Proceedings of the 23rd International Conference on Principles of Knowledge Representation and Reasoning
Main Track

Defeasible Conditional Obligation in a Two-tiered Preference-based Semantics

Xavier Parent
Institute of Logic and Computation, TU Wien, Vienna, Austria
xavier.parent@tuwien.ac.at

Abstract

In response to a concern raised by Horty, this paper develops a two-tiered, preference-based semantic framework for modeling defeasible conditional obligations. The paper extends a Hansson–Lewis style preference semantics for dyadic deontic logic by incorporating a nonmonotonic reasoning mechanism that enables previously derived obligations to be withdrawn when new, potentially conflicting information comes in. The account is bi-preferential: two orderings−ideality and normality−on worlds are employed to address shortcomings in earlier approaches, with a separate ranking method for each. At the nonmonotonic layer, a number of postulates are considered, including antecedent strengthening, inclusion and no-drowning. A connection is established with so-called constrained input/output (I/O) logic−an existing standard for normative reasoning based on a different methodology.

1 Introduction

Obligations are inherently conditional and take the form of conditional statements. Their logical analysis is the focus of the logics of conditional obligation or dyadic deontic logics. These can be classified into two categories: preference-based systems and norm-based systems. Preference-based systems analyse the conditional obligation operator within the framework of possible world semantics. A preference relation ranks possible worlds in terms of betterness or ideality. These systems can be seen as the deontic cousin of the KLM systems for nonmonotonic inference (Kraus, Lehmann, and Magidor 1990).[1] Notable systems in this category include: (Hansson 1969)'s DSDL systems; (van Fraassen 1972)'s system CD; (Lewis 1973)'s system VN; (Goble 2003)'s system DP; (Åqvist 2002)'s systems E, F and G; and (Kratzer 2012)'s semantics. In contrast, norm-based systems analyse the conditional obligation operator with reference to a set of explicit norms representing the normative system. The semantics is operational rather than model-theoretic. The meaning of the conditional obligation operator is given through a set of procedures that yield outputs based on given inputs. Prominent systems in this category are: (Horty 2012)'s deontic default logic; (Makinson and van der Torre 2000)'s input/output (I/O) logic; and (Hansen 2008)'s and (Sergot 2025)'s imperativist semantics for deontic logic.

The methodologies in these two groups differ substantially, leading to distinct semantic frameworks. Each type excels in different areas. Preference-based systems are particularly effective at handling contrary-to-duty (CTD) reasoning, such as (Chisholm 1963)'s paradox–a core problem in deontic logic dealing with norm violation. Norm-based systems are well-suited for defeasible, or nonmonotonic, reasoning: they can accommodate norms with exceptions, allowing conclusions to be revised or withdrawn as new information becomes available. Unifying these approaches so as to combine their respective strengths remains a central challenge, identified as an open problem by (Horty 2014).

This paper focuses on concerns raised by Horty (Horty 1993; Horty 1994; Horty 2007; Horty 2012; Horty 2014) concerning the ability of preference-based systems to accurately model defeasible normative reasoning. The objections are summarized, and a new ordering semantics is proposed to address them. The approach is two-tiered in two respects:

(i) A preference-based dyadic deontic logic is extended with a nonmonotonic reasoning mechanism; nonmonotonicity is obtained by (indirectly) parameterizing an otherwise monotonic consequence relation with a set of norms representing the normative system;

(ii) Two orderings–ideality and normality–on worlds are employed to address issues in previous implementations, notably the fallacy of the prohibited exception (see §2.1); each ordering uses a distinct ranking method.

The deontic fragment of the proposed framework is faithfully embedded into constrained input/output (I/O) logic, a well-established norm-based system due to (Makinson and van der Torre 2001).[2] Earlier approaches either focus on so-called unconstrained I/O logic, which does not support nonmonotonic nor CTD reasoning (Makinson and van der Torre 2000; Bochman 2005; Parent and van der Torre 2014; Stolpe 2020; Parent 2021; Ciabattoni and Rozplokhas 2023), or adopt a proof-theoretic perspective based on adaptive logic or argumentation (Straßer, Beirlaen, and Putte 2016; Arieli, van Berkel, and Straßer 2024). This paper adopts a model-theoretic, KLM-style perspective on constrained

[1](Makinson 1993) describes them as a nonmonotonic formalism that was "ahead of its time" (*avant la lettre*) and identifies them as the fifth—and "most complex" (p. 374)—face of minimality.

[2]Overview in (Parent and van der Torre 2013).

I/O logic, extending previous findings for the unconstrained case. To my knowledge, it is the first endeavor of its kind.[3]

2 Preliminaries

Horty's main point is that normative reasoning is defeasible, and thus nonmonotonic: a conclusion drawn can be revised by new information. Preference-based dyadic deontic logic was developed as a species of modal and conditional logic. The entailment (or consequence) relation is monotonic: once conclusions are drawn, they cannot be revoked.

Nonmonotonicity may be understood as a property of either the conditional, in the object language, or the entailment relation, in the meta-language. Following Horty and others, this paper adopts the latter perspective, justifying it with reference to what may be called the strengthening problem, a central issue in normative reasoning

2.1 The Strengthening Problem

$\bigcirc(B/A)$ may be read as "It ought to be the case that B if A". The principle of *Strengthening of the Antecedent* (SA) allows one to infer $\bigcirc(B/A \wedge C)$ from $\bigcirc(B/A)$. This principle is central to normative reasoning: $\bigcirc(B/A)$ is general to the extent that A can be "freely" strengthened. In practice, A occurs with some background condition C. The question is whether B remains obligatory. By default, it does−unless C cancels the obligation. Without the ability to strengthen antecedents, obligations would fail as cues for action.

The strengthening problem concerns finding a middle ground between unrestricted strengthening and none at all. A child told "Do not eat with your fingers" naturally assumes that the prohibition extends to more specific contexts, like "if eating asparagus," unless instructed otherwise. Accordingly, antecedents can be strengthened as much as possible, treating the obligation as generally as possible—two ways of expressing the same idea—unless doing so conflicts with available evidence. Implementing this idea requires a nonmonotonic entailment relation, which existing norm-based systems can handle easily. The question is whether preference-based dyadic deontic logic can accommodate the same kind of defeasible reasoning.

The proposed approach differs from the "enthymematic" treatment of the strengthening problem (Goble 2004; van Benthem, Grossi, and Liu 2014). The latter keeps a monotonic entailment relation, but restricts (SA) suitably. A seemingly valid strengthening is accounted for by making a "hidden" premise−typically, a permission–explicit. Due to space limitation, a discussion of the enthymematic approach must be deferred to another occasion.

To illustrate the strengthening problem, Horty uses the following example.[4]

Example 1 (The asparagus, (Horty 1993))**.** *Consider the following rules of etiquette:*

(i) *You ought not to eat with your fingers:* $\bigcirc\neg f$*;*

(ii) *You ought to put your napkin on your lap:* $\bigcirc n$*;*

(iii) *If you are served asparagus, you ought to eat with your fingers:* $\bigcirc(f/a)$*.*

Horty notes that, assuming (i) exhausts the premise set, (iv) should be derivable:

(iv) *If you are served asparagus, you ought not to eat with your fingers:* $\bigcirc(\neg f/a)$*.*

He goes on to observe that, if the premise set is (i)-(iii), then (iv) should not be derivable−(iv) contradicts (iii). Accordingly, (SA) cannot be endorsed as it is. However, some strengthening is necessary: just as one wants to infer (iv) from (i) alone, one also wants to infer (v) from (i)–(iii).

(v) *If you are served asparagus, you ought to put your napkin on your lap:* $\bigcirc(n/a)$*.*

The reason is that (iii) defeats or overrides (i), but not (ii). Hence, a conditionalization on a *should be permitted for (ii), but not for (i).*

Techniques were developed in Knowledge Representation and Reasoning (KR) to resolve a similar problem.[5] While one might consider applying them to deontic logic, Horty dismisses this idea:

> "Although I can offer no argument to show that this cannot be done, I have not seen it done successfully, and I am skeptical: being served cold asparagus does not seem like the sort of thing that should force us to consider nonideal worlds." (Horty 2014, p. 449)

Horty's reference to "being served cold asparagus" highlights the following point:

> "In the classical [viz., preference-based] framework [...] we cannot have $\bigcirc\neg f$ and $\bigcirc(f/a)$ without $\bigcirc\neg a$, because the first two statements entail the third." (Horty 2014, p. 446)

The truth-conditions put $\bigcirc(B/A)$ true if all the best A-worlds are B-worlds. Accordingly, $\bigcirc\neg f$ states that f is false in all the best $\top$-worlds, and $\bigcirc(f/a)$ states that all the best a-worlds are f-worlds. It follows that a is false in all the best $\top$-worlds; that is, $\bigcirc\neg a$.[6] Hence, a-worlds are "nonideal" overall. The exception ("being served asparagus") is treated as a prohibition for the only reason that it is an exception.

[3] Omitted proofs can be found in a companion document available on-line (Parent 2026).

[4] As usual $\bigcirc A$ is short for $\bigcirc(A/\top)$, where $\top$ is a tautology.

[5] The asparagus scenario resembles what Pearl calls the problem of "property inheritance [blocking] from classes to subclasses" (Pearl 1990, p. 128). If a subclass of C (e.g., penguins as birds) is exceptional with respect to one property (flying), it may fail to inherit other typical properties of C (such as having wings). This illustrates what (Lehmann 1995) calls "presumptive reasoning": properties of a class are assumed to hold for its members unless there is evidence to the contrary. Intuitively, one wants to infer "penguins have wings" from "birds have wings". The formalization in (Casini, Meyer, and Varzinczak 2019, Ex. 4) further clarifies the analogy between the two cases.

[6] Suppose not. Then a best $\top$-world w is an a-world. Since $\bigcirc\neg f$, w makes f false. However, by $\bigcirc(f/a)$, no best a-world makes f false, so w is not a best a-world. It follows that w is not a best $\top$-world, contrary to our assumption. The argument is informal, and can be verified more rigorously using the evaluation rule that identifies the best worlds with maximal ones (cf. §3).

