Perspectival obligation and extensionality in an alethic-deontic setting

Dominik Pichler and Xavier Parent [1]

Favoritenstraße 9/E192-05 (Stiege 2/3.Stock) A-1040 Wien, Austria

Abstract

The theme of extensionality in first-order deontic logic has been thoroughly studied in the past, but not in the context of a combination of different types of modalities. An operator is extensional if it allows substitution *salva veritate* of co-referential terms within its scope and intensional if it does not. It can be argued that one distinctive feature of "ought" (as opposed to the other modalities) is that it is extensional. The question naturally arises as to whether it is possible to combine extensionality and intensionality of different modal operators in the same semantics without creating the deontic collapse. We answer this question within a particular framework, Åqvist's system **F** for conditional obligation. We develop in full detail a perspectival account of obligation (and related notions), as was done for Standard Deontic Logic (SDL) by Goble. It is called "perspectival", because one always evaluates the content of an obligation in one world from the perspective of another one, hence using some form of cross-world evaluation. The proposed framework allows for a more nuanced way of approaching first-order deontic principles.

Keywords: First-order reasoning, extensionality, conditional obligation, 2-dimensional semantics, preferences, perspectivism

1 Introduction

The past 15 years have seen a renewed interest in so-called relativism or perspectivism in the philosophy of language. Relativist or perspectivist accounts have been put forth to explain discourse about knowledge, epistemic possibility, matters of taste, contingent future events, modalities (including the deontic ones) and the like. Here relativism is usually taken to be, or to presuppose, a semantic thesis. Understanding how some discourses function requires recognizing that speakers express propositions whose truth or falsity are relative to parameters or perspectives in addition to a possible world−see Kölbel [20] for a

[1] Emails: dominik@logic.at; xavier@logic.at. The research of the second author is supported by the Austrian Science Fund (FWF) through the ANCoR project [M-3240-N]. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. This paper was originated by the first author's master thesis [25]. The material was modified, and expanded. We thank three anonymous reviewers for their valuable comments.

thorough defense of this view, and also MacFarlane [22]. The approach is often called "perspectivism" as it has a less negative connotation than "relativism", and we will stick to this term.

The purpose of the present paper is to show some of the usefulness of this view for normative reasoning. We believe it may shed light on a topic that has been overlooked in the recent papers devoted to first-order deontic reasoning, e.g. [7,8,28]. This is the topic of extensionality of "ought". We do not claim to be original, as we will pick up on a proposal made long ago by Goble [12,13,14]. It can be summarized thus. An operator is extensional if it allows substitution *salva veritate* of co-referential terms within its scope, and intensional if it does not. It can be argued that one distinctive feature of "ought" (as opposed to the other modalities) is that it is extensional. The problem is: a deontic logic in which "ought" is extensional can be shown to collapse to triviality. Goble developed his own solution to this problem, and we will refer to it as the original "perspectival" account. The basic idea is that the content of an obligation at one world is to be evaluated from the perspective of another one, so that some form of cross-world evaluation is made possible. This idea of cross-world evaluation is familiar from the literature on multi-dimensional modal logic (see e.g. [3,11,18,29]). Other works in multi-dimensional deontic logic we are aware of focus on the propositional case [6,9,10,17]. The novelty lies in linking the perspectival idea to first-order considerations.

Our goal is to improve the original account in two ways. By doing so, we hope to strengthen the case for the perspectival idea, and provide more credibility to it.

- The original account is cast within the framework of Standard Deontic Logic (SDL) [31], which is known to be plagued by the deontic paradoxes, in particular the paradox of contrary-to-duty (CTD) obligation [4]. We will recast the account within the framework of preference-based dyadic deontic logic [1,5,15,16,23]. Dyadic deontic logic is the logic for reasoning with dyadic obligations "it ought to be the case that ψ if it is the case that φ" (notation: $\bigcirc(\psi/\varphi)$). Its semantics is in terms of a betterness relation. Initially devised to resolve the CTD paradox, dyadic deontic logic is a recognized standard for normative reasoning. The idea of making it two-dimensional is not entirely new: Lewis [21, p. 63] suggested to analyze conditionals within the framework of two-dimensional modal logic, but his motivations were different.
- The original account does not allow for different types of modalities to interact. We will lift this restriction, and look at the question of whether it is possible to combine extensionality and intensionality of different modal operators in the same semantics without creating the collapse. We will use Åqvist's mixed alethic-deontic preference-based logic **F** [1,23,24]. The language of **F** has an extra modal operator $\Box$ ("it is settled that"), allowing to capture some fundamental principles of normative reasoning, like "strong factual detachment" [26]. Among the systems proposed by Åqvist, **F** is also the weakest one in which the collapse arises. The first-order ex-

tension of **F** will be called $\mathbf{F}^{\forall}$. (One could object that, in **F**, $\Box$ is a *soi disant* modality, definable in terms of $\bigcirc(-/-)$. In $\mathbf{F}^{\forall}$, it will become a first-class citizen, viz. a primitive modality.)

The paper is organized as follows. Section 2 sets the stage, and defines a list of basic requirements to be met by the logic. Section 3 develops in full semantic detail the perspectival account of obligation (and related notions) alluded to above. Section 4 shows how the requirements are met. Section 5 concludes.

2 Setting the stage

We give a list of basic requirements that we think an adequate first-order (FO) deontic logic should meet. The problem dealt with in this paper will be to devise a framework meeting them. For ease of readability, we formulate the requirements within the language of a monadic deontic logic. Our list is not meant to be exhaustive.

2.1 Requirements

Requirement 1 (Extentionality for "ought") $\bigcirc$ *("It ought to be the case that ...") should validate the principle of substitution salva veritate (E-$\bigcirc$), where φ is a formula, t and s are terms, and $\varphi_{t\hookrightarrow s}$ is the result of replacing zero up to all occurrences of t, in φ, by s:*

$$t = s \rightarrow (\bigcirc\varphi \leftrightarrow \bigcirc\varphi_{t\hookrightarrow s}) \qquad \text{(E-}\bigcirc\text{)}$$

Intuitively: two co-referential terms may be interchanged without altering the truth-value of the deontic formula in which they occur.

A modal operator is usually said to be referentially transparent, when it satisfies the principle of substitution *salva veritate*, and referentially opaque otherwise. As pointed out by Castañeda [2] there are good reasons to believe that deontic operators are referentially transparent. For instance, the inference from (1) and (2) to (3) is intuitively valid:

(1) The Pope ought to live a life of exceptional sanctity: $\bigcirc S(\imath x Po(x))$

(2) Jose is the Pope: $j = \imath x Po(x)$

(3) Jose ought to live a life of exceptional sanctity: $\bigcirc S(j)$

$\imath x Po(x)$ is a so-called definite description, and is read "the x that is Po" ("the Pope"). Definite descriptions are used to refer to what a speaker wishes to talk about. Castañeda (rightly) says: "a man's obligations are *his* [the author's emphasis] regardless of his characterizations". In other worlds, they are independent of the way he is referred to.
In daily conversations, one casually switches between a proper name and the definite description associated with it. When using one instead of the other, we are still talking about the same individual, referring to him using different descriptions (the Pope, the direct successor of St Peter, ...). This would just not be possible if "ought" was not referentially transparent.
However, it may be questioned whether the inference from (1) and (2) to (3)

is valid intuitively.[2] For one can consistently add to the premises set

(2′) Jose ought not to be the Pope: $\bigcirc(j \neq \imath x Po(x))$

Does (3) still follow? It seems not. Two comments are in order. First, it may be thought that a finer-grained version of the principle is needed. To add

(2″) Jose ought to be the Pope: $\bigcirc(j = \imath x Po(x))$

would resolve the problem, but would make (2) superfluous. For (3) follows from (1) and (2″) using the standard principles of deontic logic and first-order logic. To add

