How to Take Heroin (if at all). Holistic Detachment in Deontic Logic

Frederik Van De Putte
Ghent University
frederik.vandeputte@ugent.be

Stef Frijters
Ghent University
stef.frijters@ugent.be

Joke Meheus
Ghent University
joke.meheus@ugent.be

Abstract

The aim of the present paper is to investigate the logic of *holistic detachment*, i.e. detachment that is triggered by *all and only* those circumstances that are fixed (unalterable, unavoidable). To this end, we present the (monotonic) modal logic **HD** that captures the distinction between mere facts and fixed circumstances, and features a non-normal "all and only"-operator. We give a sound and (strongly) complete axiomatization of **HD** and discuss its most salient properties. We show that **HD** is rather weak when applied to realistic scenarios and we argue against what we call the "enthymematic approach" to mitigate this weakness. In contrast, it is shown that **HD** can and should be strengthened non-monotonically, in order to capture deontic reasoning.

Keywords: deontic logic, conditional oughts, detachment, all that is fixed

We are indebted to Christian Straßer, John Horty, Eric Pacuit, and Federico Faroldi for valuable comments on previous drafts of the paper.

1 Introduction

Consider the following dialogue, inspired by [22]:

David: Caroline ought not to take heroin. However, if she does take heroin, then she ought to take a light dose and use a clean needle.
Lou: Caroline does take heroin.
David: Then she ought to use a clean needle and take a light dose.
Lou: No. She ought not to care about needles or dosage – she simply ought not to take heroin in the first place!
David: But if she does not use a clean needle, she might get HIV-infected...

Arguably, Lou and David are talking past one another in this little dialogue. What David means is that, if Caroline's taking heroin is taken for granted, if this is a fixed feature of the situation she is in, then Caroline ought to use a clean needle and take a light dose. In contrast, Lou entertains the possibility of Caroline *not* taking heroin, and that possibility is clearly preferable to her taking a light dose of heroin and using a clean needle.

This point is not new. One of the insights brought up by the discussion on contrary-to-duties and the associated paradoxes is that, whether one can ratioinally detach a conditional obligation in the light of factual information, depends on what one takes to be "fixed circumstances" [13; 10; 9; 22; 7]. Not all facts have the same status: one may e.g. consider Caroline's current unemployment as a fixed fact, but not her drug abuse. Unfortunately, it turns out to be very hard to pin down when one should consider some truth φ merely contingent and when φ is a circumstance that can trigger detachment.[1] The obvious way out of this puzzling question – at least for the deontic logician – is to build this distinction into one's logic. Once there, detachment can be formalized by relying on the following principle:[2]

> **Restricted Detachment:** one should take *only* fixed circumstances[3] into account, when detaching oughts.

[1] See [7, pp. 283-284] for an attempt to do so.

[2] Restricted detachment (RD) is a refinement of *factual* detachment (FD), i.e. the derivation of "actual" or "situational" obligations from conditional obligations and information about the facts at hand [23, p. 118]. Factual detachment is usually opposed to *deontic* detachment (DD), which concerns the derivation of actual obligations from conditional obligations and other actual obligations. It is well-known, at least since [1] that (FD) and (DD) cannot easily be combined. Whereas some argue that (RD) can be combined with (DD), we leave the issue of (DD) for another occasion.

[3] In the remainder, we simply use "fixed circumstances" as shorthand for "those circumstances that are taken to be fixed by the person who reasons about the situation in question".

An obvious next question is: *how much* of the fixed circumstances do we need to take into account when applying detachment? To continue the above example: suppose that Caroline has a severe heroin addiction. This means that if she takes only a small dose of heroin, in the absence of medical supervision, she will suffer from dangerous withdrawal effects. Let us agree on the following conditional, and let us moreover agree that the factual claims below it are fixed:

(C) If Caroline takes heroin, and if she is a heroin addict, then she ought to take a sufficiently large dose of heroin.
(F1) Caroline takes heroin.
(F2) Caroline is a heroin addict.

In this extension of our first scenario, it seems Caroline ought *not* to take a light dose of heroin; in fact, she ought to take a dose that is sufficiently large, in order to avoid serious withdrawal effects. This intuition can be explained by the following credo:[4]

> **Holistic Detachment:** one should take *all and only* the fixed circumstances into account, when detaching oughts.

Although there are various logics that capture the principle of restricted detachment in one way or another, only few have attempted to target its holistic counterpart and explicate it in exact, formal terms.[5] This is the aim of the present paper. We present the (monotonic) modal logic **HD** which captures the distinction between mere facts and fixed circumstances, and which validates a rule of detachment triggered by "all that is fixed" (Section 2). In Section 3, we show that this logic is rather weak when applied to realistic scenarios and we argue against what we call the "enthymematic approach" to mitigate this weakness. In contrast, we show that **HD** can and should be strengthened non-monotonically. Section 4 gives a brief survey of related work and Section 5 concludes the paper.

Before we proceed, two disclaimers are in order. First, unlike many other existing accounts, ours yields a conservative extension of Standard Deontic Logic. This implies that it inherits some paradoxes of the latter, but also all its inferential

[4] In his discussion of ethical reasoning, Jonsen seems to argue in favour of a principle akin to our holistic detachment, where he writes that "[...] the ultimate view of the case and its appropriate resolution comes, not from a single principle, nor from a dominant theory, but from the converging impression made by all of the relevant facts and arguments [...]" [26, p. 245]. Simplifying Jonsen's view somewhat, one can take arguments to be deontic conditionals, and relevant facts to be the fixed circumstances.

[5] We consider approaches similar or related to ours in Section 4.

power. In the current paper, we will remain mostly silent on those paradoxes, since we consider their solution to be orthogonal to the issue of detachment. We do however believe that giving up on the inferential power of **SDL** should not be taken too lightly.

Second, in this paper we focus on conditional, defeasible oughts that concern one particular agent – often these claims can be interpreted as forms of "advice" to the person in question. So we focus on tentative claims of the type "if you do X, then you ought to do Y" (where the agent is the same in the antecedent and consequent, or neither involves any agency). This can be contrasted with legal claims such as "if you drive above the speed limit, and if you are not in circumstance X_1, ..., or X_n, then you must be fined". In the latter case, the exceptions are usually made explicit in the normative system, and the consequent of the conditional concerns an action of the legislator or an agent-independent proposition, not an action of the one who violates the norm in question. We will not discuss this distinction here in detail, but merely flag it to avoid any confusion.[6]

2 The monotonic logic HD

In this section we present a monotonic logic for holistic detachment. Even though its underlying intuitions seem straightforward, they give rise to a rich system with some surprising interaction principles (cf. Section 2.3).

2.1 Formal language

Fix a countable set $\mathcal{S} = \{p, q, r, \ldots\}$ of sentential variables. The set of wffs of **HD**, $\mathcal{W}$, is obtained by closing $\mathcal{S} \cup \{\top, \bot\}$ under the classical truth-functional connectives $\neg, \vee, \wedge, \rightarrow, \leftrightarrow$, the unary operators $\mathsf{U}, \vec{\Box}, \overleftrightarrow{\Box}, \mathsf{O}$ and the binary operator $\mathsf{O}(.|.)$. We treat $\bot$, $\neg$ and $\vee$ as primitive, $\top$ and the other classical connectives are defined in the usual way.

U is a global modality in the sense of [18]; it simplifies the axiomatization of the logic and will turn out highly useful in defining the non-monotonic extensions of **HD**.[7] $\vec{\Box}$ is a normal modality of the type **KT**, and is used to express the properties of the situation that are fixed – one may also call those properties unalterable or unavoidable, cf. [7]. $\mathsf{O}(.|.)$ allows us to express the conditional oughts that are used in our deliberation, in order to determine the obligations that apply to the case at hand – the latter are then formalized using O, which is a normal modality of the type

[6] Due to space limitations we had to omit the Appendix with meta-proofs in this manuscript. They are included in the online version of this article, available at `www.clps.ugent.be/`.

[7] See our definition of Δ_2 on page 12.

KD. Following [22], we read O as an operator for "situation-specific obligation", or more briefly, "situational obligation".

$\overleftrightarrow{\Box}$ is an "all and only" modality in the sense of [28] and [25]. The formula $\overleftrightarrow{\Box}\varphi$ allows us to express that φ is *all* that is fixed, and plays a crucial role in the detachment rule of **HD** (see the axiom (DET) in Section 2.3).