In (Ryle 1938)'s terminology, such an outcome constitutes a category mistake—akin to asserting that "The number two is blue"—and may be termed the "fallacy of the prohibited exception." Given its centrality, particularly in (Horty 2014), it is presented below as a separate observation. This problem affects the classical semantics, but also some of the more recent works in nonmonotonic logic mentioned above, like (Lehmann 1995)'s lexicographic entailment relation. However, it does not apply to (Delgrande 2020)'s account. But the cost is significant. The obligation $\bigcirc\neg a$ is lost only because the more general obligation $\bigcirc\neg f$ is also eliminated; in context $\top$, the general obligation simply disappears. This is the drowning effect (Broersen and van der Torre 2012; Parent and van der Torre 2017; Parent and van der Torre 2018).[7] This poses a problem for norm-compliance checking. In certain applications, it may be desirable to determine whether $\bigcirc(B/A)$ has been violated or fulfilled. Intuitively, such a check assumes that the obligation holds, but in the model, the obligation does not hold.[8]

Example 2 (Prohibited exception, drowning effect)**.** *From*

(i) *You ought not to eat with your fingers:* $\bigcirc\neg f$*;*

(ii) *If you are served asparagus, you ought to eat with your fingers:* $\bigcirc(f/a)$*;*

one wants to be able to infer

(i) *You ought not to eat with your fingers:* $\bigcirc\neg f$*;*

but not

(iv) *You ought not to be served asparagus:* $\bigcirc\neg a$*.*

A bi-ordering or two-dimensional semantics is employed to address both problems, separating ideality and normality rankings. The language of dyadic deontic logic will thus be supplemented with a normality conditional $\Rightarrow$, expressing what typically holds, relative to a notion of normal or non-exceptional situation.[9] For $A \Rightarrow B$, read "If A, then normally B". The basic idea is to constrain the orderings on worlds in a model further using a designated set of $\bigcirc$- and $\Rightarrow$-formulas. While the use of designated obligations to constrain an ordering on worlds is not new, our approach introduces a second ordering together with a separate set of (normality) conditionals associated with it.

[7]This is distinct from the "drowning" problem described in (Benferhat et al. 1993), which is another term for Pearl's problem of property inheritance.

[8]The drowning effect (for deontics) was originally discussed in (Prakken and Sergot 1996, p. 98ff), although they do not use this term. There it is described as a problem for solutions to the contrary-to-duty (CTD) problem based on nonmonotonic logics (Ryu and Lee 1994; McCarty 1994). In these approaches, an obligation disappears when violated. The so-called cottage regulation example states that there should be no fence, and that if there is a fence it should be white. If we can no longer infer that there should be no fence because a fence exists, how do we formally represent that a violation has occurred?

[9]A classic reference on the modal approach to normality is (Boutilier 1994).

normal:	$\bar{r}\bar{a}$		
exceptional:	r, a	$r, \bar{a}$	$\bar{r}, a$

Figure 1: f-DIS ranking. A bar "‾" over a propositional letter means "$\neg$". Falsifying a more specific conditional does not reduce a world's normality.

2.2 Ranking Worlds

This subsection offers an independent justification for employing distinct ranking methods for deontic and normality conditionals, a key component of the solution to the strengthening problem. In what follows, the generic term 'rule' and the symbol r are used to refer indistinctly to a conditional obligation and a normality conditional.

For normality, (Lehmann 1995)'s lexicographic ordering method will be used, while for ideality (Delgrande 2020)'s method will be used. They will be called LEX and DEL, respectively. Both are variations on a method of ranking that is familiar from conditional logic and deontic logic. Assessing the normality (resp. goodness) of a world involves examining how well it conforms to the normality conditionals (resp. obligations). Fewer violations correspond to a higher degree of normality (resp. ideality). This ranking method is called f-DIS, for flat distance-based. The relation captures an abstract notion of distance to "full" normality (resp. ideality). The adjective "flat" is meant to emphasize the absence of hierarchy among conditionals (resp. obligations).[10]

A problem arises with rules allowing for exceptions. Consider the pair of conditionals "you normally do not eat asparagus" ($\top \Rightarrow \neg a$) and "if you are at the restaurant, you normally eat asparagus" ($r \Rightarrow a$). A normality conditional (resp. an obligation) r is said to be falsified (resp. violated) in a world if its antecedent is true and its consequent false in this world. In this example, f-DIS yields that a $r \wedge a$-world and a $r \wedge \neg a$-world are equally exceptional. This is shown in Fig. 1. Paradoxically, falsifying a more specific conditional does not make a world less normal compared to falsifying a more general one–the two are treated equally.

This outcome is counter-intuitive because, while specificity may represent an exception to a general rule, it still encodes a meaningful expectation.[11] It is noteworthy that, under the standard truth-conditions for $\Rightarrow$ in terms of maximality (viz. "all the most normal A-worlds are B-worlds"), the resulting

[10]Three accounts based on f-DIS are: the semantics of the inference relation of a Poole system (Freund 1998, §3.1), (Kratzer 2012)'s semantics and (van Benthem, Grossi, and Liu 2014)'s theory of P-graphs (for the special case of a "flat" P-graph). (Makinson 1999) discusses how f-DIS relates to (Jørgensen 1937)'s dilemma−the problem of whether norms bear a truth-value and whether a logic of norms can exist.

[11]The point is even clearer in Tweety's well-known example: "birds normally fly" ($b \Rightarrow f$) and "penguins (like Tweety) normally do not fly" ($p \wedge b \Rightarrow \neg f$). Although penguins are exceptions to the general flying rule for birds, the more specific conditional "penguins do not fly" still conveys an expectation. Falsifying it (even artificially, with the help of artificial wings) results in an even less normal world.

model makes $r \Rightarrow a$ false. Indeed, a most normal r-world is a $\neg a$-world.

LEX addresses this problem by ranking worlds lexicographically rather than by a simple count. A world is more normal if it falsifies fewer conditionals, giving precedence to more specific conditionals; falsifications of less specific conditionals are considered only in case of a tie on more specific conditionals. One assumes a partition of the set of (explicitly given) conditionals into levels of specificity, and applies f-DIS level by level, starting with the set of most specific conditionals and proceeding up to the less specific ones. The ranking is shown in Fig. 2 in Ex. 3.

LEX is appropriate for normality conditionals, but not for deontic conditionals. Consider: "you should not eat with your fingers" ($\bigcirc\neg f$) and "if you are served asparagus, you should eat with your fingers" ($\bigcirc(f/a)$). f-DIS and LEX yield that the $\neg a \wedge \neg f$-world is strictly better than the $a \wedge f$-world. This clashes with the specificity principle, which states that the more specific obligation overrides and "suspends" the more general one. $\bigcirc(f/a)$ states that, if a is the case, then f is obligatory. Intuitively, $\bigcirc\neg f$ no longer applies. Why, then, consider this obligation violated, if $a \wedge f$? This is counter-intuitive. DEL corrects this problem by stipulating that $\bigcirc(B/A)$ is not considered violated in world w—even when A holds and B fails—if w triggers and complies with a more specific obligation overriding it. As a result, the $\neg a \wedge \neg f$-world and the $a \wedge f$-world are ranked as equally good.

3 Framework

In addition to a set of propositional letters, the Boolean connectives and the modal operator $\Box$, the language $\mathcal{L}$ has (as primitives) a normality conditional operator $\Rightarrow$ and a conditional obligation operator $\bigcirc(-/-)$. $\Rightarrow$ and $\bigcirc(-/-)$ are read as before. For $\Box A$, read "It is necessary that A". For simplicity's sake, it is assumed that $\Rightarrow$ and $\Box$ do not occur within the scope of $\bigcirc$ (and vice-versa) and that $\Box$ does not occur within the scope of $\Rightarrow$ (and vice-versa). Iterations of $\bigcirc$ or $\Rightarrow$ are ruled out too.[12] As mentioned, r (possibly indexed by a natural number) is a schematic symbol that may stand for either a conditional obligation or a normality conditional. Given some r, $h(r)$ is the head (or consequent) of r, and $b(r)$ is its body (or antecedent). As usual, $\Diamond A$ is short for $\neg\Box\neg A$. $\models_{\mathrm{S5}}$ denotes the entailment relation in modal logic S5 (Chellas 1980), and $\models_{\mathrm{PL}}$ the entailment relation in (classical) propositional logic.

3.1 Preliminary Notions

This section introduces notions needed for subsequent developments. In what follows, $\mathbf{R} = (\mathbf{R}^{\Rightarrow}, \mathbf{R}^{\circ})$, where $\mathbf{R}^{\Rightarrow}$ and $\mathbf{R}^{\circ}$ are finite sets of $\Rightarrow$-formulas and $\bigcirc$-formulas, respectively. The formulas in $\mathbf{R}^{\Rightarrow}$ can be thought of as expressing "soft" constraints, allowing for exceptions. $\mathbf{R}^{\circ}$ is the normative system under consideration. Alethic formulas are used to encode hard (i.e., non-defeasible) information about the domain. For instance, $\Box(A \to \neg B)$ states that A and B cannot both hold.

[12] An example of a logic allowing nested occurrences of $\Rightarrow$ is (Lamarre 1991)'s system C4. Nested conditionals are not directly relevant to the problem at hand (see, e.g., (Makinson 1993) for a general discussion).

Definition 1 (Overridingness or defeat)**.** *Given $r_i, r_j \in \mathbf{R}^{\circ}$, and a set Γ of alethic formulas, r_j is said to override r_i (notation: $r_j \rhd r_i$) whenever*

(i) $\{h(r_i), h(r_j)\} \cup \Gamma \models_{\mathrm{S5}} \bot$

(ii) $b(r_j) \models_{\mathrm{PL}} b(r_i)$ *and* $b(r_i) \not\models_{\mathrm{PL}} b(r_j)$

(iii) $h(r_i), b(r_j) \not\models_{\mathrm{PL}} \bot$.

(i) and (ii) may be found in (Horty 1993; Delgrande 2020). (i) says that the heads of the obligations are inconsistent in S5, given the hard information at our disposal. (ii) says that the overriding obligation is more specific. (iii) is proposed by (Tan and van der Torre 1995). It may be motivated as follows. A contrary-to-duty (CTD) obligation says what should be done if another (primary) obligation is violated. (iii) tells us that r_j is not a CTD of r_i. Consider (Forrester 1984)'s gentle murder paradox: *Do not kill, but if you kill, do it gently!*. Without (iii), the CTD obligation *If you kill, do it gently!* ($\bigcirc(k \wedge g/k)$) would override the primary obligation *Do not kill!* ($\bigcirc\neg k$). This is usually considered counter-intuitive.

