Logic Journal of IGPL Advance Access published September 9, 2009

A complete axiom set for Hansson's deontic logic DSDL2

XAVIER PARENT, *Individual and Collective Reasoning Group, Department of Computer Science and Communications, University of Luxembourg.*
E-mail: xavier.parent@uni.lu

Abstract

The main contribution of this paper is a (strong) completeness result for an axiomatization of Hansson [13]'s deontic system DSDL2, whose semantics involves a non-necessarily transitive betterness relation. Reference is made to a deductive system put forth by Åqvist [2, 3].

Keywords: deontic logic; completeness; Hansson's system DSDL2; non-transitive betterness relation; optimality language for dyadic deontic logic

1 Introduction

This paper is concerned with so-called preferential semantics for deontic logic. 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 [13]. There they are used to give a semantical 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., [3, 6, 9, 16, 18, 20, 22, 26, 28]).

In what follows, I shall focus on the problem of how to axiomatize the classes of structures outlined by Hansson in the aforementioned pioneering paper.[1] Previous work has almost exclusively dealt with Hansson's 'official' system DSDL3, which makes relatively strong assumptions about the agent's rationality. A weakly complete axiomatization of DSDL3 can be found in Spohn [24] and Åqvist [2, 3]. This weak completeness result has been strengthened into a strong one in Parent [21]. Here my attention will be devoted to Hansson's weaker system DSDL2. It is much like DSDL3, except that the betterness relation is no longer required to be transitive. The requirement of transitivity has been criticized by a number of moral philosophers and contemporary decision theorists.[2] The question naturally comes into mind if DSDL2 can be axiomatized. Åqvist [2, 3] conjectured an axiomatic basis using an optimality operator language for dyadic deontic logic, whose symbol is "Q" ('optimally', 'ideally', ...).[3] Unfortunately this conjecture has been settled in the negative by Ardeshir

[1]The systems proposed by Hansson (he confidently calls them 'dyadic standard systems of deontic logic' - DSDL) are purely semantical. Syntactical issues such as questions of axiomatization are put to one side.

[2]Cf., e.g., Temkin [25] and Sen [23, sec. 10].

[3]The conditional obligation connective used by Hansson can be defined from that operator, plus other logical apparatus (see section 2 below). The counterpart of Q within conditional logic is the so-called circumstantial

doi:10.1093/jigpal/jzp050

and Nabavi in a recent article in this journal [4]. There it is argued that the proposed axiomatic basis is incomplete, i.e. that there is a formula that is not a theorem of the system even though it is valid in the corresponding class of models. Indeed, the following is such a sentence − this is simply Sen's property γ (see [23]) in modal logical notation:

$$(QA \wedge QB) \rightarrow Q(A \vee B) \tag{γ}$$

The main purpose of this paper is to show that completeness can be regained if (γ) is added to Åqvist's axiom set. The completeness result is a strong completeness theorem for the system. Moreover, the arguments in the proof carry over to the case where no specific restrictions are put on the betterness relation.

Readers should be warned that there is far less standardization in preference semantics than in the usual Kripke-style semantics for deontic logic, and more room for variation. This is due to the fact that there are several factors that must be juggled all at once. In this paper I will stick to the Åqvist account. Those who wish to get a general overview of the possible approaches that can be taken might find it useful to consult Makinson [17] and Goble [11, 12].

The plan of this paper is as follows. Section 2 presents the syntax, semantics and proof theory of the framework being used. Section 3 establishes completeness using canonical models. Section 4 lists a number of open questions that our main result raises.

2 Syntax, Semantics and Proof Theory

The syntax is generated by adding the following three unary modal operators to the syntax of propositional logic: $\Box$ (for necessity); $\Diamond$ (for possibility) ; and Q (for optimality). For QA read: "ideally A" or "A under the best circumstances". The set of well-formed formulae (wffs) is defined in the straightforward way. There are no restrictions as to iterations of 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

$$M = (W, \succeq, v)$$

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) $v: \mathrm{Prop} \rightarrow \mathcal{P}(W)$ (v is an assignment, which associates a set of possible worlds to each propositional letter p).

The definition of truth at a point in a model is as usual except for the modal clauses. I use the following evaluation rules, where x and y are in W:

$$\begin{array}{lll} M, x \models \Box A & \text{iff} & \forall y\,(M, y \models A) \\ M, x \models \Diamond A & \text{iff} & \exists y\,(M, y \models A) \\ M, x \models QA & \text{iff} & M, x \models A \;\&\; \forall y\,(M, y \models A \Rightarrow x \succeq y) \end{array}$$

operator "$\star$" (*ceteris paribus*) due to Åqvist (see [1]), and further discussed by Lewis in section 2.8 of [15], and Humberstone (see [14]) among others.