One can express various types of violations in $\mathcal{W}$. $\mathsf{O}(\varphi|\top) \wedge \neg\varphi$ stands for "the general obligation that φ is violated"; $\mathsf{O}(\varphi|\top) \wedge \vec{\Box}\neg\varphi$ expresses that this violation is fixed. $\mathsf{O}\varphi \wedge \neg\varphi$ should be read as "the situational obligation that φ is violated". Finally, $\mathsf{O}(\varphi|\top)\wedge\mathsf{O}\neg\varphi$ expresses that one has the situational obligation to violate the general obligation that φ. Each of these expressions are contingent in our logic (for contingent φ). One may further refine the formal language and distinguish various levels of "fixedness" (essentially generalizing the picture drawn in [7]), but we leave that aside here.

One may also define a different monadic operator for obligation:

$$\mathsf{O}_a\varphi =_{\mathsf{df}} \mathsf{O}\varphi \wedge \neg\vec{\Box}\varphi$$

The operator O_a speaks of situational obligations that are not vacuous, in the sense that $\mathsf{O}_a\varphi$ is true iff φ is true in all acceptable worlds relative to the case at hand, but φ is not fixed. Such a φ may however still be more or less specific. As a result, O_a still suffers from some paradoxes akin to the Ross paradox in Standard Deontic Logic.[8] As noted in the introduction, we will largely ignore those paradoxes, and focus on O in most of what follows. We do however briefly return to O_a in Section 4.

2.2 Semantics

Definition 1. *An* **HD**-model *is a tuple* $M = \langle W, R, f, V\rangle$ *where*

(C1) $W \neq \emptyset$ *is the* domain *of* M
(C2) $R \subseteq W \times W$ *is reflexive*
(C3) $f : W \times \wp(W) \to \wp(W)$ *is a function that satisfies the following conditions:*
  (C3.1) *where* $w \in W$ *and* $\emptyset \neq X \subseteq W$: $f(w, X) \neq \emptyset$
  (C3.2) *where* $w \in W$ *and* $X \subseteq W$: $f(w, X) \subseteq X$
(C4) $V : \mathcal{S} \to \wp(W)$ *is a valuation function*

[8] For instance, we have the following variant of the Ross paradox: $\mathsf{O}_a\varphi \wedge \neg\vec{\Box}(\varphi \vee \psi) \vdash \mathsf{O}_a(\varphi \vee \psi)$. So if (according to this reading) it is obligatory that one mails the letter, and if mailing the letter or burning it is not fixed, then it is obligatory that one mails or burns it.

Definition 2. *Where $M = \langle W, R, f, V\rangle$ is an* **HD***-model and $w \in W$,*

(SC0) $M, w \models \varphi$ *iff* $w \in V(\varphi)$ *for all* $\varphi \in \mathcal{S}$
(SC1) $M, w \not\models \bot$
(SC2) $M, w \models \neg\varphi$ *iff* $M, w \not\models \varphi$
(SC3) $M, w \models \varphi \vee \psi$ *iff* $M, w \models \varphi$ *or* $M, w \models \psi$
(SC4) $M, w \models \mathsf{U}\varphi$ *iff for all* $w' \in W$, $M, w' \models \varphi$
(SC5) $M, w \models \vec{\Box}\varphi$ *iff* $R(w) \subseteq \|\varphi\|^M$
(SC6) $M, w \models \overleftrightarrow{\Box}\varphi$ *iff* $R(w) = \|\varphi\|^M$
(SC7) $M, w \models \mathsf{O}(\varphi|\psi)$ *iff* $f(w, \|\psi\|^M) \subseteq \|\varphi\|^M$
(SC8) $M, w \models \mathsf{O}\varphi$ *iff* $f(w, R(w)) \subseteq \|\varphi\|^M$

where, $\|\varphi\|^M = \{w' \in W \mid M, w' \models \varphi\}$ *and* $R(w) = \{w' \in W \mid Rww'\}$.

Semantic consequence ($\Gamma \Vdash \varphi$) and validity ($\Vdash \varphi$) are defined in the standard way. The interesting (since non-standard) clauses are those for $\overleftrightarrow{\Box}$ (which corresponds to our intuitive reading of "all that is fixed"), and the one for O, which refers to both R and f.

Intuitively, $R(w)$ corresponds to the set of worlds $w' \in W$ that are available, in view of those circumstances that are fixed at w. The requirement that R is reflexive is motivated by the idea that, from the perspective of the person who reasons about a situation, whatever is fixed also obtains in the current world.

The function f is used to interpret deontic conditionals. Intuitively, $f(w, X)$ is the set of worlds $w' \in X$ that would be acceptable from the viewpoint of w, if X would coincide with one's options at w. We require that, unless φ is impossible in M, $f(w, \|\varphi\|^M) \neq \emptyset$. In other words: conditional on a proposition that is possible, one cannot be obliged to do the impossible.

Semantic clause (SC8) shows that situational obligations are a function of conditional obligations and the fixed circumstances. Note that, since R is reflexive, $R(w)$ is guaranteed to be non-empty, and hence so is $f(w, R(w))$ by condition (C3.1). In view of (SC8) this guarantees that one gets a normal modal operator of type **KD**. As a result, **HD** is a conservative extension of Standard Deontic Logic.

Recall that, in our example from the introduction, David and Lou had different views on what counts as fixed circumstances for their deontic reasoning. Such differences correspond, in our semantics, to a difference concerning the set $R(w)$. At one extreme, everything that happens to be true in our current world is fixed, and hence $R(w) = \{w\}$. This will trivialize the concept of situational obligation, since $\varphi \rightarrow \mathsf{O}\varphi$ becomes valid under this condition. At the other extreme, one only considers those circumstances fixed that are logically unavoidable: $R(w) = W$. This implies that $\mathsf{O}\varphi$ becomes equivalent to $\mathsf{O}(\varphi|\top)$. Note that in general, $\mathsf{O}\varphi$ and $\mathsf{O}(\varphi|\top)$ are logically

independent in **HD**. Whereas $\mathsf{O}\varphi$ expresses that φ is obligatory in view of the fixed circumstances, $\mathsf{O}(\varphi|\top)$ can be read as "absent further information, φ is obligatory".

We do not explicitly model the difference in view between David and Lou at the object level of our logic. Rather, we see this as a difference in the premises they endorse, or alternatively, as a difference in the models each of them considers. $\vec{\Box}$ and $\overleftrightarrow{\Box}$ should hence be interpreted here in a metaphysical, not in an epistemological or doxastic sense: they express what is true of the situation at hand, not what a given agent knows or believes to be true. Accordingly, O does not represent belief-based or knowledge-based obligation in the sense of [32; 12].

2.3 Axiomatization

A sound and strongly complete axiomatization of **HD** is obtained by closing a complete axiomatization for classical logic together with all instances of the axiom schemata in Table 1 under necessitation for U and modus ponens.[9] φ is an **HD**-theorem ($\vdash \varphi$) iff φ can be derived from the **HD**-axioms and rules. φ is **HD**-derivable from Γ ($\Gamma \vdash \varphi$) iff there are $\psi_1, \ldots, \psi_n \in \Gamma$ such that $\vdash (\psi_1 \wedge \ldots \wedge \psi_n) \rightarrow \varphi$.[10]

	S5 for U	(UB)	$\mathsf{U}\varphi \rightarrow \vec{\Box}\varphi$
	KT for $\vec{\Box}$	(UC)	$\mathsf{U}\varphi \rightarrow \mathsf{O}(\varphi\|\psi)$
	KD for O	(BO)	$\vec{\Box}\varphi \rightarrow \mathsf{O}\varphi$
(CG)	$\mathsf{U}(\varphi \leftrightarrow \psi) \rightarrow (\mathsf{O}(\tau\|\varphi) \rightarrow \mathsf{O}(\tau\|\psi))$	(AO1)	$\overleftrightarrow{\Box}\varphi \rightarrow \vec{\Box}\varphi$
(CK)	$(\mathsf{O}(\psi\|\varphi) \wedge \mathsf{O}(\psi \rightarrow \tau\|\varphi)) \rightarrow \mathsf{O}(\tau\|\varphi)$	(AO2)	$(\vec{\Box}\varphi \wedge \overleftrightarrow{\Box}\psi) \rightarrow \mathsf{U}(\psi \rightarrow \varphi)$
(CP)	$\neg\mathsf{U}\neg\varphi \rightarrow (\mathsf{O}(\psi\|\varphi) \rightarrow \neg\mathsf{O}(\neg\psi\|\varphi))$	(AO3)	$\mathsf{U}(\varphi \leftrightarrow \psi) \rightarrow (\overleftrightarrow{\Box}\varphi \leftrightarrow \overleftrightarrow{\Box}\psi)$
(CI)	$\mathsf{O}(\varphi\|\varphi)$	(DET)	$(\overleftrightarrow{\Box}\varphi \wedge \mathsf{O}(\psi\|\varphi)) \rightarrow \mathsf{O}\psi$
		(ATT)	$(\overleftrightarrow{\Box}\varphi \wedge \mathsf{O}\psi) \rightarrow \mathsf{O}(\psi\|\varphi)$

Table 1: Axiom schemata for **HD**.