A model (referred to as an $\mathbf{R}$-ordered model) has the form

$$M = (W, \geq_N, \geq_I, v)$$

where W is a (non-empty) set of worlds, $\geq_N$ and $\geq_I$ denote a normality and ideality (betterness) ordering on the worlds, and v a valuation assigning to each world the set of propositional letters true in this world. "$w \geq_N v$" is read as "w is at least as normal as v", and "$w \geq_I v$" as "w is at least as good (or ideal) as v". "$>_N$" (the strict counterpart of "$\geq_N$") is defined by $w >_N v$ if and only if (iff) $w \geq_N v$ and $v \not\geq_N w$. "$>_I$" (the strict counterpart of "$\geq_I$") is defined analogously. Two worlds w and v are equally normal iff $w \geq_N v$ and $v \geq_N w$. Given $X \subseteq W$, $\max_{\geq_N}(X)$ is $\{w \in X : \forall v \in X\ (v \geq_N w \to w \geq_N v)\}$. Equal goodness is defined analogously. $|\cdot|$ is the cardinality of $\cdot$.

Given a model M, and a world w, the satisfaction relation $\models \subseteq W \times \mathcal{L}$ is defined inductively by "$w \models p$ iff $w \in v(p)$", with the usual clauses for the Boolean connectives. Those for the other connectives are introduced in the next sections.

Occasionally the focus will be on models that are replete in the sense of (Goble 2019). Call a Boolean formula A PL-consistent, if it is consistent in (classical) propositional logic, viz. $A \not\models_{\mathrm{PL}} \bot$.

Definition 2 (Repleteness)**.** *A model M is replete if, whenever a Boolean formula A is PL-consistent, there is a world w in M such that $M, w \models A$.*

Definition 3 (Falsification set)**.** *Set*

$$F(w) = \{r_i \in \mathbf{R}^{\Rightarrow} : w \models b(r_i) \wedge \neg h(r_i)\}$$

$F(w)$ gathers all the conditionals in $\mathbf{R}^{\Rightarrow}$ falsified by w.

Definition 4 (Violation set)**.** *Set*

$$V(w) = \{r_i \in \mathbf{R}^{\circ} : w \models b(r_i) \wedge \neg h(r_i) \text{ and } w \not\models b(r_j)\ \forall r_j \in \mathbf{R}^{\circ} \text{ s.t. } r_j \rhd r_i\}$$

$V(w)$ gathers all the obligations violated by w. The clause "$w \not\models b(r_j)\ \forall r_j \in \mathbf{R}^{\circ}$ s.t. $r_j \rhd r_i$" ensures that $\bigcirc(B/A)$ is not considered violated by w—even when A holds and B fails—if w triggers and complies with a more specific obligation overriding it (cf. §2.2).

3.2 Normality

This section describes (Lehmann 1995)'s lexicographic ordering method for normality. Given $X \subseteq \mathbf{R}^{\Rightarrow}$, denote by $\mathrm{m}(X)$ the materialization of X, i.e., $\{B \rightarrow C : B \Rightarrow C \in X\}$, where $\rightarrow$ is material implication. A Boolean formula A is said to be exceptional for X if $\mathrm{m}(X) \models_{\mathrm{PL}} \neg A$. A conditional $A \Rightarrow B$ is said to be exceptional for X if its antecedent A is. Let $\varepsilon(X)$ be the set of $A \Rightarrow B$s in $\mathbf{R}^{\Rightarrow}$ that are exceptional for $X \subseteq \mathbf{R}^{\Rightarrow}$. Define a sequence $\mathcal{E}_0$, ..., $\mathcal{E}_\infty$ of subsets of $\mathbf{R}^{\Rightarrow}$ as follows: (i) $\mathcal{E}_0 = \mathbf{R}^{\Rightarrow}$; (ii) $\mathcal{E}_i = \varepsilon(\mathcal{E}_{i-1})$, for $0 < i < m$ and $\mathcal{E}_\infty = \mathcal{E}_m$, where m is the smallest k for which $\mathcal{E}_k = \mathcal{E}_l$ for all $l > m$. We call m the order of $\mathbf{R}^{\Rightarrow}$. That m always exists follows from the fact that $\mathbf{R}^{\Rightarrow}$ is finite, and the following fact:

Fact 1. $(\mathcal{E}_i)_{i\geqslant 0}$ *is (strictly) decreasing:*

$$\mathcal{E}_{i+1} \subset \mathcal{E}_i \textit{ for all } i \geqslant 0$$

$(\mathcal{E}_i)_{i\geqslant 0}$ is called the LM-sequence (after (Lehmann and Magidor 1992)).

For technical reasons, an assumption of coherence of $\mathbf{R}^{\Rightarrow}$ is made. This assumption is intuitively plausible.

Definition 5 (Coherence)**.** *$\mathbf{R}^{\Rightarrow}$ is said to be coherent if $\neg\exists X \subseteq \mathbf{R}^{\Rightarrow}$ such that $\mathrm{m}(X) \models_{\mathrm{PL}} \bigwedge_{r\in X} \neg b(r)$, and incoherent otherwise.*

Assumption 1. *$\mathbf{R}^{\Rightarrow}$ is coherent.*

For instance, $\{A \Rightarrow B, A \Rightarrow \neg B\}$ and $\{A \Rightarrow \neg A\}$ are incoherent, but $\{A \Rightarrow B, A \wedge C \Rightarrow \neg B\}$ is coherent.

The following follows:

Fact 2. $\mathcal{E}_\infty = \emptyset$.

Proof. Suppose, to reach a contradiction, that $\mathcal{E}_\infty = \mathcal{E}_m \neq \emptyset$, where m is the smallest k for which $\mathcal{E}_k = \mathcal{E}_l$ for all $l > m$. Let $\mathcal{E}_m = \{r_1, ..., r_n\} \subseteq \mathbf{R}^{\Rightarrow}$. Since $\mathcal{E}_m = \mathcal{E}_{m+1}$,

$$\mathcal{E}_{m+1} = \{r_1, ..., r_n\}.$$

Since $\mathcal{E}_{m+1} = \varepsilon(\mathcal{E}_m)$,

$$\mathrm{m}(\mathcal{E}_m) \models_{\mathrm{PL}} \bigwedge_{i=1}^{n} \neg b(r_i)$$

As $\mathcal{E}_m \subseteq \mathbf{R}^{\Rightarrow}$, $\mathbf{R}^{\Rightarrow}$ is incoherent. $\square$

From this, we define the ranked partition of $\mathbf{R}^{\Rightarrow}$ as $\langle \Delta_0, ..., \Delta_m \rangle$ where $\Delta_i = \mathcal{E}_i - \mathcal{E}_{i+1}$, for $i = 0, ..., m-1$, and $\Delta_m = \mathcal{E}_\infty$. Intuitively, each Δ_i gathers the elements of $\mathbf{R}^{\Rightarrow}$ with the same level of specificity. Δ_0 is the set of less specific conditionals, Δ_1 is the set of second-less specific ones, and so forth up to Δ_{m-1}, the set of most specific ones. Given some $\mathbf{R}^{\Rightarrow}$ of order m, to every subset $X \subseteq \mathbf{R}^{\Rightarrow}$ may be associated a tuple of natural numbers: $\langle n_1, ..., n_m \rangle$, where $n_i = |\Delta_{m-i} \cap X|$, for all $i = 1, ..., m$. These Xs may be ordered using the natural lexicographic ordering on their associated tuples. Let the relation be denoted by $\gtrsim$. Its definition may be stated formally thus, where $\langle n_1, ..., n_m \rangle$ and $\langle n'_1, ..., n'_m \rangle$ are the tuples associated with X and Y, respectively:[13]

most normal:	$\bar{r}\bar{a}^{<0,0>}$
intermediate:	$ra^{<0,1>}$ $\bar{r}a^{<0,1>}$
least normal:	$r\bar{a}^{<1,0>}$

Figure 2: LEX ranking. Falsifying a more specific conditional makes a world less normal (cf. §2.2)

$$\begin{aligned} X \gtrsim Y \text{ iff } & \langle n_1, ..., n_m \rangle \geqslant^{lex} \langle n'_1, ..., n'_m \rangle \\ & \text{iff } n_i = n'_i \text{ for all } i \text{ such that } 1 \leqslant i \leqslant m, \text{ or} \\ & \quad n_i < n'_i \text{ for the first } i \text{ such that } n_i \neq n'_i \end{aligned}$$

The normality relation $\succeq_N$ can now be introduced. Intuitively, a world is more normal if it falsifies fewer conditionals, giving precedence to more specific conditionals; falsifications of less specific conditionals are considered only in case of a tie on more specific conditionals.

Definition 6 (Normality relation)**.** *Given a set $\mathbf{R}^{\Rightarrow}$ of normality conditionals, and $w_1, w_2 \in W$, set*

$$w_1 \succeq_N w_2 \textit{ iff } F(w_1) \gtrsim F(w_2)$$

Example 3 (Exception, normality)**.** *Let $\mathbf{R}^{\Rightarrow} = \{\top \Rightarrow \neg a, r \Rightarrow a\}$. Fig. 2 shows the worlds' ranking, with tuples indicated by superscripts.*

Remark 1. *(i) $\succeq_N$ is reflexive, transitive and total (for all $w_1, w_2 \in W$, $w_1 \succeq_N w_2$ or $w_2 \succeq_N w_1$); (ii) $\succ_N$ is irreflexive, transitive and modular (for all $w_1, w_2, w_3 \in W$, if $w_1 \succ_N w_2$ then $w_1 \succ_N w_3$ or $w_3 \succ_N w_2$).*

$\|A\|^M$ is the truth-set of A in M, viz., the set of worlds where A holds. M is omitted when no confusion can arise.

Remark 2. *If $\mathbf{R}^{\Rightarrow} = \emptyset$, then $\succeq_N = W \times W$ and* $\max_{\succeq_N}(\|A\|) = \|A\|$.