(2⋆) Jose may be the Pope: $P(j = \imath x Po(x))$–P: "it is permitted that"

would resolve the problem, and not make (2) superfluous. Thus, one way to address the above problem is to introduce the following permitted version of (E-$\bigcirc$):

$$t = s \wedge P(t = s) \rightarrow (\bigcirc\varphi \leftrightarrow \bigcirc\varphi_{t \hookrightarrow s}) \qquad \text{(Permitted E-}\bigcirc\text{)}$$

Second, it may make a difference whether the substitution is done in the consequent or the antecedent of a conditional obligation. Consider:

(4) If the Pope does not live a life of exceptional sanctity, we should elect a new one: $\bigcirc(\exists y(El(y) \wedge y \neq \imath x Po(x))/\neg S(\imath x Po(x)))$

(5) If Jose does not live a life of exceptional sanctity, we should elect a new Pope: $\bigcirc(\exists y(El(y) \wedge y \neq \imath x Po(x))/\neg S(j))$

The antecedent of (4) refers to a sub-ideal world where (1) is violated. Intuitively, (4) and (5) seem equivalent, even in the presence of (2′). Thus, (4) and (5) are two different ways to say the same thing. If "ought" is not referentially transparent, then (4) and (5) are not synonymous, since they have a different antecedent. If so, one would need

- the permitted version of the principle for any substitution done in the consequent (proviso: $P(t = s)$);
- the unrestricted version for any substitution done in an antecedent.

We leave it as a topic of future research to investigate how to implement these suggestions.

For simplicity's sake, $\Box$ will be read as "It is necessary that ...". Whether it is historical necessity or some other type of necessity is not germane for our discussion.

Requirement 2 (Intensionality for "necessarily") $\Box$ *should not validate the principle of substitution salva veritate, where t and s are terms (either a constant or a definite description):*

$$t = s \rightarrow (\Box\varphi \leftrightarrow \Box\varphi_{t \hookrightarrow s}) \qquad \text{(E-}\Box\text{)}$$

[2] We owe this objection from an anonymous referee.

It is usually thought that $\Box$ should not verify (E-$\Box$). The reason why is best illustrated with the following well-known example. Intuitively, (6) and (7) do not imply (8):[3]

(6) Number of planets $= 8$

(7) $\Box(8 = 8)$

(8) $\Box$(Number of planets $= 8$)

If $\Box$ means "settled" in the sense of outside of the agent's control, then (8) is fine. But if $\Box$ means metaphysical necessity, settledness in the sense of historical necessity, or knowledge, then (8) is clearly unwanted. Indeed, before 2006, (6) was false.[4]

A second, independent argument against (E-$\Box$) will be given in Prop. 2.2.

Requirement 3 (No collapse) *The logic should avoid the deontic collapse. That is, the formula* $\varphi \leftrightarrow \bigcirc\varphi$ *should not be derivable.*

This requirement is taken from Goble [12,13,14]. A separate section is devoted to this requirement.

The *raison d'être* of our last requirement is this: obligations are there to make the world a better place; they are constantly violated, but should not be so. Therefore, our account should make the notion of definite description well-behaved with respect to negation. That is to say:

Requirement 4 (Self-negation) *Given E-*$\bigcirc$*, the logic should be able to account for the meaningfulness of a deontic statement denying a property of an individual identified using that very same property.*

Here is an example:

(9) The tyrant has an obligation not to be a tyrant: $\bigcirc\neg T(\imath x T(x))$

Self-negation like the one in (9) cannot be accounted for in (a straightforward FO extension of) SDL. (9) tells us that in the best of all possible worlds the tyrant x is not a tyrant. But this is a contradiction (assuming that such an x exists). Of course, the claim is not that in the best of all possible worlds the tyrant x is not a tyrant. Rather$-$to anticipate our solution$-$the claim is that the individual x that is a tyrant in the actual world is not a tyrant in all the best worlds. This is a relation among objects in possible worlds that cannot be captured in the standard possible world semantics. The semantic analysis of (9) calls for a "cross-world" mode of evaluation.

In itself, the above point is independent of the question of whether $\bigcirc$ is extensional or not. However (9) may very well follow from an application of the principle of substitution *salva veritate.* Premises:

---

[3] Quine argues for this requirement in his [27]. Notoriously, Kripke [19] defended the view that (E-$\Box$) holds for constants (proper names are rigid designators). We do not make this assumption in this paper.

[4] Since then, Pluto is no longer considered a planet of the solar system (cf. `https://www.iau.org/public/themes/pluto`)

(10) Sam has an obligation not to be a tyrant: $\bigcirc\neg T(s)$

(11) Sam is a tyrant: $s = \imath x T(x)$

Conclusion:

(12) The tyrant has an obligation not to be a tyrant: $\bigcirc\neg T(\imath x T(x))$

One could object that (9) may alternatively be rendered as $\exists x(T(x)\wedge\bigcirc\neg T(x))$. This formalisation is unproblematic. First, we point out that as a spin-off of the extensionality of the deontic operator the principles of universal instantiation and existential generalisation hold unrestrictedly (viz. even if t is inside the scope of a deontic operator).

$$\exists x(x = t) \rightarrow (\forall x\varphi(x) \rightarrow \varphi(t)) \tag{UI}$$

$$\exists x(x = t) \rightarrow (\varphi(t) \rightarrow \exists x\varphi(x)) \tag{EG}$$

Given the assumption $\exists x(x = \imath y T(y))$, the two formalisations are equivalent. Thus the principle of extensionality turns an apparently unproblematic formula ($\exists x(T(x) \wedge \bigcirc\neg T(x))$ into a problematic one ($\bigcirc\neg T(\imath x T(x))$). Our task is to account of the meaningfulness of the later formula. The following two derivations show the equivalence between the two formalisations. We use $\exists!$ for the uniqueness quantification defined as $\exists! x\varphi := \exists x\forall y(\varphi \leftrightarrow y = x)$.

(a) $\exists x(x = \imath y T(y))$	(Hypothesis)
(b) $\exists x(T(x) \wedge \bigcirc\neg T(x))$	(Hypothesis)
(c) $\exists! x T(x)$	(a)
(d) $\exists! x(T(x) \wedge \bigcirc\neg T(x))$	(FO + b + c)
(e) $\forall x(T(x) \rightarrow \bigcirc\neg T(x))$	(FO + d)
(f) $T(\imath y T(y)) \rightarrow \bigcirc\neg T(\imath y T(y))$	(e + UI)
(g) $T(\imath y T(y))$	(a)
(h) $\bigcirc\neg T(\imath y T(y))$	(f + g)

Derivation 1

(a) $\exists x(x = \imath y T(y))$	(Hypothesis)
(b) $\bigcirc\neg T(\imath y T(y))$	(Hypothesis)
(c) $T(\imath y T(y))$	(a)
(d) $T(\imath y T(y)) \wedge \bigcirc\neg T(\imath y T(y))$	(b + c)
(e) $\exists x(T(x) \wedge \bigcirc\neg T(x))$	(d + EG)

Derivation 2

2.2 Collapse

We explain in more detail how the collapse mentioned in requirement 3 arises. The discussion draws on Goble [12,13,14]. We say the deontic collapse arises in a logic if the formula $\varphi \leftrightarrow \bigcirc\varphi$ is derivable (for every formula φ). This would mean that everything that is true is obligatory and vice versa. Goble pointed out that, if the principle of substitution *salva veritate* holds, then the deontic collapse follows. We reiterate and amplify his main points.

The derivation of $\bigcirc\varphi \to \varphi$ presupposes that of $\varphi \to \bigcirc\varphi$. We start with the former one. It appeals to the law of contraposition, the law of double negation elimination, and the **D** axiom for $\bigcirc$:

(a) $\bigcirc\varphi$	(Hypothesis)
(b) $\neg\bigcirc\neg\varphi$	(**D** axiom)
(c) $\neg\neg\varphi$	($\varphi \to \bigcirc\varphi$ and contraposition)
(d) φ	(Double $\neg$ elimination)

Derivation 3

One may be tempted to block this derivation by just abandoning the principle of contraposition or the principle of double $\neg$ elimination. However, this would not block the derivation of $\varphi \to \bigcirc\varphi$, which in itself is counter-intuitive. We turn to this implication. We do not give the original argument, but a variant one, which highlights the role of $\Box$.