The first two clauses are self-explanatory. The third one says that QA is true at x whenever x is among the *best* (according to $\succeq$) A-worlds. I shall usually drop reference to M when it is clear what model is intended.

The syntax of Hansson's system DSDL2 is based on a dyadic operator $O(B/A)$ for "B is an obligation conditionally on A" rather than Åqvist's monadic operator Q. Hansson's dyadic operator is clearly definable in the language as $O(B/A) = \Box(QA \to B)$ ("in all the best A-worlds, B is true").

The comparative goodness relation $\succeq$ may be constrained by suitable conditions as desired. I shall focus on the class of (Åqvist's terminology) H_2-models, whose betterness relation fulfills the following two requirements:

- reflexivity:

  For all $x \in W, x \succeq x$ (δ_1)

- limit assumption:

  If $[\![A]\!]^M \neq \emptyset$ then $\{x \in [\![A]\!]^M : (\forall y \in [\![A]\!]^M)\, x \succeq y\} \neq \emptyset$, (δ_2)

  where $[\![A]\!]^M$ denotes the 'truth-set' of A, i.e.

  the set of worlds in which A holds

The class of H_2-models corresponds to Hansson's system DSDL2. Semantic consequence, validity and satisfiability are defined as usual relative to the latter class.

Some readers familiar with Hansson's paper [13] might wonder why (in the introductory remarks) DSDL2 is described as corresponding to the non-necessarily transitive case. Two other systems are discussed by Hansson. One is DSDL1, whose betterness relation is only reflexive. The other is DSDL3, which is obtained from DSDL2 by requiring the betterness relation also be transitive and total ("for all $x, y \in W$, either $x \succeq y$ or $y \succeq x$"). In fact, there is some redundancy in adding the latter requirement. For, in the finite case (i.e. when the language is generated from finitely many propositional letters), (δ_2) entails the total order assumption.[4] It therefore makes sense to say that DSDL2 is distinguished from DSDL3 by just letting the transitivity condition go. Note, however, that the question of whether (δ_2) also entails the total order assumption in the infinite case remains an open problem.

I now briefly turn to the proof theory. The axioms are the tautologies, the S5-schemata for $\Box$ and $\Diamond$, and the sentences

$$\Box(A \equiv B) \to \Box(QA \equiv QB) \qquad \text{(RE)}$$

$$QA \to A \qquad \text{(T)}$$

$$QA \wedge B \to Q(A \wedge B) \qquad \text{(Ch)}$$

$$\Diamond A \to \Diamond QA \qquad \text{(CD)}$$

$$(QA \wedge QB) \to Q(A \vee B) \qquad (\gamma)$$

The rules are modus ponens and the rule of necessitation for $\Box$. ('Ch' is mnemonic for Chernoff [8], and the abbreviation 'CD' is taken from Chellas [7].) The axiomatic system comprising all instances of these axiom schemata and rules will be called simply Θ. Theoremhood, deducibility, and consistency are defined as usual relative to the latter system.

[4]See Parent [21, p. 192]. The argument presented there holds under the assumption that the universe contains no worlds x and y that are duplicates in the sense that x and y agree on exactly the same wffs.

Ardeshir and Nabavi [4] showed that (γ) cannot be proved from the remaining axioms of Θ. This observation settled an open problem put forth by Åqvist [2, p. 174] and [3, p. 243]. It is the problem of whether the system obtained by deleting (γ) from our Θ [5] provides a complete axiomatization of DSDL2.

Soundness of Θ with respect to the class of H_2-models is easy to demonstrate, and may be left to the reader. Completeness is shown in the next section using the canonical model method. Definitions and facts relative to maximal consistent sets (mcs's, for short) will be taken for granted (see, e.g., Chellas [7]).

3 Completeness Result

I begin by stating formally the principle result that is to be established here.

Theorem 3.1.
Θ is (strongly) complete with respect to the class of H_2-models.

Proof. The key idea is to work with a point-generated canonical model. Let $W^\star$ be the set of all mcs's of Θ. Let w be a fixed element of $W^\star$. Define the canonical model $M^w = (W, \succeq, v)$ generated by w as follows:

(i) $W = \{x \in W^\star : \{A : \Box A \in w\} \subseteq x\}$;
(ii) $\succeq$ is the binary relation on W defined by

$$x \succeq y \begin{cases} \text{if either} & x = y \\ \text{or} & \{A : QA \in x\} \cap y \neq \emptyset \end{cases}$$

(iii) v is the valuation function defined by $v(p) = \{x \in W : p \in x\}$ for all p in Prop.