Infinite chains of increasingly more normal worlds are ruled out; $\succeq_N$ verifies the condition called "smoothness" by (Kraus, Lehmann, and Magidor 1990) and "stoppering" by (Makinson 1988):

Proposition 1 (smoothness)**.** *If $w_1 \models A$, then $w_1 \in \max_{\succeq_N}(\|A\|)$ or $\exists w_2 \succ_N w_1$ s.t. $w_2 \in \max_{\succeq_N}(\|A\|)$.*

Intuitively, $\Box A$ holds in world w when A holds in all worlds, and $A \Rightarrow B$ holds in w when the most normal A-worlds are all B-worlds. Formally:

Definition 7 (Truth-conditions, part I)**.**

$$M, w \models \Box A \textit{ iff } W = \|A\|^M$$

$$M, w \models A \Rightarrow B \textit{ iff } \max_{\succeq_N}(\|A\|^M) \subseteq \|B\|^M$$

Remark 3 (Systems S5 and R)**.** *(i) $\Box$ validates the axioms and rules of S5; (ii) $\Rightarrow$ validates the rules of (Lehmann and Magidor 1992)'s system R (recast as a Hilbert-style axiom system).*

[13] $n_0 = |\Delta_m \cap X|$ does not need to be considered as a first coordinate, since under Assumption 1 it is always equal to 0.

3.3 Ideality

Definition 8 (Ideality relation)**.** *For a set $\mathbf{R}^{\circ}$ of obligations, and $w_1, w_2 \in W$, set:*

$$w_1 \geq_I w_2 \text{ iff } V(w_1) \subseteq V(w_2)$$

Intuitively, w_1 is at least as good as w_2 if the set of obligations violated by w_1 is included in the set of obligations violated by w_2.

Remark 4. *$\geq_I$ is reflexive, transitive and smooth.*

The truth-conditions of $\bigcirc$ are a variation on ones commonly used in a mono-preferential setting. To indicate what the variation is, a specific lifting relation must be introduced.[14] Below the superscript s is for "set".

Definition 9. *For $U, V \subseteq W$, $U \geq_I^s V$ (reading: U is at least as good as V) iff $\forall v \in V\ \exists u \in U$ s.t. $u \geq_I v$.*

Proposition 2. *$\geq_I^s$ is reflexive and transitive.*

In deontic logic, it is usual to interpret $\bigcirc(B/A)$ as expressing a (strict) preference of $A \wedge B$ over $A \wedge \neg B$. $\bigcirc(B/A)$ is equivalent with $A \wedge B > A \wedge \neg B$, where the connective $>$ (on formulas) is defined by putting $A > B$ as true at a world iff $\|B\| \not\geq_I^s \|A\|$. Intuitively, to determine the truth-value of $\bigcirc(B/A)$ amounts to verifying whether the set of $A \wedge \neg B$-worlds is not at least as good as the set of $A \wedge B$-worlds, using an appropriate lifting relation. Instead of working with the set of worlds verifying the relevant formulas, this paper suggests working with the subset of those that are the most normal. One gets: $\bigcirc(B/A)$ is true, whenever the set of most normal $A \wedge \neg B$-worlds is not at least as good as the set of most normal $A \wedge B$-worlds. This kind of variant rule, where normality and ideality are combined, is familiar from the deontic logic literature−see e.g. (Jackson 1985; van der Torre 1997).

Definition 10 (Truth-conditions, part II)**.**

$$w \models \bigcirc(B/A) \text{ iff } \max_{\geq_N}(\|A \wedge \neg B\|) \not\geq_I^s \max_{\geq_N}(\|A \wedge B\|)$$

Taken together, Def. 9 and 10 yield the following derived truth-conditions. They state that $\bigcirc(B/A)$ holds, if there is a most normal $A \wedge B$-world that is not worse than (or weakly dominated by) any most normal $A \wedge \neg B$-world.

Proposition 3 (Derived truth-conditions)**.**

$$w \models \bigcirc(B/A) \text{ iff } \exists v \in \max_{\geq_N}(\|A \wedge B\|)\ \ \forall u \in \max_{\geq_N}(\|A \wedge \neg B\|)\ \ u \not\geq_I v$$

A few words about zooming in on the most normal worlds are in order. The assumption is that when I say "it ought to be that B, given A," I typically do so under the presumption that things are as normal as possible. In other words, exceptional or abnormal worlds are treated as irrelevant when assessing whether the obligation holds. This assumption is in itself plausible. As shown in §3.5, it prevents the drowning effect.

[14](Halpern 1997, p. 4) credits this lifting relation to (Lewis 1973).

It is now possible to explain why Def. 9 was chosen. When $\mathbf{R}^{\Rightarrow} = \emptyset$, the truth-conditions for $\bigcirc$ coincide with those employed in deontic logic to accommodate unresolved conflicts, namely situations in which both $\bigcirc A$ and $\bigcirc\neg A$ hold without either taking precedence (van Fraassen 1972; Lewis 1974; Prakken and Sergot 1997). This pattern of evaluation may be referred to as the $\exists\forall$ truth-conditions. The use of this approach to accommode conflicts is advocated by (Goble 2003; Goble 2013). The rule reads: $\bigcirc(B/A)$ is true in w iff there is some $A \wedge B$-world such that, as we go up in the ordering, the material implication $A \rightarrow B$ always holds.[15] Formally,

Proposition 4 ($\exists\forall$ truth-conditions)**.** *Let $\mathbf{R}^{\Rightarrow} = \emptyset$. Then:*

$$w \models \bigcirc(B/A) \text{ iff } \exists v \big(v \models A \wedge B \text{ and } \forall u\, (u \geq_I v \twoheadrightarrow u \models A \rightarrow B)\big)$$

Proof. This follows at once from Prop. 3 and Rem. 2. □

$\bigcirc$ as defined in Prop. 4 satisfies the axioms and rules of (Goble 2003)'s system DP.[16]

Looking back, this consideration is what primarily motivated the use of Def. 9. It is reasonable to expect that an established theory of conditional obligation should emerge as a limiting case when normality considerations are set aside.

3.4 Nonmonotonic Layer

This section adds a nonmonotonic layer to the framework.

It is helpful to adopt Horty's terminology. His entailment relation $|\!\sim$ operates on a theory Δ, and outputs formulas (e.g., obligations) "entailed" by the theory (see, e.g., (Horty 2012, p. 80)). In what follows, Γ is a (finite) set of ordinary (Boolean or alethic) formulas. A theory Δ is a pair $(\Gamma, \mathbf{R})$, where $\mathbf{R} = (\mathbf{R}^{\Rightarrow}, \mathbf{R}^{\circ})$. $|\!\sim$ is defined as in (Prakken and Sergot 1997). The definition refers to the class $\mathcal{C}(\mathbf{R})$ of all $\mathbf{R}$-ordered models, whose normality and ideality orderings are determined by $\mathbf{R}$ (*as per* §3.2 and §3.3). Accordingly, one can say that $|\!\sim$ is (indirectly) parametrized by $\mathbf{R}$.

Definition 11. *$\Delta\ |\!\sim A$ ("Δ entails A") iff $\Gamma \models_{\mathcal{C}(\mathbf{R})} A$.*

Intuitively, Def. 11 says that theory Δ entails A if in any $\mathbf{R}$-ordered model M, and any world w in M, if all the formulas in Γ hold in w, then so does A. Def. 11 does not require that every normality conditional $A \Rightarrow B$ in $\mathbf{R}^{\Rightarrow}$ be satisfied in world w. Prop. 5 will demonstrate that this requirement is automatically met. Similarly, the definition imposes no such requirement on the obligations $\bigcirc(B/A)$ in $\mathbf{R}^{\circ}$; Prop. 7 and 8 specify when this too holds automatically.

[15]Compared to the standard evaluation pattern in terms of "best" (Hansson 1969), the main advantage of the $\exists\forall$ rule is that it can accommodate conflicts without generating a deontic collapse. In particular, the set $\{\bigcirc A, \bigcirc\neg A\}$ is satisfiable in a model. But, unlike the best-based evaluation rule, the $\exists\forall$ rule does not warrant the inference from $\bigcirc A$ and $\bigcirc\neg A$ to $\bigcirc B$ for an arbitrary B, implying the collapse of all normative distinctions in the face of a conflict.

[16]DP differs from (Lewis 1973)'s system VN in not containing the axiom $\bigcirc(B/A) \rightarrow \neg\bigcirc(\neg B/A)$ nor the principle AND, which is present in (Kraus, Lehmann, and Magidor 1990)'s system P.

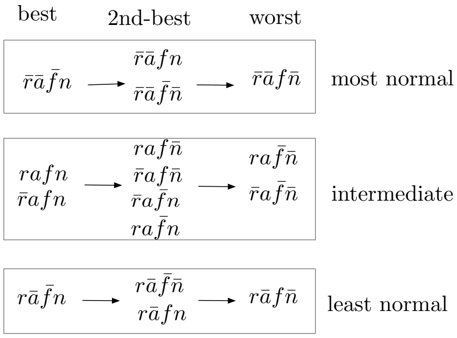

Figure 3: The asparagus

To capture the property of monotonicity in our setting, the usual notion of containment between theories must be redefined.

Definition 12. $\Delta_1 = (\Gamma_1, \mathbf{R}_1)$ *is contained in* $\Delta_2 = (\Gamma_2, \mathbf{R}_2)$ *(notation:* $\Delta_1 \sqsubseteq \Delta_2$*) whenever* $\Gamma_1 \subseteq \Gamma_2$, $\mathbf{R}_1^{\Rightarrow} \subseteq \mathbf{R}_2^{\Rightarrow}$, *and* $\mathbf{R}_1^{\circ} \subseteq \mathbf{R}_2^{\circ}$.

Def. 12 means that Δ_2 contains all the facts, hard information, explicit obligations and explicit normality conditionals already present in Δ_1. Monotonicity is the property that if $\Delta_1 \mid\!\sim A$ and $\Delta_1 \sqsubseteq \Delta_2$, it follows that $\Delta_2 \mid\!\sim A$.

3.5 Example

A specificity structure is represented as $\{\bigcirc A, \bigcirc(\neg A/B), \top \Rightarrow \neg B\}$. On the technical side, $\top \Rightarrow \neg B$ is used to generate a normality ordering in the model. Conceptually, $\top \Rightarrow \neg B$ fixes $\neg B$ as the default context. Since $\neg B$ is declared holding by default, the default obligation in that context is $\bigcirc A$, defeasibly overridden by $\bigcirc(\neg A/B)$.