Proposition 2.1 *Consider a deontic logic containing (i) the usual principles of first-order logic (FO), (ii) the principle of substitution salva veritate for "ought" (E-$\bigcirc$), $t = s \to (\bigcirc\varphi \leftrightarrow \bigcirc\varphi_{t\hookrightarrow s})$ (iii) the principle $\Box\varphi \to \bigcirc\varphi$ ($\Box 2\bigcirc$) and (iv) the principle of inheritance "If $\vdash \varphi \to \psi$ then $\vdash \bigcirc\varphi \to \bigcirc\psi$" (In). Then $\varphi \to \bigcirc\varphi$ is derivable from $\Box\exists y(y = t)$.*

Proof. In this derivation we assume that x and y do not occur free in φ:

(a) φ	(Hypothesis)
(b) $\Box\exists y(y = t)$	(Hypothesis)
(c) $t = \imath x(x = t \wedge \varphi)$	(FO + a)
(d) $\bigcirc\exists y(y = t)$	($\Box 2\bigcirc$ + b)
(e) $\bigcirc\exists y(y = \imath x(x = t \wedge \varphi))$	(E-$\bigcirc$ + c + d)
(f) $\bigcirc\varphi$	(In + e)

Derivation 4

$\Box$

Some comments are in order:

- We show $\varphi \to \bigcirc\varphi$, where the original argument shows $\bigcirc\psi \to (\varphi \to \bigcirc\varphi)$.
- Our derivation starts from the supposition $\Box\exists y(y = t)$. This may be read as t necessarily denotes. We take this supposition to be harmless. We do not even want the collapse under this assumption.
- Line (c) "drags" φ inside the scope of the definite description to write "the-unique-x-identical-with-t-and-φ". Line (f) "drags" φ outside the scope of the definite description. The move is allowed in first-order logic.
- The principle (E-$\bigcirc$) is used on line (d), where t is replaced by the co-referential term "the-unique-x-identical-with-t-and-φ". The formula (e)

seems already counter-intuitive. However, as we will see in Section 4.3 the two-dimensional semantics presented in this paper gives an unproblematic reading to this formula.

- Line (f) is obtained by applying (In). This final move is explained in more detail in derivation 5.

To avoid the deontic collapse, the following ways out suggest themselves:

Option 1 Revise the laws of first-order logic;

Option 2 Abandon ($\Box 2\bigcirc$);

Option 3 Abandon (In), or restrict its application.

We will go with option 3. Thus, in derivation 4, the move from (e) to (f) is blocked. A good reason for choosing this path is that option 2 alone would not block the original derivation of the collapse in a mono-modal setting, which uses (In) and the laws of first-order logic. Note that in Åqvist's system **F**, (In) is not a primitive rule, but is derivable from ($\Box 2\bigcirc$) and two extra principles:

- the principle of necessitation for $\Box$: "If $\vdash \varphi$, then $\vdash \Box\varphi$" (N-$\Box$)
- the K axiom for $\bigcirc$: $\bigcirc(\varphi \to \psi) \to (\bigcirc\varphi \to \bigcirc\psi)$ (K-$\bigcirc$)

This is easily verified. The move from (e) to (f) is explained thus:

(a) $\vdash \exists y(y = \imath x(x = t \wedge \varphi)) \to \varphi$	(FO)
(b) $\vdash \Box[\exists y(y = \imath x(x = t \wedge \varphi)) \to \varphi]$	(N-$\Box$)
(c) $\vdash \bigcirc[\exists y(y = \imath x(x = t \wedge \varphi)) \to \varphi]$	($\Box 2\bigcirc$)
(d) $\vdash \bigcirc\exists y(y = \imath x(x = t \wedge \varphi)) \to \bigcirc\varphi$	(K-$\bigcirc$)

Derivation 5

Ultimately, the solution will consist in restricting the application of (N-$\Box$). However, the final effect will be the same: (In) will go away in its plain form. Prop. 2.2 provides an independent argument for keeping $\Box$ intensional (cf. requirement 2):

Proposition 2.2 *Consider the same deontic logic as in Prop. 2.1, but with (E-$\bigcirc$) replaced with (E-$\Box$). In such a logic, $\varphi \to \bigcirc\varphi$ is derivable from $\Box\exists y(y = t)$.*

Proof. As before we assume that x and y do not occur free in φ:

(a) φ	(Hypothesis)
(b) $\Box\exists y(y = t)$	(Hypothesis)
(c) $t = \imath x(x = t \wedge \varphi)$	(FO + a)
(d) $\Box\exists y(y = \imath x(x = t \wedge \varphi))$	(E-$\Box$ + b + c)
(e) $\bigcirc\exists y(y = \imath x(x = t \wedge \varphi))$	($\Box 2\bigcirc$)
(f) $\bigcirc\varphi$	(In)

Derivation 6

$\Box$

3 The perspectival account

In this section, we develop in full detail our perspectival account. The basic idea is that the content of an obligation at one world is to be evaluated from the perspective of another one. What we mean by this is the following. Formulas will be evaluated with respect to two dimensions, or pair of worlds (v, w). World v is where the evaluation takes place, and world w is the one from the perspective of which formulas are evaluated (call it the reference or actual world, if you wish). Throughout the paper the reference world will be represented as an upper index in the notation $v \models^w$. What is meant by "φ is evaluated in v from w's perspective" is this: when determining the truth-value of φ in v, the terms occurring in φ get the same denotation as in w.
To get a more flexible framework, we introduce two alethic modal operators, $\boxdot$ and $\boxtimes$. The first will be extensional, and the second intensional. Our prime interest is in combining extensionality for $\bigcirc$ and intensionality for $\Box$. However, there are readings of $\Box$ under which extensionality remains desirable. Hence we allow for both.

Definition 3.1 The language $\mathcal{L}$ contains:

- A countable set of variables $V := \{x, y, z, ...\}$
- A countable set of constants $C := \{c, d, e, ...\}$
- Two propositional connectives $\wedge, \neg$
- Three first-order logic symbols $\forall, \imath, =$
- A binary obligation operator $\bigcirc(-/-)$
- Two unary alethic operators $\boxdot$ and $\boxtimes$
- For each $n \in \mathbb{Z}^+$ a countable set of n-place predicate symbols $\mathbb{P} := \{A^n, B^n, ...\}$

We can now define inductively the well-formed terms and formulas used in our logic and their respective complexity ($\ulcorner ... \urcorner$).

Definition 3.2 [Terms and formulas]

- **Terms:**
  - · Every element of $V \cup C$ is a term of complexity 0
  - · If φ is a formula and $x \in V$ then $\imath x\varphi$ is a term with $\ulcorner \imath x\varphi \urcorner := \ulcorner \varphi \urcorner + 1$
- **Formulas:**
  - · If $R^n \in \mathbb{P}$ is a n-place predicate symbol and $t_1, ..., t_n$ are terms then $R^n(t_1, ..., t_n)$ is a formula with $\ulcorner R^n(t_1, ..., t_n) \urcorner := \sum_{i=1}^{n} \ulcorner t_i \urcorner$
  - · If φ is a formula and $x \in V$ then $\forall x\varphi$ is a formula with $\ulcorner \forall x\varphi \urcorner := \ulcorner \varphi \urcorner + 1$
  - · If t_1 and t_2 are terms then $t_1 = t_2$ is a formula with $\ulcorner t_1 = t_2 \urcorner := \ulcorner t_1 \urcorner + \ulcorner t_2 \urcorner + 1$
  - · If φ is a formula then $\neg\varphi$ is a formula with $\ulcorner \neg\varphi \urcorner := \ulcorner \varphi \urcorner + 1$
  - · If φ is a formula then $\boxdot\varphi$ is a formula with $\ulcorner \boxdot\varphi \urcorner := \ulcorner \varphi \urcorner + 1$
  - · If φ is a formula then $\boxtimes\varphi$ is a formula with $\ulcorner \boxtimes\varphi \urcorner := \ulcorner \varphi \urcorner + 1$
  - · If φ and ψ are formulas then $\varphi \wedge \psi$ is a formula with $\ulcorner \varphi \wedge \psi \urcorner := \ulcorner \varphi \urcorner + \ulcorner \psi \urcorner + 1$

- If φ and ψ are formulas then $\bigcirc(\psi/\varphi)$ is a formula with $\ulcorner\bigcirc(\psi/\varphi)\urcorner := \ulcorner\varphi\urcorner + \ulcorner\psi\urcorner + 1$
- Nothing else is a formula

Definition 3.3 [Derived connectives] Let t be a term. We define $E(t)$ as $\exists x(x = t)$, where x is the first element of V not appearing in t. The symbols $\vee, \bot, \top, \rightarrow, \leftrightarrow, \Diamond\varphi, \Diamond\varphi, P(./.), \exists, \exists!$ and $\neq$ are introduced the usual way.