(CG) follows from the fact that the function f operates on sets of worlds, rather than formulas. The axioms (CK) and (UC) together with necessitation for U imply that, given that one holds the antecedent φ fixed, one can read $\mathsf{O}(.|\varphi)$ as a normal modal operator. (CP) and (CI) correspond to conditions (C3.1), respectively (C3.2) in Definition 1.

(UB) and (UC) follow from the fact that U is a global modality, and that both $\vec{\Box}$ and $\mathsf{O}(.|\varphi)$ (for fixed φ) are normal modalities. The bridging principle (BO) follows

[9] Note that this entails necessitation for $\vec{\Box}$ and O as well, in view of axiom (UB), resp. (BO).

[10] Note that this syntactic consequence relation is by definition compact.

from the semantic clause for O: if a given alternative is acceptable, conditional on all that is fixed, then it must be one that is still available given all that is fixed; hence it must make all the fixed circumstances true.

(AO1) and (AO2) express interactions between the normal modal operator $\vec{\Box}$ and its "all and only"-counterpart. Together with Necessitation for U, (AO3) entails that $\overleftrightarrow{\Box}$ is a classical operator in the sense of [11]. Note that, using (AO1)-(AO3), one can derive the following theorem:

(AO4) $(\overleftrightarrow{\Box}\varphi \wedge \overleftrightarrow{\Box}\psi) \rightarrow \mathsf{U}(\varphi \leftrightarrow \psi)$

(DET) corresponds to our notion of holistic detachment. It can be seen as an introduction rule for O and as an elimination rule for $\mathsf{O}(.|.)$. Interestingly, with the current semantics we also get an elimination rule for O that allows us to introduce new conditionals, viz. (ATT) (for "attachment"). This rule says that, if you are in a situation where ψ is an unconditional obligation, and if φ provides an adequate and complete description of the fixed circumstances in that situation, then the conditional $\mathsf{O}(\psi|\varphi)$ is true.

Before closing this section, let us briefly mention some possibilities for varying on the above semantics. First, one may consider weaker or stronger requirements on the accessibility relation R, in line with traditional distinctions in normal modal logics. In [7, p. 291] it is argued that fixed propositions need not be true. Technically, this option – i.e. to give up reflexivity and the associated T-schema for $\vec{\Box}$ – poses no problems; the completeness proof can be run just as before. Alternatively, one may consider more restricted classes of models, where e.g. R is required to be transitive and/or symmetric. For instance, the logic of all models where R is an equivalence relation is characterized by adding the **S5**-axioms for $\vec{\Box}$ to our axiomatization of **HD**, together with the following two axiom schemata for $\overleftrightarrow{\Box}$:[11]

(S5$\overleftrightarrow{\Box}$-1) $\overleftrightarrow{\Box}\varphi \leftrightarrow \overleftrightarrow{\Box}\overleftrightarrow{\Box}\varphi$
(S5$\overleftrightarrow{\Box}$-2) $\neg\overleftrightarrow{\Box}\varphi \rightarrow \vec{\Box}\neg\overleftrightarrow{\Box}\varphi$

Another type of variation would be obtained by imposing additional requirements on the deontic function f. In particular, one may require that f is constant, in the sense that for all $w, w' \in W$ and all $X \subseteq W$, $f(w, X) = f(w', X)$. This condition makes it possible to capture (an abstract form of) reasoning from cases to conditional norms, as it validates all instances of the following schema:

$$\neg\vec{\Box}\neg(\overleftrightarrow{\Box}\varphi \wedge \mathsf{O}\psi) \rightarrow \mathsf{O}(\psi|\varphi)$$

[11] We sketch the completeness proof for this variant in the Appendix to the full version of this paper, available at `www.clps.ugent.be/`.

An exploration of these and other frame conditions, and the axiomatization of the resulting logics is left for future work.

3 Strengthening HD

Although **HD** has interesting features and is very expressive (cf. supra), it is also inferentially weak in at least two respects. We first explain why, after which we consider various ways one can strengthen **HD** and thus allow for a more realistic formalization of reasoning with deontic conditionals.[12]

3.1 All that is fixed?

If we were to formalize David's view in the example from the introduction in **HD**, the following premises seem natural:

(P1) $\mathsf{O}(\neg p|\top)$ — "In general, Caroline ought not to take heroin."
(P2) $\mathsf{O}(q \wedge s|p)$ — "If Caroline takes heroin, then she ought to take a light dose and use a clean needle."
(P3) p — "Caroline takes heroin."
(P4) $\vec{\Box}p$ — "It is fixed that Caroline takes heroin."

Note that Lou agrees with David on (P1)-(P3), but rejects (P4). Now, does it follow from (P1)-(P4) that Caroline ought to take a light dose and use a clean needle? In other words, does $\mathsf{O}(q \wedge s)$ follow from these premises? Intuitively it might, but it does not in **HD**. The reason is simple: the premises leave it open that there are propositions other than (and independent of) p that are also fixed. Fully in line with the idea behind **HD**, one needs to know that p expresses *all* that is fixed, before one can apply detachment. But unfortunately, one can never derive $\vec{\Box}p$ from (P1)-(P4). More generally:

Theorem 1. *Let $\Gamma \subseteq \mathcal{W}$ be an* **HD***-consistent set such that $\vec{\Box}$ occurs in no member of Γ. Then there is no φ such that $\Gamma \vdash \vec{\Box}\varphi$.*

In view of this theorem, there is a logical gap between formulas that express fixed circumstances and formulas of the form $\vec{\Box}\varphi$. So if we formalize David's reasoning, and if we want to arrive at the appropriate conclusion using **HD**, then we should add $\vec{\Box}p$ as a premise. We will return to this point below, but first consider a different problem for **HD**.

[12] What we write below applies just as well to the stronger logics obtained by imposing one or more frame conditions like the ones we discussed at the end of Section 2.3. So this is really a problem of the approach in general, not of the specificities of **HD**.

3.2 Applying general norms to specific cases

Suppose that we add the following premise to (P1)-(P4):

(P5) $\overset{\leftrightarrow}{\Box}(p \wedge r)$ — "Caroline takes heroin and has a child, and this is all that is fixed"

In this case, we do have information about all that is fixed. Still, we cannot detach that Caroline ought to use a clean needle and take a light dose of heroin. The reason is that the deontic conditional $\mathsf{O}(q \wedge s|p)$ only speaks about those situations in which all that is fixed coincides with p. One obvious way out would be to assume that conditional obligations are closed under strengthening of the antecedent (henceforth, SA): from $\mathsf{O}(\varphi|\psi)$, to infer $\mathsf{O}(\varphi|\psi \wedge \tau)$. Semantically, this corresponds to the condition: if $X \subseteq Y$, then $f(w, X) \subseteq f(w, Y)$.

The problem with this move is that it implies a very strong reading of the deontic conditional: what advice can one ever give that is not overruled in certain very specific circumstances? Consider our second example from the introduction: there we have a clear exception to the general rule concerning heroin, which is made explicit only after the general rule was stated. This exception does *not* generate a conflict at the level of the eventual advice one will give: it simply *blocks* the application of the more general rule. If (SA) is built into the logic, then either one must rule out the possibility of such posterior exceptions – and hence, have all exception clauses built into one's deontic conditionals from the start –, or one should treat exceptions as merely "other considerations" that are on equal footing with the specific variant of the general rule that can be derived by (SA). Moreover, if all exception clauses are explicitly stated as part of the general rule, then one should also assume that all the negations of those clauses are fixed circumstances, in order to solve the problem noted in Section 3.1.

3.3 Tacit premises?

Each of the above problems can easily be tackled if we just add certain premises to our formalization of the examples in question. For the problem of specificity (Section 3.2), this means one would add the following premise:

(P6) $\mathsf{O}(q|p \wedge r)$ — "If Caroline takes heroin and has a child, then she ought to take a light dose."

In other words, the argument from (P1)-(P5) to $\mathsf{O}q$ is treated as an *enthymeme*: an argument that draws on a tacit premise – i.c. (P6) – that is endorsed by anyone who reasons about the example.