Example 4. *[The asparagus, extended version] Consider a theory* $\Delta = (\Gamma, \mathbf{R})$ *where* $\Gamma = \varnothing$ *and* $\mathbf{R}^{\Rightarrow} = \{\top \Rightarrow \neg a, r \Rightarrow a\}$ *and* $\mathbf{R}^{\circ} = \{\bigcirc\neg f, \bigcirc(f/a), \bigcirc n\}$. *Given* $\mathbf{R}^{\Rightarrow}$, *the set of all possible worlds is partitioned into three equivalence classes: the class of those with* $\neg a$ *(=the most normal worlds), the class of those with* $r \wedge a$ *or* $\neg r \wedge a$ *(=the second most normal worlds) and the class of those with* $r \wedge \neg a$ *(=the least normal worlds). Fig. 3 shows the resulting* $\mathbf{R}$*-ordered model* M*. The "*y*-axis" represents normality. The higher up a world is, the more normal it is. The "*x*-axis" represents ideality. The more to the right a world is, the farther away to full ideality it is. The account gives the expected solutions. First, each of* $\bigcirc(f/a)$, $\bigcirc n$, $\top \Rightarrow \neg a$ *and* $r \Rightarrow a$ *holds in a given world* w *in* M*, and so each is defeasibly entailed by* Δ*. Second, we have that*

$$\Delta \mid\!\sim \bigcirc(n/a) \quad (1) \qquad \Delta \not\mid\!\sim \bigcirc\neg a \text{ (witness: } afn\text{)} \quad (3)$$

$$\Delta \not\mid\!\sim \bigcirc(\neg f/a) \quad (2) \qquad \Delta \mid\!\sim \bigcirc\neg f \text{ (witness: } \bar{r}\bar{a}\bar{f}n\text{)} \quad (4)$$

(1) and (2) show that the "right" amount of strengthening is obtained. (3) and (4) show that the fallacy of the prohibited exception and the drowning effect are avoided.

We can now see why zooming in on the most normal worlds in Def. 10 prevents the drowning effect. As mentioned, the idea is that, when I say e.g. "you ought not to eat with your fingers," I do so assuming that circumstances are as normal as possible. Abnormal or exceptional worlds—such as those where you are served asparagus—are disregarded when evaluating whether the obligation holds. Fig. 3 illustrates that, if such worlds were included, $\bigcirc\neg f$ would be false. For $\|f\| \succeq_I^s \|\neg f\|$ while $\max_{\succeq_N}(\|f\|) \not\succeq_I^s \max_{\succeq_N}(\|\neg f\|)$. Thus, obligations do not drown because normality acts as an extra filter.

In this setting, the fallacy of the prohibited exception does not arise, thanks to DEL, and the fact that the ideality and normality relations are independent. As mentioned in §2.2, the key point is that, other things being equal with respect to other obligations, a world in which the general obligation is violated and the more specific one fulfilled ($a \wedge f$) is as good as a world in which the general obligation is fulfilled ($\neg a \wedge \neg f$). This remains true when referring to the most normal $a \wedge f$- and $\neg a \wedge \neg f$-worlds, even though they lie at different levels of normality.

4 Properties

This section generalizes the points made in the previous example, by identifying postulates characteristic of the system at the nonmonotonic layer. They serve as laws for the explicitly given obligations and normality conditionals.

The first postulate is the principle of inclusion,[17] stating that explicit conditionals are defeasibly entailed. For normality conditionals, inclusion holds without restriction (subject to a consistency proviso).

Proposition 5 (Inclusion, normality)**.** *Consider a theory* Δ *such that* $A \Rightarrow B \in \mathbf{R}^{\Rightarrow}$ *and* $A \wedge \neg B$ *is PL-consistent. Then (under the assumption of repleteness)*

$$\Delta \mid\!\sim A \Rightarrow B$$

As a corollary of Prop. 6 below, the inclusion property extends to obligation; it is a special case of strengthening of the antecedent (SA), which holds as a defeasible principle.

Proposition 6 (Defeasible SA, obligation)**.** *Suppose* Δ *is such that* $\bigcirc(B/A) \in \mathbf{R}^{\circ}$, $\neg\exists r \in \mathbf{R}^{\circ}$ *s.t.* $r \rhd \bigcirc(B/A)$ *and* $\Diamond(A \wedge B \wedge C) \in \Gamma$*. Then,*

$$\Delta \mid\!\sim \bigcirc(B/A \wedge C)$$

Proof. We show that, for all $\mathbf{R}$-ordered model M, and all w in M, if $w \models \bigwedge \Gamma$, then $w \models \bigcirc(B/A \wedge C)$. Since $w \models \Diamond(A \wedge B \wedge C)$, $\exists u$ s.t. $u \models A \wedge B \wedge C$. By smoothness, $\exists v \in \max_{\succeq_N}(\|A \wedge B \wedge C\|)$. Let $z \in \max_{\succeq_N}(\|A \wedge C \wedge \neg B\|)$. We have $v \models A \wedge B$ and $z \models A \wedge \neg B$. Since $\neg\exists r \in \mathbf{R}^{\circ}$ s.t. $r \rhd \bigcirc(B/A)$, $\bigcirc(B/A) \in V(z)$. But $\bigcirc(B/A) \notin V(v)$. Hence $z \not\succeq_I v$. □

The case where $C := \top$ yields inclusion. This property holds under the assumption that the normative system contains no more specific obligation overriding the inferred one. This qualification is justified, since a more specific obligation may override the general obligation in that context. This condition is sufficient, though not necessary.[18]

[17] The name is from (Casini, Meyer, and Varzinczak 2019).

[18] Since Γ is not assumed to be closed under logical consequence, Prop. 7 is not, strictly speaking, an instance of Prop. 6; nevertheless, a similar argument establishes it.

Proposition 7 (Inclusion, obligation)**.** *Suppose Δ is such that $\bigcirc(B/A) \in \mathbf{R}^{\circ}$, $\neg\exists r \in \mathbf{R}^{\circ}$ s.t. $r \rhd \bigcirc(B/A)$ and $\Diamond(A \wedge B) \in \Gamma$. Then*

$$\Delta \mid\!\sim \bigcirc(B/A)$$

Nonmonotonicity follows from Prop. 6 and Ex. 4.

Example 5 (Nonmonotonicity)**.** *Consider $\Delta_1 = (\Gamma_1, \mathbf{R}_1)$, with $\Gamma_1 = \{\Diamond(a \wedge \neg f)\}$ and $\mathbf{R}_1^{\circ} = \{\bigcirc\neg f\}$. By Prop. 6, $\Delta_1 \mid\!\sim \bigcirc(\neg f/a)$. Let Δ_2 be as in Ex. 4, except that we now set $\Gamma_2 = \{\Diamond(a \wedge \neg f)\}$. $\Delta_1 \sqsubseteq \Delta_2$, but $\Delta_2 \not\mid\!\sim \bigcirc(\neg f/a)$ (witness: either of the two leftmost worlds in the intermediate layer in Fig. 3).*

Complementing Prop. 7, Prop. 8 gives sufficient conditions under which an overridden obligation does not drown. The condition $\Delta \mid\!\sim \neg(\top \Rightarrow (A \rightarrow B))$ indicates that the obligation in question is not expected to be necessarily fulfilled in the most normal worlds—arguably a separate issue.

Proposition 8 (No-drowning)**.** *Let Δ be such that $\bigcirc(B/A) \in \mathbf{R}^{\circ}$, $\exists r \in \mathbf{R}^{\circ}$ s.t. $r \rhd \bigcirc(B/A)$, $\top \Rightarrow \neg b(r) \in \mathbf{R}^{\Rightarrow}$ for all $r \in \mathbf{R}^{\circ}$ s.t. $r \rhd \bigcirc(B/A)$, $\Diamond(A \wedge B) \in \Gamma$ and $\Delta \mid\!\sim \neg(\top \Rightarrow (A \rightarrow B))$. Then*

$$\Delta \mid\!\sim \bigcirc(B/A)$$

Proof. Suppose $w \models \bigwedge\Gamma$. As before, since $w \models \Diamond(A \wedge B)$, $\exists s \in \max_{\geq_N}(\|A \wedge B\|)$. Let $u \in \max_{\geq_N}(\|A \wedge \neg B\|)$. Since $\Delta \mid\!\sim \neg(\top \Rightarrow (A \rightarrow B))$, $w \models \neg(\top \Rightarrow (A \rightarrow B))$. Hence $\exists v$ in M s.t. $v \in \max_{\geq_N}(W)$ and $v \models A \wedge \neg B$. Let $t \in W$ be s.t. $t \geq_N u$. By totality, either $t \geq_N v$ or $v \geq_N t$. In both cases, $v \geq_N t$, since $v \in \max_{\geq_N}(W)$. By totality again, either $u \geq_N v$ or $v \geq_N u$. In both cases, $u \geq_N v$, since $v \models A \wedge \neg B$ and $u \in \max_{\geq_N}(\|A \wedge \neg B\|)$. By transitivity, $u \geq_N t$, and so $u \in \max_{\geq_N}(W)$. By Prop. 5, $w \models \top \Rightarrow \neg b(r)$ for all r s.t. $r \rhd \bigcirc(B/A)$. So $u \not\models b(r)$ for all such r. Hence $\bigcirc(B/A) \in V(u)$. Since $s \models A \wedge B$, $\bigcirc(B/A) \notin V(s)$. It follows that $u \not\geq_I s$. Hence $w \models \bigcirc(B/A)$. □

5 Embedding into Constrained I/O logic

5.1 Basic Definitions

This section begins with a brief description of unconstrained I/O logic (Makinson and van der Torre 2000), defined on top of classical propositional logic (PL). For readability, lowercase letters a, x, ... are used to represent Boolean formulas. A normative code is a set N of conditional norms, defined as pairs of Boolean formulas (a, x), where a is the body of the norm and x the head of the norm. The pair (a, x) can be read as "if a, then x is obligatory." The main semantic construct is of the form $x \in out(N, a)$. This is interpreted as follows: given an input a (state of affairs), x (obligation) is included in the output under norms N. out is defined in terms of detachment (or modus ponens). This paper uses the reusable throughput operation out_4^+ as the underlying I/O operation. It is based on the I/O operation out_4, which is the strongest I/O operation among those studied by (Makinson and van der Torre 2000). out_4^+ is obtained from out_4 by allowing inputs to reappear as outputs and adding (a, a) as an axiom. This leads to a collapse into classical logic. Specifically, $x \in out_4^+(N, a)$ iff $x \in out_4(N \cup \{(b, b) \mid b \text{ is a formula}\}, a)$, which in turn is equivalent to $\{a\} \cup \mathrm{m}(N) \models_{\mathrm{PL}} x$, where $\mathrm{m}(N)$ is the materialization of N, with the definition straightforwardly adapted to the I/O setting.[19]

To deal with contrary-to-duties (CTDs) and defeasible reasoning, constrained I/O logic (cIOL) adds a nonmonotonic mechanism that filters excess output via a set C of formulas, called constraints.[20] To avoid excess output–namely, inconsistency–when too many norms apply, the set of norms is cut back to just below the threshold of yielding excess; this is the "threshold" idea,[21] captured by the notions of maxfamily and outfamily (here, out is out_4^+):

- maxfamily: *maxf*(N, a, C) is the set of $\subseteq$-maximal subsets H of N such that $out(N, a)$ is PL-consistent with C;
- outfamily: *outf*$(N, a, C) = \{out(H, a) \mid H \in$ *maxf*$(N, a, C)\}$.