Definition 3.4 [Frames] $\mathcal{F} = \langle W, \succeq, D\rangle$ is called a frame, where

- $W \neq \emptyset$ is a set of worlds
- $\succeq \subseteq W \times W$ is a binary relation called the betterness relation [5]
- D is a function which maps every world $w \in W$ to a non-empty set D_w

D is called the domain function, and D_w is called the domain of w.
$\mathbb{D} := \bigcup_{w\in W} D_w$ is called the "actual" domain and $\mathbb{D}^+ := \mathbb{D} \cup \{\mathbb{D}\}$ the (whole) domain.

The individual domains $(D_w)_{w\in W}$ contain all objects which are within the range of the universal quantifier at a world w. The actual domain $\mathbb{D}$ is not contained in the domain of any world [6] and is used as the value assigned to definite descriptions that do not designate (uniquely).

Definition 3.5 [Models] $\mathcal{M} = \langle W, \succeq, D, I\rangle$ is called a model (on the frame $\mathcal{F} = \langle W, \succeq, D\rangle$), where I is a function (called interpretation function) such that:

- for $c \in C$ and $w \in W$: $I(c, w) \in \mathbb{D}^+$
- for $R^n \in \mathbb{P}$ and $w \in W$: $I(R^n, w) \subseteq (\mathbb{D}^+)^n$

$I(c, w) = a$ says that a is the denotation of c in w. In our semantics a constant may not denote, and it does not need to designate the same entity in every possible world. In Kripke's terminology, proper names are not rigid designators. We have not investigated the effects of making this assumption.

Definition 3.6 [Variable assignment] Given a model $\mathcal{M} = \langle W, \succeq, D, I\rangle$ we call a function $g : V \times W \mapsto \mathbb{D}^+$ a variable assignment (of $\mathcal{M}$).

Notice that the assignment is world-dependent. Roughly speaking, $g(x, w) = a$ says that a is the denotation of x in w. Note that $g(x, w)$ does not have to be an element of the domain of w. [7] We amend the usual notion of an x-variant as follows. An *x-variant of some variable assignment g at a world w* is a variable assignment h that agrees with g on all values except for x, whose value in every world remains constant, and an element of D_w. Formally:

Definition 3.7 [x-variant] Assume a model $\mathcal{M} = \langle W, \succeq, D, I\rangle$, a variable assignment g of $\mathcal{M}$ and an element of the whole domain $d \in \mathbb{D}^+$. We write $g_{x\Rightarrow d}$

[5] When $w \succeq v$, we say that a world w is at least as good as world v.
[6] $\mathbb{D} \notin \mathbb{D}$.
[7] The element a does not even have to be contained in the actual domain.

for the variable assignment which replaces the value assigned to x at any world by d:

$$g_{x\Rightarrow d}(z,v) := \begin{cases} d & \text{if } (z,v) \in \{x\} \times W \\ g(z,v) & \text{otherwise} \end{cases}$$

A variable assignment h is an *x-variant of g at w* iff $h = g_{x\Rightarrow d}$ for some $d \in D_w$.

"Best", in terms of which the truth-conditions for $\bigcirc(-/-)$ are cast, is defined by:

Definition 3.8 [best] Given a model $\mathcal{M} = \langle W, \succeq, D, I\rangle$ and a set of worlds $X \subseteq W$ we define

$$\mathit{best}(X) := \{w \in X : \forall v \in W(v \in X \Rightarrow w \succeq v)\}$$

$\mathit{best}(X)$ is the set of worlds in X that are at least as good as every member of X.

The construct "$\mathcal{M}, v \models_g^w \varphi$" can be read as "$v$ forces φ under g if looked at from the point of view of (an inhabitant of) w". We stress that $\mathcal{M}, v \models_g^w$ does not convey a truth value for the formula φ per se, but it is used to define the truth conditions of φ by induction. We put $||\varphi||_{g,w}^{\mathcal{M}} := \{v \in W : \mathcal{M}, v \models_g^w \varphi\}$.

Definition 3.9 Let $\mathcal{M} = \langle W, \succeq, D, I\rangle$ be a model, g a variable assignment, $x \in V$ and $c \in C$. We define

- $I_g^w(x) := g(x,w)$
- $I_g^w(c) := I(c,w)$
- $I_g^w(\imath x\varphi) := \begin{cases} h(x,w) & \text{if } h \text{ is the } \textbf{unique } x\text{-variant of } g \text{ at } w \\ & \text{such that } \mathcal{M}, w \models_h^w \varphi \\ \mathbb{D} & \text{otherwise} \end{cases}$

The forcing relation $\models$ can be defined inductively as follows:

- $\mathcal{M}, v \models_g^w R^n(t_1, ..., t_n) :\Leftrightarrow \langle I_g^w(t_1), ..., I_g^w(t_n)\rangle \in I(R^n, v)$
- $\mathcal{M}, v \models_g^w \neg\varphi :\Leftrightarrow \mathcal{M}, v \not\models_g^w \varphi$
- $\mathcal{M}, v \models_g^w \varphi \wedge \psi :\Leftrightarrow \mathcal{M}, v \models_g^w \varphi$ and $\mathcal{M}, v \models_g^w \psi$
- $\mathcal{M}, v \models_g^w \forall x\varphi :\Leftrightarrow \mathcal{M}, v \models_h^w \varphi$ for all x-variants h of g at v
- $\mathcal{M}, v \models_g^w t_1 = t_2 :\Leftrightarrow I_g^w(t_1) = I_g^w(t_2)$
- $\mathcal{M}, v \models_g^w \boxdot\varphi :\Leftrightarrow \forall u \in W\ \mathcal{M}, u \models_g^w \varphi$
- $\mathcal{M}, v \models_g^w \boxtimes\varphi :\Leftrightarrow \forall u\ \forall v' \in W\ \mathcal{M}, u \models_g^{v'} \varphi$
- $\mathcal{M}, v \models_g^w \bigcirc(\psi/\varphi) :\Leftrightarrow \mathit{best}(||\varphi||_{g,w}^{\mathcal{M}}) \subseteq ||\psi||_{g,w}^{\mathcal{M}}$

We drop the reference to $\mathcal{M}$ when it is clear what model is intended.

Definition 3.10 [Truth in $\mathbf{F}^{\forall}$] Given a model $\mathcal{M} = \langle W, \succeq, D, I\rangle$, a variable assignment g, a formula φ and a world w we define what it means that φ is true in $\mathcal{M}$ at w under g (in symbols: $\mathcal{M}, w \models_g \varphi$) as

$$\mathcal{M}, w \models_g \varphi :\Leftrightarrow \mathcal{M}, w \models_g^w \varphi$$

The meaning of $\boxdot$, $\boxtimes$ and $\bigcirc$ is easier to explain using the following derived truth conditions.