Obviously $\succeq$ as so defined is reflexive. Before establishing that $\succeq$ also fulfills the limit assumption (δ_2), it might help to extend the 'truth = membership' equation to arbitrary formulae.

Lemma 3.2 (Truth Lemma)
For all formulae A and x in W,

$$M^w, x \models A \text{ iff } A \in x$$

Proof. By induction on A; we only show the inductive case when $A = QB$, supposing the lemma to hold for B.

For the right-to-left direction, let $QB \in x$. By (T), $B \in x$. Thus, by the inductive hypothesis, $x \models B$. Let $y \in W$ be such that $y \models B$. By the inductive hypothesis again, $B \in y$. So, by definition of $\succeq$, $x \succeq y$. This suffices for $x \models QB$.

For the left-to-right direction, we argue contrapositively. We assume that $QB \notin x$, and prove that $x \not\models QB$. If $x \not\models B$, we are done. So assume $x \models B$. By the inductive hypothesis, $B \in x$. There are two possibilities:

Case 1: $\nexists A$ s.t. $QA \in x$. Put $y^- = \{A : \Box A \in w\} \cup \{QB\}$. Note that $\{A : \Box A \in w\} \neq \emptyset$ since $\top \in \{A : \Box A \in w\}$ because $\Box \top \in w$ by necessitation. Now suppose, to reach a contradiction, that

[5] Here I am referring to Åqvist's system $S5^N_{Qm0}$+ every β_i with $i = 0, 1, 2, 3$.

y^- is not consistent. Then for $m \geq 1$ there are sentences A_1, ..., A_m in $\{A : \Box A \in w\}$ such that $\vdash (A_1 \wedge \ldots \wedge A_m) \rightarrow \neg QB$. By S5 we get $\vdash (\Box A_1 \wedge \ldots \wedge \Box A_m) \rightarrow \Box \neg QB$. For each A_i $(1 \leq i \leq m)$, $\Box A_i \in w$; hence $\Box A_1 \wedge \ldots \wedge \Box A_m \in w$, and so $\Box \neg QB \in w$, i.e., $\neg \Diamond QB \in w$. So by (CD) $\neg \Diamond B \in w$, i.e., $\Box \neg B \in w$. But then by construction $\neg B \in x$, contrary to the consistency of x. Therefore y^- must be consistent. By Lindenbaum's lemma, y^- has a maximal consistent extension, call it y. Obviously $y \in W$ and $B \in y$ by the (T) schema. The inductive hypothesis gives $y \models B$. From supposition $\{A : QA \in x\} = \emptyset$, and so a fortiori $\{A : QA \in x\} \cap y = \emptyset$. Furthermore, $x \neq y$ because $QB \in y$ and $QB \notin x$. So $x \not\succeq y$, which suffices for $x \not\models QB$.

Case 2: $\exists A$ s.t. $QA \in x$. The set $z^- = \{A : \Box A \in w\} \cup \{\neg A : QA \in x\} \cup \{B\}$ is consistent. For suppose otherwise. Then for $m, n \geq 1$ there are sentences A_1, ..., A_m in $\{A : \Box A \in w\}$ and sentences $\neg B_1$, ..., $\neg B_n$ in $\{\neg A : QA \in x\}$ such that

$$\vdash (A_1 \wedge \ldots \wedge A_m \wedge \neg B_1 \wedge \ldots \wedge \neg B_n) \rightarrow \neg B$$

Equivalently,

$$\vdash (A_1 \wedge \ldots \wedge A_m) \rightarrow (B \rightarrow (B_1 \vee \ldots \vee B_n))$$

By S5,

$$\vdash (\Box A_1 \wedge \ldots \wedge \Box A_m) \rightarrow \Box (B \rightarrow (B_1 \vee \ldots \vee B_n))$$

For each A_i $(1 \leq i \leq m)$, $\Box A_i \in w$; hence $\Box A_1 \wedge \ldots \wedge \Box A_m \in w$, and thus $\Box (B \rightarrow (B_1 \vee \ldots \vee B_n)) \in w$. Hence by the properties of $\Box$ again,