The so-called full meet constrained output under N given input a, $\cap($*outf*$(N, a, C))$, is the final output.

Here, $C = \{a\}$. This choice is motivated by (Hansson 1969)'s view of the input as "settled": what has occurred cannot be undone, and the output must therefore be consistent with the input (*ought* implies *can*).

5.2 Embedding

Let $\mathbf{R}^{\Rightarrow} = \varnothing$ and $\mathbf{R}^{\circ} = \{r_1, ..., r_n\}$. The basic idea is to rewrite $\bigcirc(x/a)$ as an I/O pair (a, x) with a more detailed body; this one also contains the negation of the bodies of all its defeaters, viz. all the more specific obligations overriding it. Let $\mathbf{R}^{\circ}_{\rhd}$ represent the result of rewriting the obligations in $\mathbf{R}^{\circ}$ in this manner, and $\bigcirc_{\mathrm{H}}(x/a)$ stand for the Hanssonian obligation operator, defined by putting $\bigcirc_{\mathrm{H}}(x/a)$ as true whenever $\max_{\geq_I}(\|a\|) \subseteq \|x\|$.[22] Let $\mathbf{R} = (\varnothing, \mathbf{R}^{\circ})$. The main result of this section is that $(\varnothing, \mathbf{R}) \mid\!\sim \bigcirc_{\mathrm{H}}(x/a)$ iff x is in the full meet constrained output under $\mathbf{R}^{\circ}_{\rhd}$ given input a, with out_4^+ as the underlying I/O operation and $C = \{a\}$. Thus, the embedding is faithful, viz. it preserves and reflects validity.[23] Validity-reflection means that the translation does not output more obligations. From this, a partial result (of validity-reflection) for $\bigcirc(-/-)$ under the $\exists\forall$ rule (cf. Prop. 4) follows. Even thought the result is partial, it gives some insights into the relationship between the two accounts.

The result and the main steps of the proof are given below; full details are provided in Appendix A. It is assumed that, on the preference side, the given model is replete.

To each $r_i := \bigcirc(x/a)$, one associates the set of all its

[19] Given $X \subseteq N$, $\mathrm{m}(X)$ is the set of all $b \rightarrow y$ with $(b, y) \in X$. When X is a singleton set, curly brackets are omitted.

[20] See (Makinson and van der Torre 2001; Parent 2011).

[21] For a comparison with Horty's approach, see (Parent 2011).

[22] Cf. (Hansson 1969; Makinson 1993).

[23] Validity-preservation (sometimes called soundness) is the left-to-right direction of the bi-conditional in Th. 3; validity-reflection is the right-to-left direction, for which no standard name exists.

defeaters $D(r_i)$: $D(r_i) = \{r_j \in \mathbf{R}^{\circ} : r_j \rhd r_i\}$. Define

$$r_i^{\rhd} = \begin{cases} (a \wedge \bigwedge_{r_j \in D(r_i)} \neg b(r_j), x) \text{ if } D(r_i) \neq \varnothing \\ (a, x) \text{ otherwise} \end{cases}$$

Set $\mathbf{R}^{\circ}_{\rhd} = \{r_1^{\rhd}, ..., r_n^{\rhd}\}$.

Th. 1 and 2 provide the central bridge: an a-world is best iff there exists a set in the maxfamily such that the world satisfies the conjunction of all formulas obtained by converting each I/O pair in that set into a material implication.

Theorem 1 (cIOL2Pref). *If $w \models a$ and $w \models \bigvee_{H \in maxf(\mathbf{R}^{\circ}_{\rhd}, a, a)} (\bigwedge \mathrm{m}(H))$, then $w \in \max_{\geq_I}(\|a\|)$, with out_4^+ as the background I/O operation.*

Theorem 2 (Pref2cIOL). *If $w \in \max_{\geq_I}(\|a\|)$, then both $w \models a$ and $w \models \bigvee_{H \in maxf(\mathbf{R}^{\circ}_{\rhd}, a, a)} (\bigwedge \mathrm{m}(H))$, with out_4^+ as the background I/O operation.*

The following follows:

Theorem 3 (Faithfulness, Hanssonian truth-conditions). *$(\varnothing, \mathbf{R}) \mid\!\sim \bigcirc_{\mathrm{H}}(x/a)$ iff $x \in \cap(outf(\mathbf{R}^{\circ}_{\rhd}, a, a))$, with out_4^+ as the background I/O operation.*

Theorem 4 ($\exists\forall$ truth-conditions). *$(\{\Diamond a\}, \mathbf{R}) \mid\!\sim \bigcirc(x/a)$ if $x \in \cap(outf(\mathbf{R}^{\circ}_{\rhd}, a, a))$, with out_4^+ as the background I/O operation.*

This embedding implies that the entailment problem for $\bigcirc_{\mathrm{H}}$ is in Π_2^P (utilizing the corresponding result for I/O logic (Sun and Robaldo 2017)). This matches the complexity of similar nonmonotonic formalisms, like default logic under skeptical reasoning (Gottlob 1992).

6 Related and Future Work

The combination of normality with deontic modalities has been considered before, but mostly within a monotonic setting; see, e.g., (Chellas 1974), although $\Rightarrow$ is not interpreted there as expressing normality. The strengthening problem can be related to what Lehmann calls the "presumption of typicality," alluded to in f.n. 5: properties of a class are assumed to hold for all its members unless there is evidence to the contrary. To quote (Lehmann 1995, §3.1), "the presumption of typicality begins where rational monotonicity [the principle that characterizes R] leaves off," and its modeling therefore requires going beyond the traditional KLM systems. One such approach is Lehmann's lexicographic entailment or the system proposed by (Delgrande 2020). Other nonmonotonic candidates include the multi-preference closure developed for description logic in, e.g., (Giordano and Dupré 2020). To the best of my knowledge, the strengthening problem in deontic logic has not been explicitly treated within a preference-based approach.

The present framework provides the basis for a number of substantial extensions. Chief among them is the design of a conflict resolution procedure based on a priority relation among obligations, intended to capture varying degrees of strength (Horty 2007; Hansen 2006; Sergot 2025). A key problem is to relate this priority relation to the ideality ordering on worlds in such a way as to yield prioritized counterparts of the postulates mentioned above, or suitable variants thereof (Prakken and Sergot 1997; Delgrande 2020). Another promising direction is the integration of permission-as-exception (Bulygin 1986). A further direction is the generalization of the embedding—not only to prioritized I/O logic (Parent 2011), but also to other norm-based formalisms. Finally, the mechanization of reasoning tasks in Isabelle/HOL, building on the automation of the monotonic basis of the deontic fragment of the logic (Benzmüller, Parent, and van der Torre 2020; Parent and Benzmüller 2024), remains an open problem.

Acknowledgements

Research funded by the Austrian Science Fund (FWF), LuCi project [10.55776/PAT1007025]. I thank the anonymous reviewers for valuable comments. Special thanks to R. Booth, J. Horty, P. McNamara and R. Ramanayake for feedback.

AI Declaration

The author has not employed any Generative AI tools.

A Proof of Th. 3 and 4

Lemma 1. *$r_i \in V(w)$ iff $w \not\models \mathrm{m}(r_i^{\rhd})$.*

Lemma 2. *If $w >_I v$, then $\exists r_i \in \mathbf{R}^{\circ}$ s.t. $w \models \mathrm{m}(r_i^{\rhd})$ and $v \not\models \mathrm{m}(r_i^{\rhd})$.*

Theorem 1 (cIOL2Pref). *If $w \models a$ and $w \models \bigvee_{H \in maxf(\mathbf{R}^{\circ}_{\rhd}, a, a)} (\bigwedge \mathrm{m}(H))$, then $w \in \max_{\geq_I}(\|a\|)$, with out_4^+ as the background I/O operation.*

Proof. Let $w \models a$ and $w \models \bigvee_{H \in maxf(\mathbf{R}^{\circ}_{\rhd}, a, a)} (\bigwedge \mathrm{m}(H))$. So $w \models \bigwedge \mathrm{m}(H)$ for some $H \in maxf(\mathbf{R}^{\circ}_{\rhd}, a, a)$.