Remark 3.11 [Derived truth conditions]

- $\mathcal{M}, w \models_g \boxdot\varphi :\Leftrightarrow \forall v \in W\ \mathcal{M}, v \models_g^w \varphi$
- $\mathcal{M}, w \models_g \boxtimes\varphi :\Leftrightarrow \forall u\ \forall v \in W\ \mathcal{M}, u \models_g^v \varphi$
- $\mathcal{M}, w \models_g \bigcirc(\psi/\varphi) :\Leftrightarrow best(||\varphi||_{g,w}^{\mathcal{M}}) \subseteq ||\psi||_{g,w}^{\mathcal{M}}$

When evaluating the truth-value of $\boxdot\varphi$ at w, one moves to an arbitrary world v, and determines the truth-value of φ in v from w's perspective. This means giving to the terms occurring in φ the denotation they have in w. When evaluating the truth-value of $\boxtimes\varphi$ at w, one moves to an arbitrary world u, and evaluates φ in u from every other world's v perspective. For obligation, the idea is similar. The standard evaluation rule puts $\bigcirc(\psi/\varphi)$ as true whenever all the best φ-worlds are ψ-worlds. The φ-worlds and the ψ-worlds in question are those according to w's perspective. This is how the principle of substitution *salva veritate* will be validated for $\bigcirc$ and $\boxdot$, and invalidated for $\boxtimes$.

Definition 3.12 Given a model $\mathcal{M} = \langle W, \succeq, D, I\rangle$. $\succeq$ is reflexive if $\forall w \in W(w \succeq w)$, and $\succeq$ fulfils the limitedness condition if for every φ, g and $w \in W$ we have

$$||\varphi||_{g,w}^{\mathcal{M}} \neq \emptyset \Rightarrow best(||\varphi||_{g,w}^{\mathcal{M}}) \neq \emptyset$$

$\mathcal{U}$ is the class of models in which $\succeq$ is reflexive and fulfils limitedness.

Intuitively, the limitedness condition validates the dyadic version of the **D** axiom (with $\Diamond$ replaced with $\Diamond$) involved in derivation 3 of the collapse (see Subsect. 2.2).

Definition 3.13 [Validity in $\mathbf{F}^{\forall}$]

- φ is valid at w in a model $\mathcal{M}$ (notation: $\mathcal{M}, w \models \varphi$) if for every variable assignment g, we have that $\mathcal{M}, w \models_g \varphi$;
- φ is valid in a model $\mathcal{M}$ (notation: $\mathcal{M} \models \varphi$) if for every world w we have $\mathcal{M}, w \models \varphi$;
- φ is valid in a class $\mathbb{M}$ of models (notation: $\mathbb{M} \models \varphi$) if for every model $\mathcal{M} \in \mathbb{M}$ we have $\mathcal{M} \models \varphi$;
- φ is valid (notation: $\models \varphi$) if φ is valid in the class $\mathcal{U}$ as defined above.

4 Benchmarking

We test the account introduced in Sect. 3 against the requirements discussed in Sect. 2.

4.1 Extensionality / intensionality / self-negation

A proof of the principle of extensionality in its general form is given in Subsect. 4.2. For simplicity's sake, here we only discuss the examples considered in Sect. 2.

Proposition 4.1 (Extensionality of $\bigcirc$, requirement 1) *We have:*

$j = \imath x P(x) \rightarrow (\bigcirc(S(\imath x P(x)) \leftrightarrow \bigcirc S(j))$

$$j = \imath x P(x) \rightarrow [\bigcirc(El(\imath y(y \neq \imath x P(x)))/\neg S(\imath x P(x))) \leftrightarrow \bigcirc(El(\imath y(y \neq \imath x P(x)))/\neg S(j))]$$

Proof. When a formula does not contain a free variable its truth condition does not depend on which variable assignment is assumed. Therefore for this and all future proofs (in which no free variable is involved) we always deal with an arbitrary variable assignment. Now, if $w \models_g^w j = \imath x P(x)$, then for every $u \in best(||\top||_{g,w}^{\mathcal{M}})$

$$u \models_g^w S(\imath x P(x)) \Leftrightarrow u \models_g^w S(j)$$

This is because the terms on both sides get the denotation they have in w. Similarly:

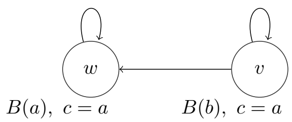
$$\begin{aligned} &best(||\neg S(\imath x P(x))||_{g,w}^{\mathcal{M}}) \subseteq ||El(\imath y(y \neq \imath x P(x)))||_{g,w}^{\mathcal{M}} \\ \Leftrightarrow\ &best(||\neg S(j)||_{g,w}^{\mathcal{M}}) \subseteq ||El(\imath y(y \neq \imath x P(x)))||_{g,w}^{\mathcal{M}} \end{aligned}$$

□

Proposition 4.2 (Intensionality of $\boxtimes$, requirement 2) *We do not have:*

$$c = \imath x B(x) \rightarrow (\boxtimes(c = c) \leftrightarrow \boxtimes(c = \imath x B(x)))$$

Proof. Put $\mathcal{M} = \langle W, \succeq, I, D \rangle$ with (an arrow from v to w means $v \succeq w$, and no arrow from from w to v means $w \not\succeq v$):



$W := \{w, v\}$
$\succeq :=$ the reflexive closure of $\{(v, w)\}\})$
$D_w := \{a\}, \quad D_v := \{a, b\}$
$I(B, w) := a, \quad I(B, w) := b$
$I(c, w) := a, \quad I(c, v) := a$

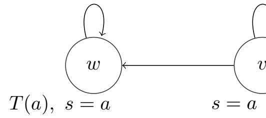
The condition of limitedness is fulfilled. We have:

- $w \models_g^w c = \imath x B(x)$ since c and $\imath x B(x)$ denote a in w
- $w \models_g^w \boxtimes(c = c)$ since $c = c$ is a tautology
- $w \not\models_g^w \boxtimes(c = \imath x B(x))$ since $w \not\models_g^v c = \imath x B(x)$ [8]

□

Proposition 4.3 (Self-negation, requirement 4) *The sentences (10), (11) and (12) are simultaneously satisfiable.*

Proof. We give a model which satisfies all three formulas in the same world.



$W := \{w, v\}$
$\succeq :=$ the reflexive closure of $\{(v, w)\}\})$
$D_w := \{a\}, \quad D_v := \{a\}$
$I(T, w) := \{a\}, \quad I(T, v) := \emptyset$
$I(s, w) := a, \quad I(s, v) := a$

[8] c and $\imath x B(x)$ do not have the same denotation in v.

As before $\succeq$ is limited. We have:

- $w \models^w_g s = \imath x T(x)$ since s and $\imath x T(x)$ denote a in w
- $w \models^w_g \bigcirc \neg T(s)$ since a is not T in v
- $w \models^w_g \bigcirc \neg T(\imath x T(x))$ since a (=the unique T in w) is not T in v

The paradox is resolved by having Sam, who is the tyrant in the actual world w, not be a tyrant in the best world v. Therefore $\bigcirc \neg T(\imath x T(x))$ can be satisfied. $\square$

4.2 Extensionality (general form)

We show the principle of extensionality in its general form. Where φ is a formula and s and t terms, let $\varphi_{t \hookrightarrow s}$ be the result of replacing zero up to all unbound occurrences of t,[9] in φ, by s. We may re-letter bound variables, if necessary, to avoid rendering the new occurrences of variables in s bound in φ.

Proposition 4.4 *Consider some g and some w in $\mathcal{M}$ such that $w \models^w_g t = s$. Then, for all v in $\mathcal{M}$,*

$$v \models^w_g \varphi \leftrightarrow \varphi_{t \hookrightarrow s} \qquad (\#)$$

provided t is not contained in the scope of the $\boxtimes$ operator in φ.