The enthymematic approach (as we shall call it) to deontic reasoning is not new. For instance, in his work on conflict-tolerant deontic logics, Goble developed logics

which allow for restricted forms of aggregation (from $O\varphi, O\psi$ to infer $O(\varphi \wedge \psi)$) and restricted forms of inheritance (from $O\varphi$ and $\varphi \vdash \psi$, to infer $O\psi$).[13] Goble's restrictions are of the type "it is possible that τ" or "it is permitted that τ". In order to make natural examples of deontic reasoning work, one then has to treat such possibility or permissibility claims as tacit premises. In a similar vain, Carmo and Jones [7] need to add the premises $\neg\vec{\Box}\varphi$ and $\neg\vec{\Box}\neg\varphi$ in order to get the inference from $\vec{\Box}\psi, \mathsf{O}(\varphi|\psi)$ to $\mathsf{O}\varphi$ off the ground.

In itself, the enthymematic approach should not be rejected: it is a fact of life that we do not always make all our premises explicit, and it is a virtue of logic that it forces us to do so. However, in the case of **HD**, logic can and should do more.[14] The (allegedly) implicit premise (P6) is not simply some "general relevant background information": it bears a specific, *formal* relation to the explicit premise (P2): (P6) can be obtained from (P2) by (SA). Likewise, (P5) has a formal relation to (P4) and to the premises (P1)-(P4) as a whole: $\overleftrightarrow{\Box}p$ entails (P4) and it is compatible with each of the other premises. As these examples show, formal tools *can* help us clarify at least some of the tacit premises that are at stake.

Such help is indispensable as soon as one considers more complex scenarios, where proper logical calculations will be required to determine which tacit premises are mutually compatible. Note that, as soon as we go to first order predicative languages, there is not even a positive test for joint consistency of the explicit premises with the tacit ones. Moreover, if we want to model the dynamics of reasoning with conditionals, we should be able to accommodate cases where explicit premises are added along the way, as we reason. In such cases, one would have to double-check consistency with previously added tacit premises, change them again, etc. Describing such a procedure in exact terms will result, essentially, in a formalism much like the one we describe below.

3.4 Going non-monotonic

In view of the preceding, one should strengthen **HD** by adding certain defeasible rules of inference. There are several ways to do so – see [29] for a reader-friendly introduction to the field. For reasons of space, we will merely give an indication of the type of system we have in mind, leaving its full exploration for future work. In doing so, we borrow terminology from Makinson's [29].[15]

[13] See e.g. [15] for an introduction to these systems.

[14] The same applies to Goble's work, as he later acknowledged [17; 16]. We believe that a similar argument can be made for the approach of Carmo and Jones, but this is beyond the scope of the current paper.

[15] For readers familiar with *Adaptive Logics* it should be noted that our proposal here can be readily translated into that framework as well. This has the immediate advantage that one obtains

The first weakness of **HD** concerns the inference from $\vec{\Box}\varphi$ to $\overleftrightarrow{\Box}\varphi$. The obvious solution would be: treat all claims of the type "if φ is a fixed circumstance, then φ is all that is fixed" as default assumptions. So our default assumptions are all members of the following set:

$$\Delta_1 = \{\vec{\Box}\varphi \rightarrow \overleftrightarrow{\Box}\varphi \mid \varphi \in \mathcal{W}\}$$

How can we define a new consequence relation $\vdash$, using **HD** and Δ_1? As a first stab, let $\Gamma \vdash \psi$ iff there are $\tau_1, \ldots, \tau_n \in \Delta_1$ such that (i) $\Gamma \cup \{\tau_1, \ldots, \tau_n\} \vdash_{\mathbf{HD}} \psi$, and (ii) $\Gamma \nvdash_{\mathbf{HD}} \neg\tau_i$ for all $i \in \{1, \ldots, n\}$. So e.g. from $\Gamma_1 = \{\vec{\Box}p\}$ we can infer $\overleftrightarrow{\Box}p$; however, from $\Gamma_1' = \{\vec{\Box}p, \vec{\Box}q, \neg\mathsf{U}(p \rightarrow q)\}$ we can no longer infer $\overleftrightarrow{\Box}p$ in view of the derived theorem (AO4). This way, we obtain a non-monotonic, but exact and formal criterion for when it is safe to assume $\overleftrightarrow{\Box}\varphi$ for some φ. One may think of criterion (i) as giving us more inferential power, whereas criterion (ii) makes sure that, whenever the premises require this, inferences are blocked to maintain consistency.

There are two problems with such an approach. The first is well-known from the general study of non-monotonic logic. That is, sometimes criterion (ii) above applies, but there is nevertheless a *disjunction* $\neg\tau_{i_1} \vee \ldots \vee \neg\tau_{i_k}$ that follows from Γ by means of **HD**. In that case, Γ will yield consequences that are jointly incompatible with Γ. Moreover, its consequences will not be closed under classical logic.

Various solutions to this first problem have been developed. One is to quantify over maximal sets of default assumptions compatible with Γ; another is to rephrase (ii) in terms of minimal disjunctions of negations of default assumptions that follow from Γ. The interested reader is kindly referred to [29] where exact definitions of these two solutions and the properties of the resulting logics are discussed.

The second problem is more specific to the current application. Let $\Gamma_2 = \{\vec{\Box}p, \vec{\Box}q\}$. Intuitively, one would expect that only $\overleftrightarrow{\Box}(p \wedge q)$ is derivable from Γ_2. However, this premise set is also compatible with the formula $\mathsf{U}(p \leftrightarrow q)$. So, as far as Γ_2 is concerned, there is no reason to block the inferences from Γ_2 to $\overleftrightarrow{\Box}p$ and $\overleftrightarrow{\Box}q$.

It seems that in order to stay closer to our logical intuitions, a different type of default assumptions should be maximized, *prior to* the assumptions in Δ_1:

$$\Delta_2 = \{\neg\mathsf{U}(\varphi \leftrightarrow \psi) \mid \varphi, \psi \in \mathcal{W}\}$$

In other words: two propositions are taken to be non-equivalent by default, i.e., unless the premises indicate otherwise.[16] Returning to our example $\Gamma_2 = \{\vec{\Box}p, \vec{\Box}q\}$,

a model-theoretic semantics and a dynamic proof theory in the sense of [2] for the resulting logics. See [37] for a study of the relation between Makinson's default assumption consequence relations and adaptive logics.

[16] Note that every member of Δ_2 is of the form $\neg\mathsf{U}\tau$, and conversely, every formula of the latter

we will thus first infer $\neg \mathsf{U}(p \leftrightarrow q)$, $\neg \mathsf{U}(p \leftrightarrow (p \wedge q))$ and $\neg \mathsf{U}(q \leftrightarrow (p \wedge q))$. This at once blocks the derivations of $\overleftrightarrow{\Box} p$ and $\overleftrightarrow{\Box} q$ in view of (AO4).

Note that in the previous paragraph, we emphasized that Δ_2 should receive priority over Δ_1. Again, there are well-studied and well-behaved formal accounts of how to impose a priority structure on default assumptions — see e.g. [37] for a framework that accommodates such refinements.

The second weakness of **HD** that we spotted was that it invalidates (SA), making it impossible to apply general conditional oughts to specific circumstances. This suggests that we use a third type of default assumptions:

$$\Delta_3 = \{\mathsf{O}(\varphi|\psi) \rightarrow \mathsf{O}(\varphi|\psi \wedge \tau) \mid \varphi, \psi, \tau \in \mathcal{W}\}$$

So, for instance, from $\Gamma_3 = \{\mathsf{O}(q|p), \overleftrightarrow{\Box}(p \wedge r)\}$ we first derive $\mathsf{O}(q|p \wedge r)$, after which we apply detachment to derive $\mathsf{O}q$. This inference is blocked in the case of $\Gamma_3' = \{\mathsf{O}(q|p), \mathsf{O}(\neg q|p \wedge r), \overleftrightarrow{\Box}(p \wedge r)\}$, since $\mathsf{O}(q|p \wedge r)$ is incompatible with the second premise. From Γ_3', one can derive $\mathsf{O}\neg q$ by our axiom (DET).[17]

At this point some readers may become suspicious about the whole enterprise of holistic detachment. We argued that one should strengthen $\overrightarrow{\Box}\varphi$ to $\overleftrightarrow{\Box}\varphi$ whenever this is possible; this inference is necessary in order to obtain the kind of information that is strong enough to license detachment. We also argued that one should have a defeasible form of (SA) in order to allow that general conditionals are applicable in more specific cases. But why then not give up on the requirement of holism, so that (SA) is not required in the first place? Doesn't that make for a much smoother logic?

Two points in defense. First, some defeasible form of (SA) is highly intuitive in itself. Regardless of the specific circumstances we are in, it seems that we can reason about the relation between conditional oughts, even if these are interpreted as defeasible pieces of advice. From "if you are in Sapporro, you should go to a sushi-bar", we are inclined to infer "if you are in Sapporro with friends, then you should go to a sushi-bar". The inference appears to be valid, regardless of where in the world one happens to be.