Assume $w \notin \max_{\geq_I}(\|a\|)$. Hence $\exists v \models a$ s.t. $v >_I w$. So $V(v) \subseteq V(w)$. By Lem. 2, $\exists r_i \in \mathbf{R}^{\circ}$ s.t. $v \models m(r_i^{\rhd})$ and $w \not\models \mathrm{m}(r_i^{\rhd})$. Hence, $\{a\} \cup \mathrm{m}(r_i^{\rhd}) \not\models_{\mathrm{PL}} \bot$, and so $H \neq \varnothing$. (If not, H would not be $\subseteq$-maximal, since $\varnothing \subseteq \{r_i^{\rhd}\}$.) Suppose to reach a contradiction that $v \not\models \bigwedge \mathrm{m}(H)$. There is $r_j^{\rhd} \in H$ such that $v \not\models \mathrm{m}(r_j^{\rhd})$ while $w \models \mathrm{m}(r_j^{\rhd})$. By Lem. 1, $r_j \in V(v)$ while $r_j \notin V(w)$. This contradicts the fact that $V(v) \subseteq V(w)$. Hence $v \models \bigwedge \mathrm{m}(H)$. But $v \models a$ and $v \models \mathrm{m}(r_i^{\rhd})$. So $out_4^+(H \cup \{r_i^{\rhd}\}, a)$ is PL-consistent with a. Since $w \models \bigwedge \mathrm{m}(H)$ and $w \not\models \mathrm{m}(r_i^{\rhd})$, $r_i^{\rhd} \notin H$. As a result, H is not $\subseteq$-maximal. Contradiction. Hence $w \in \max_{\geq_I}(\|a\|)$. ☐

Theorem 2 (Pref2cIOL). *If $w \in \max_{\geq_I}(\|a\|)$, then both $w \models a$ and $w \models \bigvee_{H \in maxf(\mathbf{R}^{\circ}_{\rhd}, a, a)} (\bigwedge \mathrm{m}(H))$, with out_4^+ as the background I/O operation.*

Proof. Let $w \models a$ and $w \not\models \bigvee_{H \in maxf(\mathbf{R}^{\circ}_{\rhd}, a, a)} (\bigwedge \mathrm{m}(H))$. So ($\star$) $w \not\models \bigwedge \mathrm{m}(H)$ for all $H \in maxf(\mathbf{R}^{\circ}_{\rhd}, a, a)$. Let $\mathbf{R}^+ = \{r^{\rhd} : w \models \mathrm{m}(r^{\rhd})\}$.

Case 1: $\mathbf{R}^+ \neq \varnothing$. Since $w \models a$, $\bigwedge \mathbf{R}^+ \wedge a$ is satisfiable, and hence PL-consistent. Hence $\mathbf{R}^+ \subseteq H$ for some $H \in maxf(\mathbf{R}^{\circ}_{\rhd}, a, a)$. By ($\star$), $\mathbf{R}^+ \subset H$. By $H \in maxf(\mathbf{R}^{\circ}_{\rhd}, a, a)$, $a \wedge \mathrm{m}(H)$ is PL-consistent. By repleteness, $\exists v$ s.t. $v \models a \wedge \bigwedge \mathrm{m}(H)$.

Let $r_i := \bigcirc(y/b) \in V(v)$. By Lem. 1, $v \not\models \mathrm{m}(r_i^{\rhd})$. Suppose $w \models \mathrm{m}(r_i^{\rhd})$. By definition, $r_i^{\rhd} \in \mathbf{R}^+ \subseteq H$. Since $v \models \bigwedge \mathrm{m}(H)$, $v \models \mathrm{m}(r_i^{\rhd})$, a contradiction. Hence $w \not\models \mathrm{m}(r_i^{\rhd})$ and so $r_i \in V(w)$, by Lem. 1. Hence $V(v) \subseteq V(w)$, and so $v \geq_I w$.

Since $\mathbf{R}^+ \subset H$, $\exists r_j^{\rhd}$ s.t. $r_j^{\rhd} \in H$ and $r_j^{\rhd} \notin \mathbf{R}^+$. Since $v \models \bigwedge \mathrm{m}(H)$, $v \models \mathrm{m}(r_j^{\rhd})$, and so by Lem. 1, $r_j \notin V(v)$. Since $r_j^{\rhd} \notin \mathbf{R}^+$, $w \not\models \mathrm{m}(r_j^{\rhd})$. So again by Lem. 1, $r_j \in V(w)$. Hence $V(w) \nsubseteq V(v)$, viz., $w \ngeq_I v$.

The two imply $v >_I w$, so that $w \notin \max_{\geq_I}(\|a\|)$.

Case 2: $\mathbf{R}^+ = \varnothing$. By Lem. 1, $V(w) = \mathbf{R}^\circ$. Intuitively, w violates all the obligations in $\mathbf{R}^\circ$. Without loss of generality, assume that $\exists H \neq \varnothing \in \mathit{maxf}(\mathbf{R}^\circ_{\rhd}, a, a)$. For, if $\mathit{maxf}(\mathbf{R}^\circ_{\rhd}, a, a) = \{\varnothing\}$, then one would get $w \models \bigwedge \varnothing$ (*aka* $\top$), contradicting ($\star$). By definition, $a \wedge \bigwedge \mathrm{m}(H)$ is PL-consistent, and thus by repleteness $\exists v$ s.t. $v \models a \wedge \bigwedge \mathrm{m}(H)$. $V(v) \subseteq V(w) = \mathbf{R}^\circ$, so $v \geq_I w$. But $H \neq \varnothing$, and so $\exists r_k^{\rhd} \in H$ s.t. $v \models \mathrm{m}(r_k^{\rhd})$. By Lem. 1, $r_k \notin V(v)$, and so $V(w) \nsubseteq V(v)$. Hence $w \ngeq_I v$, and so $v >_I w$, which suffices for $w \notin \max_{\geq_I}(\|a\|)$. □

Theorem 3 (Faithfulness, Hanssonian truth-conditions)**.** *$(\varnothing, \mathbf{R}) \mid\sim \bigcirc_{\mathrm{H}}(x/a)$ iff $x \in \cap(\mathit{outf}(\mathbf{R}^\circ_{\rhd}, a, a))$, with out_4^+ as the background I/O operation.*

Proof. Left-to-right. Let $(\{\top\}, \mathbf{R}) \mid\sim \bigcirc_{\mathrm{H}}(x/a)$. Hence, for all $\mathbf{R}$-ordered model M and w in M, $w \models \bigcirc_{\mathrm{H}}(x/a)$. Let $H \in \mathit{maxf}(\mathbf{R}^\circ_{\rhd}, a, a)$. Suppose, for a reductio, that $\{a\} \cup \mathrm{m}(H) \not\models_{\mathrm{PL}} x$. Then $\{a\} \cup \mathrm{m}(H) \cup \{\neg x\}$ is PL-consistent. By repleteness, $\exists v$ s.t. $v \models \bigwedge \mathrm{m}(H)$, $v \models a$ and $v \not\models x$. Clearly, $v \models \bigvee_{H \in \mathit{maxf}(\mathbf{R}^\circ_{\rhd}, a, a)} (\bigwedge \mathrm{m}(H))$. By Th. 1, $v \in \max_{\geq_I}(\|a\|)$, and so $v \models x$, a contradiction. Hence $\{a\} \cup \mathrm{m}(H) \models_{\mathrm{PL}} x$, which suffices for $x \in \mathit{out}_4^+(H, a)$. Hence $x \in \cap(\mathit{outf}(\mathbf{R}^\circ_{\rhd}, a, a))$.

Right-to-left. Assume $x \in \cap(\mathit{outf}(\mathbf{R}^\circ_{\rhd}, a, a))$. Let M be an $\mathbf{R}$-ordered model, and w in M. Let $v \in \max_{\geq_I}(\|a\|)$. By Th. 2, $v \models a$ and $v \models \bigvee_{H \in \mathit{maxf}(\mathbf{R}^\circ_{\rhd}, a, a)} (\bigwedge \mathrm{m}(H))$. So $\exists H \in \mathit{maxf}(\mathbf{R}^\circ_{\rhd}, a, a)$ s.t. $v \models \bigwedge \mathrm{m}(H)$. But $x \in \mathit{out}_4^+(H, a)$, viz. $\{a\} \cup \mathrm{m}(H) \models_{\mathrm{PL}} x$. So $v \models x$. This shows that $w \models \bigcirc_{\mathrm{H}}(x/a)$. Hence, $(\{\top\}, \mathbf{R}) \mid\sim \bigcirc_{\mathrm{H}}(x/a)$. □

Theorem 4 ($\exists\forall$ truth-conditions)**.** *$(\{\Diamond a\}, \mathbf{R}) \mid\sim \bigcirc(x/a)$ if $x \in \cap(\mathit{outf}(\mathbf{R}^\circ_{\rhd}, a, a))$, with out_4^+ as the background I/O operation.*

Proof. Assume $x \in \cap(\mathit{outf}(\mathbf{R}^\circ_{\rhd}, a, a))$. Let M be an $\mathbf{R}$-ordered model, and w in M be s.t. $w \models \Diamond a$. Hence, $\exists v$ s.t. $v \models a$. By smoothness, $\exists u \in \max_{\geq_I}(\|a\|)$. The same argument as used for the right-to-left direction of Th. 3 yields $w \models \bigcirc_{\mathrm{H}}(x/a)$. Hence, $u \models a \wedge x$. Let s be s.t. $s \geq_I u$ and $s \models a$. We claim that $s \in \max_{\geq_I}(\|a\|)$. Let t be s.t. $t \geq_I s$ and $t \models a$. By transitivity, $t \geq_I u$. Hence $u \geq_I t$, since $u \in \max_{\geq_I}(\|a\|)$. By transitivity again, $s \geq_I t$. So $s \in \max_{\geq_I}(\|a\|)$, and so $s \models x$. By Prop. 4, $w \models \bigcirc(x/a)$. Hence, $(\{\Diamond a\}, \mathbf{R}) \mid\sim \bigcirc(x/a)$. □

References

Åqvist, L. 2002. Deontic logic. In Gabbay, D., and Guenthner, F., eds., *Handbook of Philosophical Logic*, volume 8. Dordrecht, Holland: Kluwer Academic Publishers, 2nd edition. 147–264. Originally published in (Gabbay and Guenthner 1984, pp. 605–714).

Arieli, O.; van Berkel, K.; and Straßer, C. 2024. Defeasible normative reasoning: A proof-theoretic integration of logical argumentation. In Wooldridge, M.; Dy, J.; and Natarajan, S., eds., *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024*, 10450–10458. AAAI Press.

Benferhat, S.; Cayrol, C.; Dubois, D.; Lang, J.; and Prade, H. 1993. Inconsistency management and prioritized syntax-based entailment. In *Proceedings of IJCAI*, IJCAI'93, 640–645. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Benzmüller, C.; Parent, X.; and van der Torre, L. 2020. Designing normative theories for ethical and legal reasoning: Logikey framework, methodology, and tool support. *Artif. Intell.* 287:103348. Extended abstract presented at *KR 2020* (Recently Published Research track).