Proof. By induction on the complexity n of a formula φ. The base case, if φ is $R(t_1, ..., t_m)$ with $\ulcorner R(t_1, ..., t_m) \urcorner = 0$, follows from the definitions involved. For the inductive case, we assume (#) holds for all $k < n$, and *for all v in $\mathcal{M}$*. We only consider three cases−the other ones are left to the reader:

- $\varphi := \forall x\ \psi$. Given the restrictions put on t and s, we have the following chain of equivalences:

$$\begin{aligned} v \models^w_g \forall x\ \psi \text{ iff } & v \models^w_h \psi \quad \text{for all } x\text{-variants } h \text{ at } v \\ & v \models^w_h \psi_{t \hookrightarrow s} \quad \text{for all } x\text{-variants } h \text{ at } v \text{ (by IH)} \\ & v \models^w_g \forall x\ \psi_{t \hookrightarrow s} \end{aligned}$$

- $\varphi := \bigcirc(\chi/\psi)$.

$$\begin{aligned} v \models^w_g \bigcirc(\chi/\psi) \text{ iff } & best(||\psi||^{\mathcal{M}}_{g,w}) \subseteq ||\chi||^{\mathcal{M}}_{g,w} \\ & best(||\psi_{t \hookrightarrow s}||^{\mathcal{M}}_{g,w}) \subseteq ||\chi_{t \hookrightarrow s}||^{\mathcal{M}}_{g,w} \text{ (by IH)} \\ & v \models^w_g \bigcirc(\chi_{t \hookrightarrow s}/\psi_{t \hookrightarrow s}) \\ & v \models^w_g \bigcirc(\chi/\psi)_{t \hookrightarrow s} \end{aligned}$$

- $\varphi := R(t_1, ..., t_m)$. Assume $v \models^w_g R(t_1, ..., t_m)$. If t appears only as one of the t_i's, then we are done. So let us suppose that t appears *in* one (or more) of the t_i's. W.l.o.g. let t only appear in $t_1 = \imath x \psi$. By the IH $w \models^w_g \psi \leftrightarrow \psi_{t \hookrightarrow s}$, so $I^w_g(\imath x \psi) = I^w_g(\imath x \psi_{t \hookrightarrow s})$. Consider some $v \in W$.

[9] By an unbounded occurrence of t, we mean that no variables in t are in the scope of a quantifier or a definite description *not* in t.

We have $\langle I^w_g(\imath x\psi), ..., t_m\rangle \in I(R, v)$, so $\langle I^w_g(\imath x\psi_{t\hookrightarrow s}), ..., t_m\rangle \in I(R, v)$. Hence $v \models^w_g R(t_1, ..., t_m)_{t\hookrightarrow s}$ as required. For the converse implication, the argument is the same.

□

Corollary 4.5 (Extensionality) *The principle (E) is valid:*

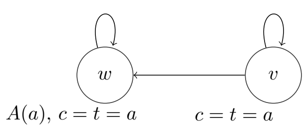
$$\models t = s \to (\varphi \leftrightarrow \varphi_{t\hookrightarrow s}) \quad \textit{if } t \textit{ is not in the scope of } \boxtimes \tag{E}$$

Proof. This follows from Prop. 4.4 putting $v = w$. □

4.3 Deontic collapse

We start by explaining how the collapse is avoided semantically. We define a model in which the formulas at steps (a)-(e) in derivation 4 are true in the actual world w but the formula at step (f) is not.

Example 4.6 Put $\varphi := A(c)$. $\mathcal{M}$ is defined by



$W := \{w, v\}$
$\succeq :=$ the reflexive closure of $\{(v, w)\}$
$D_w := \{a\}, \quad D_v := \{a\}$
$I(c, w) := I(c, v) := a$
$I(t, w) := I(t, v) := a$
$I(A, w) := \{a\}, \quad I(A, v) := \emptyset$

We have

(a) $w \models_g A(c)$ since $I(c, w) = a \in I(A, w)$
(b) $w \models_g \boxtimes\exists y(y = t)$ since $I(t, w) = I(t, v) = a \in D_w$
and $I(t, w) = I(t, v) = a \in D_v$
(c) $w \models_g t = \imath x(x = t \wedge A(c))$ since $I(t, w) = a = I^w_g(\imath x(x = t \wedge A(c)))$
(d) $w \models_g \bigcirc\exists y(y = t)$ since $I(t, w) = a \in D_v$ [10]
(e) $w \models_g \bigcirc\exists y(y = \imath x(x = t \wedge A(c)))$ since $I^w_g(\imath x(x = t \wedge A(c))) = a \in D_v$
(f) $w \not\models_g \bigcirc A(c)$ since $I(c, w) = a \notin I(A, v)$

Let it be clear that (e) means $v \models^w_g \exists y(y = \imath x(x = t \wedge A(c)))$, which says that the unique x, for which the formula $x = t \wedge A(c)$ holds in w, exists in v. However this does NOT imply $v \models^w_g \exists x(x = t \wedge A(c)))$, since there exists no element in the domain of v for which the formula $x = t \wedge A(c)$ holds in v from w's perspective. In the statements, $v \models^w_g \exists y(y = \imath x(x = t \wedge A(c)))$ and $v \models^w_g \exists x(x = t \wedge A(c))$ the two c refer to the same individual a, but in different worlds where they have different properties.

This model serves as a counter-model to the rule of inheritance. The formula $\exists y(y = \imath x(x = t \wedge A(c))) \to A(c)$ is valid, but not $\bigcirc\exists y(y = \imath x(x = t \wedge A(c))) \to \bigcirc A(c)$.

[10] By definition $v \models^w_g \exists y(y = t)$ holds if there exists an y-variant h of g at v such that $h(y, w) = I(t, w)$. This is equivalent to $I(t, w)$ being an element of D_v.

To explain how the deontic collapse is avoided proof-theoretically, we introduce the notion of "variable only" version φ^* of a formula φ. Intuitively, φ^* is obtained by substituting, in φ, a new variable for every definite description and constant occurring in φ. This ensures that φ^* contains only variables, making it impossible to apply the rule of inheritance (and necessitation) from which the collapse follows. Formally:

Definition 4.7 [Variable only version, Goble [13]] Given a formula φ, we define φ^* as the formula in which all terms $t_1, ..., t_n$, which are not variables and are occurring in the formula φ, have been replaced by $x_1, ..., x_n \in V$ respectively. The variables $x_1, ..., x_n$ are the first, pairwise different, elements of V such that $x_1, ..., x_n$ do not occur in φ.

Example 4.8 Let A, B and C be predicate symbols, $x, y, z \in V$ the first three variables of V, $c \in C$ a constant and $\varphi \in WF$ a well-formed formula:

- $A(\imath y\varphi, c)^* = A(x, z)$
- $A(\imath yB(\imath xC(x, y)), y)^* = A(z, y)$
- $\forall xA(\imath yB(y, d), x)^* = \forall xA(z, x)$
- $A(y, y)^* = A(y, y)$

Like in Goble's original treatment, the collapse is blocked by restricting the application of the rule of necessitation for $\boxtimes$, and of the principle of inheritance for $\bigcirc$. These two are now available in the following form:

$$\text{If } \models \varphi^* \text{ then } \models \boxtimes\varphi \qquad (\text{N}^*\text{-}\boxtimes)$$
$$\text{If } \models (\psi_1 \to \psi_2)^* \text{ then } \models \bigcirc(\psi_1/\varphi) \to \bigcirc(\psi_2/\varphi)) \qquad (\text{In}^\star)$$

Before continuing want to point out that the other law involved in the collapse, $\boxtimes\psi \to \bigcirc(\psi/\varphi)$, still holds. This follows at once from the following:

Proposition 4.9 *We have*

$$\models \boxtimes\psi \to \boxdot\psi \qquad (\boxtimes 2\boxdot)$$
$$\models \boxdot\psi \to \bigcirc(\psi/\varphi) \qquad (\boxdot 2\bigcirc)$$

Proof. $(\boxtimes 2\boxdot)$ is straightforward, and may be left to the reader. For $(\boxdot 2\bigcirc)$, let us assume $w \models_g \boxdot\psi$ holds for a fixed model $\mathcal{M} = \langle W, \succeq, D, I\rangle$, a world $w \in W$ and a variable assignment g. This is equivalent to $||\psi||^{\mathcal{M}}_{g,w}$ being equal to the whole set of worlds W. Hence we can infer that for any formula φ we have $best(||\varphi||^{\mathcal{M}}_{g,w}) \subseteq W = ||\psi||^{\mathcal{M}}_{g,w}$, which, by definition, means $w \models_g \bigcirc(\psi/\varphi)$. □

We now show that the rules $(\text{N}^*\text{-}\boxtimes)$ and $(\text{In}^\star)$ preserve validity. To show this we need the following two lemmas.