Second, in cases where exceptions are explicitly mentioned – such as, "if you are in Sapporro but you are allergic to fish, then you should not go to a sushi-bar" – we

form can be equivalently rephrased as a formula in Δ_2 (simply by putting $\varphi = \tau$ and $\psi = \top$). Treating formulas of the form $\neg \mathsf{U}\tau$ as default assumptions gives rise to a logic similar to that studied in [3].

[17] Although this paragraph may suggest the opposite, implementing this idea to obtain a formal, well-behaved logic does bring some complications. More specifically, one needs to restrict the logical form of the assumptions from Δ_3 in various ways, in order to overcome so-called *flip-flop problems*. As above, we leave the technical details for a follow-up paper.

do want to be able to draw the correct conclusion regarding our obligations, relative to the circumstances at hand. If I am in Saporro and I happen to be allergic to fish, then I do not want to derive the conclusion that I should go to a sushi-bar. But if one skips the holistic requirement, that conclusion will have exactly the same logical status as the conclusion that I should not go to a sushi-bar: it is a deductive consequence of the premises.

4 Related work

The literature on dyadic deontic logic and detachment is vast. For reasons of space, we focus on work that is directly linked to ours and draw some high-level comparisons. A full study of these relationships is left for future work. We first focus on traditional, possible worlds semantics, after which we consider norm-based accounts (in the sense of [20]) and other more syntactic approaches.

4.1 Possible worlds semantics that validate restricted detachment

As noted in the introduction, the idea of restricted detachment can be found in many accounts of contrary-to-duty paradoxes. Already in his [21], Hansson distinguishes between mere facts and unalterable ones in relation to detachment [21, p. 394]. Greenspan [19] seems to be the first to formalize fixed circumstances by means of a normal modal operator akin to our $\vec{\Box}$. In her account, those circumstances are tied to temporal reasons: e.g. once the time to leave has passed, it is a fixed circumstance that you will not help. However, in [34] it is shown that there are examples of detachment where the circumstances are not fixed due to temporal (or agential) reasons.

In his [13], Feldman argues in favour of restricted detachment and proposes a logic based on this idea. In a back and forth ([10; 14; 9]) Castañeda criticises Feldman's logic, but acknowledges the importance of what he calls "deontic circumstances" for a proper understanding of ought-claims. Such circumstances are contrasted with "deontic foci", i.e. the things that are the subject of obligations and permissions. Our distinction between mere facts and fixed circumstances has some parallels with Castañeda's distinction between deontic circumstances and deontic foci, but there are also some fundamental differences. Most strikingly, Castañeda argues that only *actions* (of a given agent) can be obligatory, not propositions.[18] Circumstances are represented by propositional variables, whereas formulas of the form $\mathsf{O}\varphi$ are only well-formed if φ represents an action. Castañeda deems it unac-

[18] Castañeda uses the term "prescriptions" to refer to the former.

ceptable that deontic circumstances are obligatory (as they are in our account of O, cf. our axiom (BO)), especially for cases of "determined, successfully and carefully planned wrongdoing" [10, p. 13]. Treating the defined operator O_a (cf. Section 2.1) as the "proper" operator for obligations is also unacceptable for Castañeda, since this would still imply that these instances of planned wrongdoing are neither wrong nor right [10, p. 13]. Note however that, as we argued in Section 2.1, we make a distinction between the claim that φ is obligatory given the circumstances – which one can express either by O or O_a – and the claim that the circumstances themselves violate a general obligation – e.g. expressed by $\mathsf{O}(\neg\varphi|\top) \wedge \vec{\Box}\varphi$.

In a series of articles ([6; 7; 8]) Carmo and Jones develop a theory of "fixedness" or unalterability with regards to conditional obligations. This theory was a direct source of inspiration for our **HD**. There are however a few problems with the specific route taken by Carmo and Jones. First, as shown in [27], their logic (as defined in [8]) validates a specific type of deontic explosion: if φ and ψ are independent, in the sense that neither strictly implies the other, and if $\mathsf{O}(\varphi|\top)$, then $\mathsf{O}(\psi|\neg\varphi)$. This is i.a. due to the validity of a (restricted) form of strengthening of the antecedent in their logic.[19] Second, as noted in Section 3.3, one needs to add various tacit premises in order to obtain the correct results with Carmo and Jones' system, even in simple cases. As we argued, this drawback can be overcome by extending the logic non-monotonically. The third and more serious drawback is that this logic cannot handle specificity-cases [7, p. 295] (see also [35]).

According to Van Benthem, Liu and Grossi [36], detachment should be modeled by a formula of the type $\mathsf{O}(\varphi|\psi) \rightarrow [!\psi]\mathsf{O}(\varphi|\top)$, which expresses that if φ is obligatory conditional on ψ, then, after the information is received that ψ, φ becomes an unconditional obligation. Note that here, the holism is built in automatically, since the event $!\psi$ is the announcement of ψ and *nothing but* ψ. However, as of yet, the Hansson-style semantics for dyadic deontic logic has not been equipped with an "all and only" modality, in order to model the (defeasible) inferences from a (possibly incomplete) description of the fixed circumstances to the oughts that can be detached in view of all that is fixed.

In the related field of practical, goal-oriented reasoning, Boutilier [5] proposes a way to determine "actual preferences" on the basis of conditional preference statements and a (finite) knowledge base $\mathcal{K}$. When $I(\psi|\varphi)$ represents that, conditional on φ, ψ is true in all preferred states, the agent has an actual preference for ψ iff $I(\psi|\bigwedge\mathcal{K})$ holds. Note however that this type of inference is not modeled at the object level of the logic: there is no operator for "all the agent knows", or for "what

[19] In fact, one can prove an even stronger fact: in the logic of Carmo and Jones, $\mathsf{O}(\varphi|\top), \neg\mathsf{U}\neg(\neg\varphi\wedge\psi) \vdash \mathsf{O}(\psi|\top)$. So whenever ψ is compatible with the violation of a general obligation that φ, then ψ is obligatory conditional on $\neg\varphi$.

is preferred given all the agent's knowledge". Boutilier's article is mainly focussed on the logic of the conditional preference operator.

Van der Torre [38] suggested that Boutilier's proposal could be turned into a principle of deontic reasoning, using Levesque's "all and only" modality. Van der Torre calls the resulting rule "exact factual detachment" (EFD); it is formally identical to our holistic detachment. He only discusses this rule, but develops no semantics or axiomatization for the resulting logic. He then goes on to argue that EFD leads to the validity of the truth schema $\mathsf{O}\varphi \rightarrow \varphi$ and concludes that "if EFD is accepted, then the relation between facts and absolute obligations is identical to the relation between antecedent and consequent of the conditional obligations" [38, p. 88]. This is however incorrect, at least as long as one can distinguish between mere facts and knowledge (or in our interpretation: what is fixed). Indeed, as we argued in Section 2.2, $\mathsf{O}\varphi \wedge \neg\varphi$ is perfectly satisfiable, even in a logic that validates holistic detachment.

4.2 Norm-based, syntactic accounts

There is also a wide range of accounts that do not attempt to reduce the truth or applicability of conditional norms to some external reality such as a preference relation or deontic function defined over possible worlds. Instead, these accounts take a set N of conditional norms as primitive, and use that N to construct an operational, rule-based semantics for deontic logic. Technically, N is just a set of pairs (ψ, φ) in a formal language.

One account of this type is developed in Horty's [24]. Here, norms are represented as default rules, and facts are conceived as the triggers of unconditional oughts. Conditional oughts $\mathsf{O}(\varphi|\psi)$ are interpreted in terms of what follows from a given default theory when adding ψ to the factual information. In Horty's framework, detachment is treated as a defeasible principle: one applies it as long as the result remains consistent with the given facts; moreover, no distinction is made between mere facts and what is considered fixed.

Input/output logic [30; 31] is another well-known class of systems in which conditionals (as syntactic entities) are primitive. Here again, one can define various input/output relations that, when given a set of input F and a set of conditionals N, fix an output $out(F, N)$, corresponding to our situational obligations. The input roughly plays the role of our fixed circumstances, with the difference that the input is not necessarily included in the output (contrary to our axiom (BO)). In contrast to Horty's account, input/output logic (at least in its original form [30; 31]) cannot handle specificity-cases.

In [33], Parent and van der Torre discuss exact factual detachment (cf. supra) as a property of I/O-logics. The idea here is that, if $(\varphi, \psi) \in N$, then ψ is in the output

of N under the input φ. In other words, if one has *exactly* the input φ, and if there is a conditional norm to the effect that ψ is obligatory if φ is the case, then ψ is indeed obligatory. In the original I/O-logics, and in the specific system $\mathcal{O}_3$ proposed by Parent and van der Torre in [33], exact factual detachment is a derived property; also there, specificity-cases cannot be handled in the way **HD** does.