$$\Box \big(((B_1 \vee \ldots \vee B_n) \wedge B) \equiv B \big) \in w$$

Using (RE) it follows that

$$\Box \big(Q((B_1 \vee \ldots \vee B_n) \wedge B) \equiv QB \big) \in w$$

So by construction

$$Q((B_1 \vee \ldots \vee B_n) \wedge B) \equiv QB \in x$$

But $QB \notin x$. So

$$Q((B_1 \vee \ldots \vee B_n) \wedge B) \notin x$$

Now, for each B_i $(1 \leq i \leq n)$, $QB_i \in x$. In particular, $QB_1, QB_2 \in x$; hence $QB_1 \wedge QB_2 \in x$. A first application of (γ) yields $Q(B_1 \vee B_2) \in x$, so that $Q(B_1 \vee B_2) \wedge QB_3 \in x$. A second application of (γ), then, yields $Q(B_1 \vee B_2 \vee B_3) \in x$. The number of B_i's is finite.[6] Reiterating this argument n times we arrive at the result that $Q(B_1 \vee \ldots \vee B_n) \in x$. But $B \in x$; hence $Q(B_1 \vee \ldots \vee B_n) \wedge B \in x$. Using (Ch) we may further move B inside the scope of Q obtaining the contradiction that $Q((B_1 \vee \ldots \vee B_n) \wedge B) \in x$. Therefore z^- is consistent, and has a maximal consistent

[6] It might be worth clarifying why. This follows at once from the standard characterization of deducibility and consistency in terms of theoremhood. If there is a proof that z^- is inconsistent, it starts from finitely many premises.

extension, z. It is clear that $z \in W$ and $B \in z$, and thus the inductive hypothesis yields $z \models B$. Choose any A such that $QA \in x$ – we can make such a choice since we are assuming that x contains at least one sentence of this form. By construction $\neg A \in z$, and thus $A \notin z$ by consistency of z. This shows that $\{A : QA \in x\} \cap z = \emptyset$. This also shows that $x \neq z$ since $A \in x$ by the (T) schema. So $x \not\succeq z$, and thus $x \not\models QB$ as required. This establishes Lemma 3.2. ∎

It remains to verify that $\succeq$ fulfills the limit assumption (δ_2). Suppose $[\![A]\!]^{M^w} \neq \emptyset$, and let x be in $[\![A]\!]^{M^w}$. By Lemma 3.2, $A \in x$. Put $y^- = \{B : \Box B \in w\} \cup \{QA\}$. A similar argument as before – it rests essentially on (CD) – yields that y^- is consistent, and has a maximal consistent extension, y. Obviously, $y \in W$, and $A \in y$ given the (T) schema. Using Lemma 3.2, it follows that $y \models A$. Let z be such that $z \models A$. By Lemma 3.2, $A \in z$, and so by definition of $\succeq$, $y \succeq z$ since $QA \in y$. This shows that (δ_2) is met.

Theorem 3.1, the principle result of this paper, now follows in the usual way. The argument is standard. Details are omitted. ∎

Let us now briefly consider the system Λ obtained by just removing (CD) from our Θ. As can easily be verified, Λ is sound with respect to the class of all H-models, and the class of H-models in which the betterness relation is reflexive (DSDL1). For completeness with respect to the first class of models, we can adapt the argument for Theorem 3.1 as follows. Let $M^w = (W, P, v)$ where

- W is the set of all mcs's of Λ containing every wff A for which $\Box A$ is in w;
- P is defined thus: xPy if and only if $\{A : QA \in x\} \cap y \neq \emptyset$;
- v is as usual.

To verify that Lemma 3.2 holds for Λ, we need only run through the proof of the modal part of the inductive step again, making one small variation. Note, first, that in the verification of the left-to-right direction, there is no more need to show that $x \neq y$ and $x \neq z$. Then, when handling case 1 above, put $y^- = \{A : \Box A \in w\} \cup \{B\}$. The consistency of y^- follows at once from $y^- \subseteq x$ (and the consistency of x).

It is notable that the proposed method, which works so smoothly for DSDL2 and the class of all H-models, cannot be extended to DSDL1. Let us see why not. First, there is no guarantee that P as just defined is reflexive. The simplest way around is to work with the reflexive closure of P, as we did in the proof of Theorem 3.1. But, then, when re-running the proof of Lemma 3.2, we get stuck on case 1, because there is no way to ensure that $x \neq y \supseteq \{A : \Box A \in w\} \cup \{B\}$. To rule out case 1 from the outset, it suffices to work with a submodel of the original canonical model, obtained by just removing from W all the mcs's that do not contain any formula of the form QA.[7] But, then, the argument for case 2 does not go through, because there is in general no guarantee that the mcs z contains one wff of this form. In other words, there is no guarantee that z is part of the submodel we are working with. Thus here is an unexpected difficulty.

4 Open Questions

- Can an analogous completeness result be established for DSDL1?
- Does the system DSDL2 have the finite model property?