Bochman, A. 2005. *Explanatory Nonmonotonic Reasoning*. Word Scientific.

Boutilier, C. 1994. Conditional logics of normality: A modal approach. *Artif. Intell.* 68(1):87–154.

Broersen, J., and van der Torre, L. 2012. Ten problems of deontic logic and normative reasoning in computer science. In Bezhanishvili, N., and Goranko, V., eds., *ESSLLI 2010*. Berlin, Heidelberg: Springer. 55–88.

Bulygin, E. 1986. Permissive norms and normative systems. In Martino, A., and Natali, F. S., eds., *Automated Analysis of Legal Texts*, 211–218. North-Holland: Elsevier.

Casini, G.; Meyer, T.; and Varzinczak, I. 2019. Taking defeasible entailment beyond rational closure. In Calimeri, F.; Leone, N.; and Manna, M., eds., *JELIA*, volume 11468 of *LNCS*, 182–197. Springer.

Chellas, B. 1974. Conditional obligation. In Stenlund, S.; Henschen-Dahlquist, A.-M.; Lindahl, L.; Nordenfelt, L.; and Odelstad, J., eds., *Logical Theory and Semantic Analysis*. Dordrecht: Springer Netherlands. 23–33.

Chellas, B. 1980. *Modal Logic*. Cambridge: Cambridge University Press.

Chisholm, R. 1963. Contrary-to-duty imperatives and deontic logic. *Analysis* 24:33–36.

Ciabattoni, A., and Rozplokhas, D. 2023. Streamlining input/output logics with sequent calculi. In Marquis, P.; Son, T. C.; and Kern-Isberner, G., eds., *Proceedings of KR 2023*, 146–155.

Delgrande, J. P. 2020. A preference-based approach to defeasible deontic inference. In Calvanese, D.; Erdem, E.; and Thielscher, M., eds., *Proceedings of KR 2020*, 326–335.

Forrester, J. 1984. Gentle murder, or the adverbial samaritan. *J. Philos.* 81(4):193–197.

Freund, M. 1998. Preferential reasoning in the perspective of Poole default logic. *Artif. Intell.* 98(1):209–235.

Gabbay, D., and Guenthner, F., eds. 1984. *Handbook of Philosophical Logic*, volume II. Dordrecht, Holland: Reidel, 1st edition.

Gabbay, D.; Horty, J.; Parent, X.; van der Meyden, R.; and van der Torre, L., eds. 2013. *Handbook of Deontic Logic and Normative Systems*, volume 1. London: College Publications.

Giordano, L., and Dupré, D. T. 2020. An ASP approach for reasoning in a concept-aware multipreferential lightweight DL. *Theory Pract. Log. Program.* 20(5):751–766.

Goble, L. 2003. Preference semantics for deontic logic. Part 1–Simple models. *Logique et Analyse* 46(183/184):383–418.

Goble, L. 2004. A proposal for dealing with deontic dilemmas. In Lomuscio, A., and Nute, D., eds., *Deontic Logic in Computer Science*, volume 3065 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg. 74–113.

Goble, L. 2013. Prima facie norms, normative conflicts and dilemmas. In Gabbay et al. (2013). 241–35.

Goble, L. 2019. Axioms for Hansson's dyadic deontic logics. *Filosofiska Notiser* 6(1):13–61.

Gottlob, G. 1992. Complexity results for nonmonotonic logics. *J. Log. Comput.* 2(3):397–425.

Halpern, J. 1997. Defining relative likelihood in partially-ordered structures. *Artif. Intell. Res.* 7:1–24.

Hansen, J. 2006. Deontic logics for prioritized imperatives. *Artif. Intell. Law* 14(1-2):1–34.

Hansen, J. 2008. Prioritized conditional imperatives: problems and a new proposal. *Auton. Agents Multi-Agent Syst.* 17(1):11–35.

Hansson, B. 1969. An analysis of some deontic logics. *Noûs* 3(4):373–398. Reprinted in (Hilpinen 1971, pp. 121-147).

Hilpinen, R., ed. 1971. *Deontic Logic: Introductory and Systematic Readings*. Dordrecht: Reidel.

Horty, J. 1993. Deontic logic as founded on nonmonotonic logic. *Ann. Math. Artif. Intell.* 9:69–91.

Horty, J. 1994. Moral dilemmas and nonmonotonic logic. *J. Philos. Log.* 23(1):35–65.

Horty, J. 2007. Defaults with priorities. *J. Philos. Log.* 36(4):367–413.

Horty, J. 2012. *Reasons as Defaults*. Oxford University Press.

Horty, J. 2014. Deontic modals: Why abandon the classical semantics? *Pac. Philos. Q.* 95(4):424–460.

Jackson, F. 1985. On the semantics and logic of obligation. *Mind* 94(374):177–195.

Jørgensen, J. 1937. Imperatives and logic. *Erkenntnis* 7(1):288–296.

Kratzer, A. 2012. The notional category of modality. In Kratzer, A., ed., *Modals and Conditionals*. Oxford University Press. 27–69.

Kraus, S.; Lehmann, D.; and Magidor, M. 1990. Nonmonotonic reasoning, preferential models and cumulative logics. *Artif. Intell.* 44(1-2):167–207.

Lamarre, P. 1991. S4 as the conditional logic of nonmonotonicity. In *Proceedings of KR 1991*, 357–367. Morgan Kaufmann Publishers Inc.

Lehmann, D., and Magidor, M. 1992. What does a conditional knowledge base entail? *Artif. Intell.* 55(1):1–60.

Lehmann, D. 1995. Another perspective on default reasoning. *Ann. Math. Artif. Intell.* 15(1):61–82.

Lewis, D. 1973. *Counterfactuals*. Oxford: Blackwell.

Lewis, D. 1974. Semantic analyses for dyadic deontic logic. In Stenlund, S.; Henschen-Dahlquist, A.-M.; Lindahl, L.; Nordenfelt, L.; and Odelstad, J., eds., *Logical Theory and Semantic Analysis*, volume 63 of *Synthese Library*. Netherlands: Springer. 1–14.

Makinson, D., and van der Torre, L. 2000. Input/output logics. *J. Philos. Log.* 29(4):383–408.

Makinson, D., and van der Torre, L. 2001. Constraints for input/output logics. *J. Philos. Log.* 30(2):155–185.

Makinson, D. 1988. General theory of cumulative inference. In Reinfrank, M.; de Kleer, J.; Ginsberg, M.; and Sandewall, E., eds., *Non-Monotonic Reasoning, 2nd International Workshop, Grassau*, volume 346, 1–18. Springer.

Makinson, D. 1993. Five faces of minimality. *Stud. Log.* 52(3):339–380.

Makinson, D. 1999. On a fundamental problem of deontic logic. In Mc Namara, P., and Prakken, H., eds., *Norms, Logics and Information Systems*, Frontiers in Artificial Intelligence and Applications. Amsterdam: IOS Press. 29–53.

McCarty, L. T. 1994. Defeasible deontic reasoning. *Fundamenta Informaticae* 21:125–148.

Parent, X., and Benzmüller, C. 2024. Conditional normative reasoning as a fragment of HOL. *J. Appl. Non Class. Logics* 34(4):561–592.

Parent, X., and van der Torre, L. 2013. Input/output logic. In Gabbay et al. (2013). 499–544.

Parent, X., and van der Torre, L. 2014. Aggregative deontic detachment for normative reasoning. In Baral, C.; Giacomo, G. D.; and Eiter, T., eds., *Proceedings of KR 2014*. AAAI Press.

Parent, X., and van der Torre, L. 2017. Detachment in normative systems: Examples, inference patterns, properties. *FLAP* 4(9):2995–3038.

Parent, X., and van der Torre, L. 2018. I/O logics with a consistency check. In Broersen, J. M.; Condoravdi, C.; Shyam, N.; and Pigozzi, G., eds., *Deontic Logic and Normative Systems - DEON 2018*, 285–299. College Publications.

Parent, X. 2011. Moral particularism in the light of deontic logic. *Artif. Intell. Law* 19(2-3):75–98.

Parent, X. 2021. A modal characterisation of an intuitionistic I/O operation. *FLAP* 8(8):2349–2362.

Parent, X. 2026. Defeasible conditional obligation in a two-tiered preference-based semantics (extended version). https://doi.org/10.48550/arXiv.2604.26977.

Pearl, J. 1990. System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In Parikh, R., ed., *Proceedings of TARK*, 121–135. Morgan Kaufmann.

Prakken, H., and Sergot, M. 1996. Contrary-to-duty obligations. *Stud Logica* 57(1):91–115.

Prakken, H., and Sergot, M. 1997. Dyadic deontic logic and contrary-to-duty obligations. In Nute, D., ed., *Defeasible Deontic Logic*. Dordrecht: Kluwer. 223–262.

Ryle, G. 1938. Categories. *Proceedings of the Aristotelian Society* 38(4):189–206.

Ryu, Y., and Lee, R. 1994. Defeasible deontic reasoning: a logic programming model. In Meyer, J.-J. C., and Wieringa, R. J., eds., *Deontic Logic in Computer Science: Normative System Specification*. UK: John Wiley and Sons Ltd. 225–241.

Sergot, M. 2025. Defeasible conditional imperatives. 1st Polish Congress of Logic, Torun, September 2025 (Invited lecture).

Stolpe, A. 2020. Unsettling preferential semantics. *J. Philos. Log.* 49(2):371–399.

Straßer, C.; Beirlaen, M.; and Putte, F. V. D. 2016. Adaptive logic characterizations of input/output logic. *Stud Logica* 104(5):869–916.

Sun, X., and Robaldo, L. 2017. On the complexity of input/output logic. *J. Appl. Log.* 25:69–88.

Tan, Y.-H., and van der Torre, L. 1995. Why defeasible deontic logic needs a multi preference semantics. In Froidevaux, C., and Kohlas, J., eds., *Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, 412–419. Berlin, Heidelberg: Springer.

van Benthem, J.; Grossi, D.; and Liu, F. 2014. Priority structures in deontic logic. *Theoria* 80(2):116–152.

van der Torre, L. 1997. *Reasoning about Obligations*. Ph.D. Dissertation, University of Erasmus.

van Fraassen, B. 1972. The logic of conditional obligation. *J. Philos. Log.* 1(3/4):417–438.