Straßer [35] develops a non-monotonic logic which features specific expressions of the type "the obligation to do φ, conditional on ψ, is overridden", denoted by $\bullet\mathsf{O}(\varphi|\psi)$.[20] The latter phrase receives a purely syntactic definition, in terms of the existence of other obligations and circumstances. In these systems, detachment is represented by a rule of the following type:

$$\psi, \mathsf{O}(\varphi|\psi), \neg\bullet\mathsf{O}(\varphi|\psi) \vdash \mathsf{O}\varphi$$

By adding a suitable set of axioms that govern the behavior of $\bullet\mathsf{O}(\varphi|\psi)$, and by adding a defeasible mechanism that validates the inference from $\mathsf{O}(\varphi|\psi)$ to $\neg\bullet\mathsf{O}(\varphi|\psi)$, one can then ensure that specificity-cases and contrary-to-duties are adequately handled. In this framework, no operators for "fixed circumstances" are used; also, no semantics for the $\bullet$-operator is provided.

In more recent work, Beirlaen and Straßer [4] have used structured argumentation frameworks to model deontic reasoning on the basis of conditional obligations and "fixed facts", making use of a normal modal operator to model the latter. Here, one looks at all possible arguments that can be built for a given claim (e.g. $\mathsf{O}p$) and defines various attack relations between such arguments. These attack relations in turn allow one to determine a "grounded extension" of the premises, which can be seen as a set of "safe arguments" that in turn deliver the situational obligations. As Beirlaen and Straßer show, specificity and other parameters can be readily built into this framework by adopting an appropriate definition of the attack relation.

The mentioned accounts are arguably richer and more fine-grained than our **HD**. Still, one argument in favour of more traditional approaches (such as our own), and contra norm-based or other purely syntactic approaches such as the ones mentioned above, is that we get one unified theory in which we can not only reason about what should hold given some F and N. That is, using principles such as our (ATT), we can also derive general rules $\mathsf{O}(\varphi|\psi)$ from premises that merely state what is fixed, and what we consider the "correct" normative judgement in that situation. It should however be admitted that from this viewpoint, **HD** is just a preliminary "toy logic". A more elaborate, insightful semantics and formal language should be developed in order to flesh out and solidify this argument.

[20] In fact, Straßer distinguishes two ways in which an obligation can be overridden, resulting in two operators $\bullet_{\mathsf{p}}$ and $\bullet_{\mathsf{i}}$, and two different rules of detachment.

5 Concluding remarks

What we hope to have shown is that one can model holistic detachment in deontic reasoning in a monotonic, specificity-sensitive way, making use of an "all and only"-operator. We do not reject non-monotonicity altogether, but move it more upwards in the inference chain that leads to situational obligations. In a slogan: not the rule of detachment, but its premises are defeasible.

A lot of work remains to be done in order to develop **HD** and its nonmonotonic extensions into a full-fledged, formal theory. Let us just mention three general lines of research. First, more refined accounts of holistic detachment should be studied. One may for instance require that all and only those fixed circumstances that are *relevant* should be taken into account. To explicate such a principle formally, it may be useful to apply techniques from the study of relevant or hyperintensional logics. Second, we hope to be able to work out richer, more insightful semantics for our deontic conditionals, following the traditional accounts cited in Section 4. Third, as noted in the Introduction, we believe there are distinct types of conditional norms, ranging from mere (defeasible) advice to strict legal stipulations. Logically speaking, such conditionals display very different behavior; a general theory of dyadic deontic logic should take this into account.

References

[1] Lennart Åqvist. *Deontic Logic*, volume 8 of *Handbook of Philosophical Logic*, chapter 4, pages 147–264. Kluwer Academic Publishers, 2nd edition, 2002.

[2] Diderik Batens. A universal logic approach to adaptive logics. *Logica Universalis*, 1:221–242, 2007.

[3] Diderik Batens and Joke Meheus. The adaptive logic of compatibility. *Studia Logica*, 66(3):327–348, 2000.

[4] Mathieu Beirlaen and Christian Straßer. A structured argumentation framework for detaching obligations. In Olivier Roy, Allard Tamminga, and Malte Willer, editors, *Deontic Logic and Normative Systems*, pages 32–48. College Publications, 2016.

[5] Craig Boutilier. Toward a logic for qualitative decision theory. In *Proceedings of the Fourth International Conference on Principles of Knowledge Representation and Reasoning*, KR'94, pages 75–86. Morgan Kaufmann Publishers Inc., 1994.

[6] José M. C. L. M. Carmo and Andrew J. I. Jones. A new approach to contrary-to-duty obligations. In D. Nute, editor, *Defeasible Deontic Logic*, pages 317–344. Synthese Library, 1997.

[7] José M. C. L. M. Carmo and Andrew J. I. Jones. Deontic logic and contrary-to-duties. In Dov M. Gabay and Franz Guenther, editors, *Handbook of Philosophical Logic*, volume 8, pages 147–264. Kluwer Academic Publishers, 2nd edition, 2002.

[8] José M. C. L. M. Carmo and Andrew J. I. Jones. Completeness and decidability results for a logic of contrary-to-duty conditionals. *Journal of Logic and Computation*, 23(3):585, 2013.

[9] Hector-Neri Castañeda. Moral obligations, circumstances, and deontic foci (a rejoinder to Fred Feldman). *Philosophical Studies*, 57(2):157–174, 1989.

[10] Hector-Neri Castañeda. Paradoxes of moral reparation: Deontic foci vs. circumstances. *Philoshophical Studies*, 57(1):1–21, 1989.

[11] Brian F. Chellas. *Modal Logic: An Introduction*. Cambridge University Press, 1980.

[12] Roberto Ciuni. Conditional doxastic logic with oughts and concurrent upgrades. In Alexandru Baltag, Jeremy Seligman, and Tomoyuki Yamada, editors, *Logic, Rationality, and Interaction*, pages 299–313. Springer, 2017.

[13] Fred Feldman. *Doing the Best We Can: An Essay in Informal Deontic Logic*. Philosophical Studies Series in Philosophy 35. Springer Netherlands, 1986.

[14] Fred Feldman. Concerning the paradox of moral reparation and other matters. *Philosophical Studies*, 57(1):23–39, 1989.

[15] Lou Goble. A logic for deontic dilemmas. *Journal of Applied Logic*, 3:461–483, 2005.

[16] Lou Goble. Prima facie norms, normative conflicts, and dilemmas. In Dov Gabbay, John Horty, Xavier Parent, Ron van der Meyden, and Leendert van der Torre, editors, *Handbook of Deontic Logic and Normative Systems*, volume 1, pages 241–351. College Publications, 2013.

[17] Lou Goble. Deontic logic (adapted) for normative conflicts. *Logic Journal of the IGPL*, 22(2):206–235, 2014.

[18] Valentin Goranko and Solomon Passy. Using the universal modality: Gains and questions. *Journal of Logic and Computation*, 2(1):5–30, 1992.

[19] P. S. Greenspan. Conditional oughts and hypothetical imperatives. *The Journal of Philosophy*, 72(10):259–276, 1975.

[20] Jörg Hansen. Reasoning about permission and obligation. In Sven Ove Hansson, editor, *David Makinson on Classical Methods for Non-Classical Problems*, pages 287–333. Springer Netherlands, 2014.

[21] Bengt Hansson. An analysis of some deontic logics. *Noûs*, 3(4):373–398, 1969.

[22] Sven Ove Hansson. Situationist deontic logic. *Journal of Philosophical Logic*, 26(4):423–448, 1997.

[23] Risto Hilpinen and Paul McNamara. Deontic logic: A historical survey and introduction. In Dov Gabbay, John Horty, Xavier Parent, Ron van der Meyden, and Leendert van der Torre, editors, *Handbook of Deontic Logic and Normative Systems*, volume 1, pages 3–136. College Publications, 2013.

[24] John F. Horty. *Reasons as Defaults*. Oxford University Press, 2012.

[25] I. L. Humberstone. The modal logic of "all and only". *Notre Dame J. Formal Logic*, 28(2):177–188, 1987.

[26] Albert Jonsen. Casuistry: An alternative or complement to principles. *Kennedy Institute of Ethics Journal*, 5(3):245, 1995.

[27] Bjørn Kjos-Hanssen. A conflict between some semantic conditions of Carmo and Jones for contrary-to-duty obligations. *Studia Logica*, (1):173–178, 2017.

[28] Hector J. Levesque. All I know: A study in autoepistemic logic. *Artificial Intelligence*, 42(2-3):263–309, 1990.