Lemma 4.10 *Given a formula φ and a model $\mathcal{M}$, then*

$$\mathcal{M} \models \varphi^* \Rightarrow \mathcal{M} \models \boxtimes(\varphi^*)$$

Proof. Let φ be a formula and $\mathcal{M} = \langle W, \succeq, D, I\rangle$ a model. If for every world $w \in W$ and every variable assignment g of $\mathcal{M}$ it holds that $w \models_g \varphi^*$, it follows that $w \models^w_g \varphi^*$ holds for every world $w \in W$ and every variable assignment g of $\mathcal{M}$. Now let us take two arbitrary but fixed worlds $v, w \in W$ and an arbitrary but fixed variable assignment g and define a new variable assignment $h : V \times W \to \mathbb{D}^+$ of $\mathcal{M}$ as:

$$h(x,u) := \begin{cases} g(x,w) & \text{if } \ u = v \\ g(x,v) & \text{if } \ u = w \\ g(x,u) & \text{otherwise} \end{cases}$$

Since h and g only swap how they see the variables at w and v, and φ^* does not contain constants or definite descriptions, we get $\forall u(u \models_g^w \varphi^* \Leftrightarrow u \models_h^v \varphi^*)$. Therefore from $v \models_h^v \varphi^*$, which holds by assumption, we can infer $v \models_g^w \varphi^*$. Since $v, w \in W$ and g were arbitrary we can conclude $\mathcal{M} \models \boxtimes\varphi^*$. □

Lemma 4.11 *Given a formula φ and a model $\mathcal{M}$, then*

$$\mathcal{M} \models \varphi^* \Rightarrow \mathcal{M} \models \varphi$$

Proof. This proof is done by contraposition. Suppose there are $\mathcal{M} = \langle W, \succeq, D, I\rangle$, $w \in W$ and g such that $w \not\models_g^w \varphi$. Let $t_1, ..., t_n$ be all terms in φ which are replaced by the corresponding variables $x_1, ..., x_n$ in φ^*. Then for the variable assignment

$$h(x,v) := \begin{cases} I_g^v(t_i) & \text{if } \ (x,v) \in \{x_i\} \times W \ \ \text{where } \ i \in \{1, ..., n\} \\ g(x,v) & \text{otherwise} \end{cases}$$

we have $w \not\models_h^w \varphi^*$. □

Putting those two lemmas together, we can prove the soundness of (N*-$\boxtimes$):

Lemma 4.12 *Given a formula φ and a model $\mathcal{M}$ then*

$$\mathcal{M} \models \varphi^* \quad \textit{implies} \quad \mathcal{M} \models \boxtimes\varphi$$

Proof. $\mathcal{M} \models \varphi^* \Rightarrow \mathcal{M} \models \boxtimes(\varphi^*) \Leftrightarrow \mathcal{M} \models (\boxtimes\varphi)^* \Rightarrow \mathcal{M} \models \boxtimes\varphi$. □

Theorem 4.13 *We have*

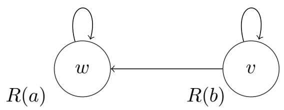
$$\textit{If } \models \varphi^* \textit{ then } \models \boxtimes\varphi \tag{N*-$\boxtimes$}$$

$$\textit{If } \models (\psi_1 \to \psi_2)^* \textit{ then } \models \bigcirc(\psi_1/\varphi) \to \bigcirc(\psi_2/\varphi) \tag{In*}$$

Proof. The first rule follows at once from Lem. 4.12. The second rule follows from the first one and Prop. 4.9. □

We end with the observation that the rule of necessitation in its plain form fails for $\boxtimes$. Here is a counter-example. The formula $\exists y(y = \imath x R(x)) \to R(\imath x R(x))$ is valid in any model. To see why, fix a model $\mathcal{M} = \langle W, \succeq, D, I\rangle$, a variable assignment g, and a world $w \in W$. Assume $w \models_g \exists y(y = \imath x R(x))$. Hence, there exists a y-variant h of g at w such that $h(y,w) = I_h^w(\imath x R(x))$. This means that $h(y,w) = a$ for some $a \in D_w$. By definition of $\imath x R(x)$, a is the unique element in D_w s.t. $a \in I(R,w)$. So $w \models_h R(\imath x R(x))$. Since y does not occur in $R(\imath x R(x))$ we conclude $w \models_g R(\imath x R(x))$ as required.
Now we define a model in which $\boxtimes[\exists y(y = \imath x R(x)) \to R(\imath x R(x))]$ is not valid:

Example 4.14 Consider the model $\mathcal{M} := \langle W, \succeq, D, I \rangle$ with



$W := \{w, v\}$
$\succeq :=$ the reflexive closure of $\{(v, w)\}$
$D_w := \{a, b\}, \quad D_v := \{a, b\}$
$I(R, w) := \{a\}, \quad I(R, v) := \{b\}$

We have $v \models_g^w \exists y(y = \imath x R(x))$, as $I_g^w(\imath x R(x)) = a \in D_v$. But $v \not\models_g^w R(\imath x R(x))$ because $I_g^w(\imath x R(x)) = a \notin I(R, v)$. So $\mathcal{M} \not\models \boxtimes[\exists y(y = \imath x R(x)) \rightarrow R(\imath x R(x))]$.

5 Concluding remarks

We have defined and studied a new perspectival account of conditional obligation. A number of requirements were identified, and shown to be met by the framework. The framework allows for a more nuanced way of approaching first-order deontic principles.
Topics for future research include:

(i) to investigate variant candidate truth-conditions for $\boxtimes$

(ii) to find a suitable axiomatic basis

Ad (i): the truth-conditions for $\boxtimes$ in Def. 3.9 allowed us to make the minimal changes to the axiomatic basis of **F**. The most significant change is that Lewis's absoluteness principle $\bigcirc(\psi/\varphi) \rightarrow \boxtimes \bigcirc (\psi/\varphi)$, stipulating that obligations are necessary, goes away. This may be considered good news. But ($\boxtimes 2 \bigcirc$) remains, and this law may be considered counter-intuitive. The following alternative truth-conditions may be used:

$$w \models_g \boxtimes\varphi \text{ iff } \forall v : \ v \models_g^v \varphi$$

Intuitively: $w \models_g \boxtimes\varphi$ holds, if φ holds at all v under the hypothesis that the terms occurring in φ take the reference they have in this very same world. With this definition of $\boxtimes$, ($\boxtimes 2 \bigcirc$) goes away, and the rule of necessitation holds without any restriction.
Ad (ii): we have identified a sound axiomatic basis for the logic. This one is shown in Appendix B. Completeness is left as a topic for future research.

Appendix A: Åqvist's system F

Axioms:
All truth-functional tautologies
S5-schemata for $\Box$ and $\Diamond$
$\bigcirc(\varphi \to \chi/\psi) \to (\bigcirc(\varphi/\psi) \to \bigcirc(\chi/\psi))$
$\bigcirc(\varphi/\psi) \to \Box \bigcirc(\varphi/\psi)$
$\Box\varphi \to \bigcirc(\varphi/\psi)$
$\Box(\varphi \leftrightarrow \psi) \to (\bigcirc(\chi/\varphi) \leftrightarrow \bigcirc(\chi/\psi))$
$\bigcirc(\varphi/\varphi)$
$\bigcirc(\varphi/\psi \wedge \chi) \to \bigcirc(\chi \to \varphi/\psi)$
$\Diamond\psi \to (\bigcirc(\varphi/\psi) \to P(\varphi/\psi))$

Rules:
If $\vdash \varphi$ and $\vdash \varphi \to \chi$ then $\vdash \chi$
If $\vdash \varphi$ then $\vdash \Box\varphi$

An explanation of the axioms can be found in [24]. The dyadic version of the **D** axiom ($\Diamond\psi \to (\bigcirc(\varphi/\psi) \to P(\varphi/\psi))$) is the distinguishing axiom of this logic. This axiom makes the system **F** the weakest system in the family of Åqvist's systems in which the collapse arises.