[7] Essentially the same trick is used by Bezzazi et al. [5, p. 615-616] in the context of preference-based semantics for non-monotonic logics, but for different purposes.

- Can a similar completeness result be proven for Åqvist's system **F** for conditional obligation (see [3, p. 248]), or one close to it?

Acknowledgements

Thanks to two anonymous referees for valuable comments. I have also greatly benefited from discussions with L. Goble, D. Makinson, and L. van der Torre. Support from the National Research Fund (*F.N.R.*) of Luxembourg (grant PDR-08-013) is gratefully acknowledged.

References

[1] L. Åqvist. Modal logic with subjunctive conditionals and dispositional predicates. *Journal of Philosophical Logic*, 2(1):1–76, 1973.

[2] L. Åqvist. *An Introduction to Deontic logic and the Theory of Normative Systems.* Biliopolis, Naples, 1987.

[3] L. Åqvist. Deontic logic. In Gabbay and Guenthner [10], pages 147–264.

[4] M. Ardeshir and F. Nabavi. On some questions of L. Åqvist. *Logic Jnl IGPL*, 14(1):1–13, 2006.

[5] H. Bezzazi, D. Makinson, and R. Pérez. Beyond rational monotony: some strong non-horn rules for nonmonotonic inference relations. *Journal of Logic and Computation*, 7(5):605–631, 1997.

[6] J. Carmo and A.J.I. Jones. Deontic logic and contrary-to-duties. In Gabbay and Guenthner [10], pages 265–343.

[7] B.F. Chellas. *Modal Logic*. Cambridge University Press, Cambridge, 1980.

[8] H. Chernoff. Rational selection of decision functions. *Econometrica*, 22:422–443, 1954.

[9] D.M. Gabbay. Reactive kripke models and contrary-to-duty obligations. In van der Meyden and van der Torre [27], pages 155–173.

[10] D.M. Gabbay and F. Guenthner, editors. *Handbook of Philosophical Logic*, volume 8. Kluwer Academic Publishers, Dordrecht, Holland, 2nd edition, 2002.

[11] L. Goble. Preference semantics for deontic logics. Part I: Simple models. *Logique & Analyse*, 46(183-184):383–418, 2003.

[12] L. Goble. Preference semantics for deontic logics. Part II: Multiplex models. *Logique & Analyse*, 47(185-188):335–363, 2004.

[13] B. Hansson. An analysis of some deontic logics. *Noûs*, 3:373–398, 1969.

[14] I.L. Humberstone. Two merits of the circumstantial operator language for conditional logics. *Australasian Journal of Philosophy*, 56(1):21–24, 1978.

[15] D.K. Lewis. *Counterfactuals*. Blackwell, Oxford, 1973.

[16] B. Loewer and M. Belzer. Dyadic deontic detachment. *Synthese*, 54(2):295–318, 1983.

[17] D.C. Makinson. Five faces of minimality. *Studia Logica*, 52(3):339–379, 1993.

[18] P.L. Mott. On Chisholm's paradox. *Journal of Philosophical Logic*, 2(2):197–211, 1973.

[19] D. Nute, editor. *Defeasible Deontic Logic*. Kluwer Academic Publishers, Dordrecht, 1997.

[20] X. Parent. Remedial interchange, contrary-to-duty obligation and commutation. *Journal of Applied Non-Classical Logics*, 13(3/4):345–375, 2003.

[21] X. Parent. On the strong completeness of Åqvist's dyadic deontic logic G. In van der Meyden and van der Torre [27], pages 189–202.

[22] H. Prakken and M.J. Sergot. Dyadic deontic logic and contrary-to-duty obligation. In Nute [19], pages 223–262.

[23] A. Sen. Choice functions and revealed preference. *Review of Economic Studies*, 38:307–317, 1971.

[24] W. Spohn. An analysis of Hansson's dyadic deontic logic. *Journal of Philosophical Logic*, 4(2):237–252, 1975.

[25] L.S. Temkin. Intransitivity and the mere addition paradox. *Philosophy and Public Affairs*, 16(2):138–187, 1987.

[26] J.E. Tomberlin. Contrary-to-duty imperatives and conditional obligation. *Noûs*, 15(3):357–375, 1981.

[27] R. van der Meyden and L. van der Torre, editors. *Deontic Logic in Computer Science.* 9th International Conference, DEON 2008, Luxembourg, July 15-18, 2008. Proceedings. Number 5076 in LNAI. Springer, Berlin Heidelberg, 2008.

[28] L.W.N. van der Torre and Y.-H. tan. The many faces of defeasibility in defeasible deontic logic. In Nute [19], pages 79–121.

Received 5 September 2008. Revised 27 June 2009.