[29] David Makinson. *Bridges from Classical to Nonmonotonic Logic*, volume 5 of *Texts in Computing*. King's College Publications, London, 2005.

[30] David Makinson and Leon van der Torre. Input/output logics. *Journal of Philosophical Logic*, 29:383–408, 2000.

[31] David Makinson and Leon van der Torre. Constraints for input/output logics. *Journal of Philosophical Logic*, 30:155–185, 2001.

[32] Eric Pacuit, Rohit Parikh, and Eva Cogan. The logic of knowledge based obligation. *Synthese*, 149(2):311–341, 2006.

[33] Xavier Parent and Leendert van der Torre. "Sing and dance!". In Fabrizio Cariani, Davide Grossi, Joke Meheus, and Xavier Parent, editors, *Deontic Logic and Normative Systems*, pages 149–165. Springer, 2014.

[34] Henry Prakken and Marek Sergot. Contrary-to-duty obligations. *Studia Logica*, 57(1):91–115, 1996.

[35] Christian Straßer. A deontic logic framework allowing for factual detachment. *Journal of Applied Logic*, 9:61–80, 2011.

[36] Johan van Benthem, Davide Grossi, and Fenrong Liu. Priority structures in deontic logic. *Theoria*, 80(2):116–152, 2014.

[37] Frederik Van De Putte. Default assumptions and selection functions: A generic framework for non-monotonic logics. In Felix Castro, Alexander Gelbukh, and Miguel Gonzalez, editors, *MICAI*, volume 8265 of *Lecture Notes in Computer Science*, pages 54–67. Springer, 2013.

[38] Leendert van der Torre. *Reasoning About Obligations*. PhD thesis, Erasmus Universiteit Rotterdam, 1997.

APPENDIX

A Soundness and strong completeness of HD

For the proof of completeness, we need to refine the usual canonical model construction in various ways. First, to deal with the universal modality U, we need to relativize the canonical model to a given maximal consistent set Γ; the idea is that we only take worlds that are equivalent to Γ for all U-formulas. Second, we need to make two copies of each maximal consistent set; this allows us to deal with cases where $R(\Delta)$ (defined in the standard way for the canonical model, given some maximal consistent set Δ) happens to be modally definable by some formula φ, but nevertheless $\overleftrightarrow{\Box}\varphi \notin \Delta$. Third and last, we need to define the function f in such a way that its interaction with R and the semantic clauses of $\overleftrightarrow{\Box}$, $\mathsf{O}(.|.)$, and O are respected.

Definition 3. *Let Γ be a maximal consistent set in* **HD**. *Let MCS_Γ denote the set of all maximal consistent sets Δ in* **HD** *such that $\{\varphi \mid \mathsf{U}\varphi \in \Delta\} = \{\psi \mid \mathsf{U}\psi \in \Gamma\}$. We define the canonical model $M_\Gamma = \langle W_\Gamma, R_\Gamma, f_\Gamma, V_\Gamma\rangle$ for* **HD** *as follows:*[21]

(i) $W_\Gamma = \{(\Theta, i) \mid \Theta \in \mathsf{MCS}_\Gamma, i \in \{1, 2\}\}$

(ii) *For all $(\Delta, i) \in W_\Gamma$:*

- *(ii.1)* *if there is a φ s.t. $\overleftrightarrow{\Box}\varphi \in \Delta$, $R_\Gamma(\Delta, i) = \{(\Theta, j) \in W_\Gamma \mid \varphi \in \Theta\}$;*
- *(ii.2)* *otherwise: $R_\Gamma(\Delta, i) = \{(\Theta, i) \mid \{\psi \mid \overrightarrow{\Box}\psi \in \Delta\} \subseteq \Theta\}$*

(iii) *For all $(\Delta, i) \in W_\Gamma$, $X \subseteq W_\Gamma$:*

- *(iii.1)* *if there is a ψ s.t. $X = \{(\Theta, j) \in W_\Gamma \mid \psi \in \Theta\}$:*
  - *(iii.1a)* *if $X = R_\Gamma(\Delta, i)$, $f_\Gamma((\Delta, i), X) = \{(\Lambda, i) \in W_\Gamma \mid \{\tau \mid \mathsf{O}\tau \in \Delta\} \subseteq \Lambda\}$*
  - *(iii.1b)* *otherwise, $f_\Gamma((\Delta, i), X) = \{(\Lambda, k) \in W_\Gamma \mid \{\tau \mid \mathsf{O}(\tau \mid \psi) \in \Delta\} \subseteq \Lambda\}$*
- *(iii.2)* *if there is no ψ s.t. $X = \{(\Theta, j) \in W_\Gamma \mid \psi \in \Theta\}$:*
  - *(iii.2a)* *if $X = R_\Gamma(\Delta, i)$, $f_\Gamma((\Delta, i), X) = \{(\Lambda, i) \mid \{\tau \mid \mathsf{O}\tau \in \Delta\} \subseteq \Lambda\}$*
  - *(iii.2b)* *otherwise, $f_\Gamma((\Delta, i), X) = X$*

(iv) *For all $\varphi \in \mathcal{S}$: $V_\Gamma(\varphi) = \{(\Theta, k) \in W_\Gamma \mid \varphi \in \Theta\}$*

Lemma 1 (Strict implication)**.** *For all φ and all $(\Delta, i) \in W_\Gamma$: $\mathsf{U}(\varphi \to \psi) \in \Delta$ iff $\{(\Theta, k) \in W_\Gamma \mid \varphi \in \Theta\} \subseteq \{(\Lambda, j) \in W_\Gamma \mid \psi \in \Lambda\}$.*

[21] We often skip brackets when referring to members of W_Γ; e.g. $R_\Gamma((\Theta, i))$ is written as $R_\Gamma(\Theta, i)$.

Proof. The proof is standard, relying on **S5**-properties of U and the construction of W_Γ. ■

Corollary 1 (Strict Equivalence)**.** *For all φ and all $(\Delta, i) \in W_\Gamma$: $\mathsf{U}(\varphi \leftrightarrow \psi) \in \Delta$ iff $\{(\Theta, k) \in W_\Gamma \mid \varphi \in \Theta\} = \{(\Lambda, j) \in W_\Gamma \mid \psi \in \Lambda\}$.*

Lemma 2. *For all maximally consistent sets Γ, M_Γ is an* **HD***-model.*

Proof. Since U is an **S5**-modality, we can easily show that $W_\Gamma \neq \emptyset$. Using Corollary 1 and the derived theorem (AO4), we can prove that R_Γ is well-defined. Also, with Corollary 1 and the axiom (CG) we can show that f_Γ is well-defined. So it suffices to check that each of the conditions (C1)-(C4) are satisfied. (C1) and (C4) are safely left to the reader. For (C2) we rely on the construction, the derivable theorem $\overleftrightarrow{\Box}\varphi \rightarrow \varphi$ (in case (ii.1) applies) and the **T**-schema for $\vec{\Box}$ (if (ii.2) applies). So we are left with establishing (C3).

Ad (C3.1). Let $(\Delta, i) \in W_\Gamma$ and suppose that $\emptyset \neq X \subseteq W_\Gamma$. If (iii.1a) or (iii.2a) applies, then we rely on the **KD**-properties of O to infer that $\{(\Theta, i) \mid \{\tau \mid \mathsf{O}\tau \in \Delta\} \subseteq \Theta\} \neq \emptyset$ and hence $f_\Gamma((\Delta, i), X) \neq \emptyset$. If (iii.1b) applies, then in view of the construction and since $X \neq \emptyset$, there exists a ψ such that $\neg\mathsf{U}\neg\psi \in \Delta$. We can rely on compactness, the **K**-axioms for $\mathsf{O}(.|\psi)$, and axiom (CP), to show that $\{(\Lambda, k) \in W_\Gamma \mid \{\tau \mid \mathsf{O}(\tau \mid \psi) \in \Delta\} \subseteq \Lambda\} \neq \emptyset$. Finally, if (iii.2b) applies then by the construction, $f((\Delta, i), X) = X \neq \emptyset$.

Ad (C3.2). This follows by the construction. For (iii.1a) and (iii.2a), apply axiom (BO) to derive that $\{(\Theta, i) \mid \{\tau \mid \mathsf{O}\tau \in \Delta\} \subseteq \Theta\} \subseteq R_\Gamma(\Delta, i)$. For (iii.1b), apply axiom (CI). For (iii.2b) the conclusion follows trivially. ■

Lemma 3 (Truth Lemma)**.** *For all φ and all $(\Delta, i) \in W_\Gamma$: $M_\Gamma, (\Delta, i) \models \varphi$ iff $\varphi \in \Delta$.*

Proof. By a standard induction on the complexity of φ. The base case is trivial. The case for $\varphi = \mathsf{U}\psi$ is likewise standard, relying on the construction and **S5**-properties of U.