Appendix B: Axiomatisation of $\mathbf{F}^{\forall}$

A sound Hilbert axiomatisation of the logic developed in this paper is shown below. In this axiomatisation, the symbol $\varphi_{x \Rightarrow t}$ is the result of replacing ALL occurrences of the variable x, in φ, by the term t. Furthermore, we write $free(\varphi)$ for the set of variables appearing in φ, which are not bound by a quantifier or a definite description.

Axioms:

All truth functional tautologies	
All axioms of system **F**	with $\Box$ replaced with $\boxdot$ and $\Diamond$ with $\diamondcdot$
S5-schemata for $\boxtimes$ and $\diamondtimes$	
$\boxtimes\varphi \to \boxdot\varphi$	
$\boxtimes\psi \to \boxtimes \bigcirc(\psi/\varphi)$	
$\boxtimes(\varphi \leftrightarrow \psi) \to \boxtimes(\bigcirc(\chi/\varphi) \leftrightarrow \bigcirc(\chi/\psi))$	
$t = s \to (\varphi \leftrightarrow \varphi_{t \hookrightarrow s})$	if t is not in the scope of the $\boxtimes$ operator
$E(t) \to (\forall x\varphi \to \varphi_{x \Rightarrow t})$	if x is not in the scope of the $\boxtimes$ operator
$\exists x \exists y(x = y)$	
$t = t$	

$t \neq s \to \boxdot t \neq s$
$\forall y((\forall x(\varphi \leftrightarrow x = y)) \to y = \imath x\varphi)$
$E(\imath x\varphi) \to \exists! x\varphi$
$\forall x(E(x) \to \varphi) \to \forall x\varphi$
$(\forall x\varphi \wedge \forall x\psi) \leftrightarrow \forall x(\varphi \wedge \psi)$

Rules:
If $\vdash \varphi$ and $\vdash \varphi \to \chi$ then $\vdash \chi$
If $\vdash \varphi^*$ then $\vdash \boxtimes\varphi$
If $\vdash \bigcirc(\varphi/\psi)$ then $\vdash \boxtimes \bigcirc (\varphi/\psi)$
If $\vdash \varphi \to t \neq x$ then $\vdash \neg\varphi$ where $x \notin$ free(φ)
If $\vdash \varphi \to \psi$ then $\vdash \varphi \to \forall x\psi$ where $x \notin$ free(φ)
If $\vdash \varphi \to \boxdot\psi$ then $\vdash \varphi \to \boxdot\forall x\psi$ where $x \notin$ free(φ)
If $\vdash \varphi \to \boxtimes\psi$ then $\vdash \varphi \to \boxtimes\forall x\psi$ where $x \notin$ free(φ)

An explanation of the first-order and definite description axioms can be found in [30].

References

[1] Åqvist, L., "An Introduction to Deontic logic and the Theory of Normative Systems," Bibliopolis, Naples, 1987.
[2] Castañeda, H., *The paradoxes of deontic logic: The simplest solution to all of them in one fell swoop*, in: R. Hilpinen, editor, *New Studies in Deontic Logic*, Springer, 1981 pp. 37–85.
[3] Chalmers, D., *Epistemic two-dimensional semantics*, Philosophical Studies **118** (2004), p. 153–226.
[4] Chisholm, R. M., *Contrary-to-duty imperatives and deontic logic*, Analysis **24** (1963), pp. 33–36.
[5] Danielsson, S., "Preference and Obligation," Filosofiska Färeningen, Uppsala, 1968.
[6] de Boer, M., D. M. Gabbay, X. Parent and M. Slavkovic, *Two dimensional standard deontic logic*, Synth. **187** (2012), pp. 623–660.
[7] Frijters, S., "All Doctors Have an Obligation to Care for Their Patients: Term-Modal Logics for Ethical Reasoning with Quantified Deontic Statements," Ph.D. thesis, Ghent (2021).
URL `http://hdl.handle.net/1854/LU-8698101`
[8] Frijters, S. and T. D. Coninck, *The Manchester twins: Conflicts between directed obligations*, in: F. Liu, A. Marra, P. Portner and F. V. D. Putte, editors, *Deontic Logic and Normative Systems - 15th International Conference, DEON 2020/21* (2021), pp. 166–182.
[9] Fusco, M., *A two-dimensional logic for diagonalization and the a priori*, Synth. **198** (2021), pp. 8307–8322.
[10] Fusco, M. and A. W. Kocurek, *A two-dimensional logic for two paradoxes of deontic modality*, Rev. Symb. Log. **15** (2022), pp. 991–1022.
[11] Gabbay, D. M., A. Kurucz, F. Wolter and M. Zakharyaschev, "Multi-dimensional modal logic," Elsevier, 2003.
[12] Goble, L., *Opacity and the ought-to-be*, Noûs (1973), pp. 407–412.

[13] Goble, L., *Quantified deontic logic with definite descriptions*, Logique et Analyse **37** (1994), pp. 239–253.
[14] Goble, L., *'Ought' and extensionality*, Noûs **30** (1996), pp. 330–355.
[15] Goble, L., *Axioms for Hansson's dyadic deontic logics*, Filosofiska Notiser **6** (2019), pp. 13–61.
[16] Hansson, B., *An analysis of some deontic logics*, Noûs (1969), pp. 373–398.
[17] Horty, J. F., *Perspectival act utilitarianism*, in: P. Girard, O. Roy and M. Marion, editors, *Dynamic Formal Epistemology*, Springer Netherlands, Dordrecht, 2011 pp. 197–221.
[18] Humberstone, L., *Two-dimensional adventures*, Philosophical Studies **118** (2004), pp. 17–65.
[19] Kripke, S., "Naming and Necessity," Harvard University Press, Cambridge, 1980.
[20] Kölbel, M., "Truth without Objectivity," Routledge, 2002.
[21] Lewis, D., "Counterfactuals," Blackwell, Oxford, 1973.
[22] MacFarlane, J., "Assessment Sensitivity: Relative Truth and its Applications," Oxford University Press, 2014.
[23] Parent, X., *Completeness of Åqvist's systems E and F*, The Review of Symbolic Logic **8** (2015), pp. 164–177.
[24] Parent, X., *Preference-based semantics for Hansson-type dyadic deontic logic*, in: D. Gabbay, J. Horty, X. Parent, R. van der Meyden and L. van der Torre, editors, *The Handbook of Deontic Logic and Normative Systems*, College Publications, 2021 pp. 1–70, volume 2.
[25] Pichler, D., "Extensionality of obligations in Åqvist's system F," Master's thesis, TU Wien (2022).
[26] Prakken, H. and M. Sergot, *Dyadic deontic logic and contrary-to-duty obligations*, in: D. Nute, editor, *Defeasible Deontic Logic*, Kluwer, Dordrecht, 1997 pp. 223–262.
[27] Quine, W. V., *Notes on existence and necessity*, The Journal of Philosophy **40** (1943), pp. 113–127.
[28] Sawasaki, T. and K. Sano, *Term-sequence-dyadic deontic logic*, in: F. Liu, A. Marra, P. Portner and F. V. D. Putte, editors, *Deontic Logic and Normative Systems - 15th International Conference, DEON 2020/21, Munich, Germany [virtual], July 21-24, 2021* (2021), pp. 376–393.
[29] Segerberg, K., *Two-dimensional modal logic*, Journal of Philosophical Logic **2** (1973), pp. 77–96.
[30] Thomason, R. H., *Some completeness results for modal predicate calculi*, in: *Philosophical problems in logic*, Springer, 1970 pp. 56–76.
[31] von Wright, G. H., *Deontic logic*, Mind **60** (1951), pp. 1–15.