrary-to-duty reasoning with preference-based dyadic obligations. Annals of Mathematics and Artificial Intelligence 27(1-4), 49–78 (1999)
28. Makinson, D.: General theory of cumulative inference. In: Reinfrank, M., Ginsberg, M.L., de Kleer, J., Sandewall, E. (eds.) Non-Monotonic Reasoning 1988. LNCS, vol. 346, pp. 1–18. Springer, Heidelberg (1989)
29. Nute, D. (ed.): Defeasible Deontic Logic. Kluwer Academic Publishers, Dordrecht (1997)