For $\varphi = \vec{\Box}\psi$, we distinguish between two cases. If (ii.1) applies, then $M_\Gamma, (\Delta, i) \models \vec{\Box}\psi$ iff [by the construction] for some τ such that $\overleftrightarrow{\Box}\tau \in \Delta$, $\{(\Theta, j) \in W_\Gamma \mid \tau \in \Theta\} \subseteq \{(\Theta, j) \in W_\Gamma \mid \psi \in \Theta\}$ iff [by Lemma 1] for some τ such that $\overleftrightarrow{\Box}\tau \in \Delta$, $\mathsf{U}(\tau \rightarrow \psi) \in \Delta$ iff [by axioms (AO1) and (UB) for left to right and by (AO2) for right to left] for some τ such that $\overleftrightarrow{\Box}\tau \in \Delta$, $\vec{\Box}\psi \in \Delta$. If (ii.2) applies, then we can use the standard argument, relying on the normality of $\vec{\Box}$, to prove the truth lemma for this case.

For $\varphi = \overleftrightarrow{\Box}\psi$, the right to left direction is easy in view of the construction, item (ii.1). For the other direction, we distinguish two cases:

Case 1: The truth set of ψ in M_Γ equals W_Γ. This means that $R_\Gamma(\Delta, i) = W_\Gamma$. In view of the construction, we cannot be in case (ii.2) and hence $\overleftrightarrow{\Box}\top \in \Delta$; applying (AO3) and Corollary 1, we obtain that $\overleftrightarrow{\Box}\psi \in \Delta$.

Case 2: ψ is not a tautology. Then, if (ii.1) applies, we can infer that ψ has the same truth set as some ψ' with $\overleftrightarrow{\Box}\psi' \in \Delta$, and hence $\overleftrightarrow{\Box}\psi \in \Delta$ by Corollary 1 and (AO3). In the other case (i.e. when (ii.2) applies), we derive a contradiction from the fact that $R_\Gamma(\Delta, i)$ is not definable by any formula τ in the model M_Γ.

For $\varphi = \mathsf{O}(\psi|\tau)$, right to left is again easy in view of case (iii.1) of the construction. The only tricky case is where (iii.1a) applies. There, we apply axiom (ATT) to derive that $\mathsf{O}(\psi|\tau) \in \Delta$.

For the other direction, suppose that $M_\Gamma, (\Delta, i) \models \mathsf{O}(\psi|\tau)$. If (iii.1a) applies, then we can infer that $\overleftrightarrow{\Box}\tau \in \Delta$ (by case $\varphi = \overleftrightarrow{\Box}\psi$) and there is some $\psi' \in \Delta$ such that $\mathsf{O}\psi'$. By axiom (ATT), $\mathsf{O}(\psi'|\tau) \in \Delta$. In view of the construction $\{(\Theta, i) \mid \psi' \in \Theta\} \subseteq \{(\Lambda, j) \mid \psi \in \Lambda\}$. Hence, $\psi' \vdash \psi$ and hence by the **K**-properties of $\mathsf{O}(.|\tau)$, $\mathsf{O}(\psi|\tau) \in \Delta$.

If (iii.1b) applies, then by the induction hypothesis, we can infer that

$$f_\Gamma((\Delta, i), \|\psi\|^{M_\Gamma}) = \{(\Lambda, k) \mid k \in \{1, 2\}, \{\tau' \mid \mathsf{O}(\tau' \mid \psi) \in \Delta\} \subseteq \Lambda\}$$

Then, applying the **K**-properties of $\mathsf{O}(.|\psi)$ and compactness, we can infer that $\mathsf{O}(\tau, \psi) \in \Delta$.

Finally, the case for $\varphi = \mathsf{O}\psi$ is completely standard; it suffices to note that, given the semantic clause for O, the only relevant items in the construction are (iii.1a) and (iii.2a), where in both cases, we can use a standard argument (relying on **K**-properties of O and the induction hypothesis) to prove that

$$\mathsf{O}\psi \in \Delta \text{ iff } f_\Gamma((\Delta, i), R_\Gamma(\Delta, i)) \subseteq \{(\Theta, j) \in W_\Gamma \mid \psi \in \Theta\}$$

■

Theorem 2. $\Gamma \vdash_{\mathbf{HD}} \varphi$ *iff* $\Gamma \models_{\mathbf{HD}} \varphi$.

Proof. Soundness is a matter of routine; it suffices to check that each of the **HD**-axioms are valid. For the completeness proof we rely on the canonical model from Definition 3. Suppose that $\Gamma \nvdash \varphi$. Hence, $\Gamma \cup \{\neg\varphi\}$ is consistent. By Lindenbaum's Lemma we can construct a maximal consistent set $\Theta \supseteq \Gamma \cup \{\neg\varphi\}$. Note that $\varphi \notin \Theta$. By Lemma 2 we know that M_Θ is an **HD**-model. By Lemma 3, $M_\Theta, (\Theta, 1) \models \psi$ for all $\psi \in \Gamma$ but $M_\Theta, (\Theta, 1) \not\models \varphi$. Hence, $\Gamma \not\Vdash \varphi$. ■

B A stronger variant

Theorem 3. *A sound and strongly complete axiomatization for the class of models where R is an equivalence relation is obtained by adding the following axioms to* **HD***:*

(B) $\varphi \rightarrow \vec{\Box}\neg\vec{\Box}\neg\varphi$
(4) $\vec{\Box}\varphi \rightarrow \vec{\Box}\vec{\Box}\varphi$
(S5$\overleftrightarrow{\Box}$-1) $\overleftrightarrow{\Box}\varphi \leftrightarrow \overleftrightarrow{\Box}\overleftrightarrow{\Box}\varphi$
(S5$\overleftrightarrow{\Box}$-2) $\neg\overleftrightarrow{\Box}\varphi \rightarrow \vec{\Box}\neg\overleftrightarrow{\Box}\varphi$

Proof. The proof is wholly analogous to that of Theorem 2, using the same construction for the canonical model. The only place where we need to amend that proof is in showing that the relation R_Γ constructed there is both transitive and symmetric. For both properties, we rely on the above four axioms. More particularly, in case (ii.1) applies we rely on (S5$\overleftrightarrow{\Box}$-1) to prove both transitivity and symmetry. In case (ii.2) applies we rely on (S5$\overleftrightarrow{\Box}$-2) and (4) for transitivity, and on (S5$\overleftrightarrow{\Box}$-2) and (B) for symmetry. ■

Note that from the above proof, we do not immediately get a completeness result for the intermediate logics where R is only symmetric (and reflexive) or only transitive (and reflexive). A full study of these systems is left for future work.

C Proof of Theorem 1

For the sake of readibility, we repeat the theorem:
Theorem 1. *Let $\Gamma \subseteq \mathcal{W}$ be a* **HD***-consistent set such that $\overleftrightarrow{\Box}$ occurs in no member of Γ. Then there is no φ such that $\Gamma \vdash \overleftrightarrow{\Box}\varphi$.*

Proof. Suppose the antecedent is true. Let $M = \langle W, R, f, V\rangle$ be a model of Γ. We construct the model $M' = \langle W', R', f', V'\rangle$ as follows (where $i \in \{1, 2\}$):

(i) $W' = \{(w, 1), (w, 2) \mid w \in W\}$
(ii) $R'(w, i) = \{(v, i) \mid v \in R(w)\}$
(iii) if there is a φ such that $X = \{(v, j) \in W' \mid v \in \|\varphi\|^M\}$, then $f'((w, i), X) = \{(u, j) \in W' \mid u \in f(w, \|\varphi\|^M)\}$
(iv) if $X = R(w, i)$, then $f'((w, i), X) = \{(v, i) \in W' \mid v \in f(w, R(w))\}$
(v) if neither (iii) nor (iv) applies, then $Ff'((w, i), X) = X$.
(vi) $V'(\varphi) = \{(w, 1), (w, 2) \mid w \in V(\varphi)\}$ for all $\varphi \in \mathcal{S}$

One can now show that M' is a **HD**-model and (by an induction on the complexity of φ) that, for all φ that do not contain $\overleftrightarrow{\Box}$, and for all $w \in W$, $M, w \models \varphi$ iff

$(M', (w, 1) \models \varphi$ and $M', (w, 2) \models \varphi)$. However, one can also show that for no $\psi \in \mathcal{W}$ and for no $(w, i) \in W'$, $M', (w, i) \models \overset{\leftrightarrow}{\Box}\psi$, since for no $(w, i) \in W'$, the set $R(w, i)$ is definable. ■

Received \jreceived