Tolerating Deontic Conflicts by Adaptively Restricting Inheritance*

Christian Straßer and Joke Meheus and Mathieu Beirlaen
Centre for Logic and Philosophy of Science, Ghent University
{christian.strasser, joke.meheus, mathieu.beirlaen}@UGent.be

August 26, 2010

Abstract

In order to deal with the possibility of deontic conflicts Lou Goble developed a group of logics (**DPM**) that are characterized by a restriction of the inheritance principle. While they approximate the deductive power of standard deontic logic, they do so only if the user adds certain statements to the premises.

By adaptively strengthening the **DPM** logics, this paper presents logics that overcome this shortcoming. Furthermore, they are capable of modeling the dynamic and defeasible aspect of our normative reasoning by their dynamic proof theory. This way they enable us to have a better insight in the relations between obligations and thus to localize deontic conflicts.

Keywords: deontic conflicts, adaptive logic, defeasible reasoning, conflict-tolerance, deontic logic

1 Introduction

Recent work in deontic logics has shown a growing interest in systems that are able to deal with deontic conflicts (e.g., [4, 5, 6, 9, 10, 12, 13, 14, 15, 19]). A deontic conflict between obligations occurs when the obligations cannot be jointly realized. Note that deontic conflicts are not just an abstruse philosophical notion, but that they occur quite commonly in our every-day moral lives (see e.g. [11, 17]). This has for instance to do with the fact that different obligations and behavioral codices may stem from different moral systems and institutions. Sartre famously reports on one of his students who found himself in an unfortunate situation. On the one hand, he felt obliged to support the French army in their resistance against Nazi Germany. On the other hand, however, there was

---

[*] Research for this paper was supported by the Research Fund of Ghent University by means of Research Project 01G01907.

the obligation to stay at home in order to support his ill mother. Obviously, it was not possible for him to fulfill both obligations simultaneously.

In deontic logics a modal operator O is used where $\mathsf{O}A$ expresses the obligation to bring about A. In order to accommodate deontic conflicts logicians are urged to develop systems that tolerate them, i.e., that do not lead to triviality when applied to conflicting obligations. Formally, conflict-tolerant deontic logics do not validate the following principle of deontic explosion:

$$\vdash (\mathsf{O}A \wedge \mathsf{O}\neg A) \supset \mathsf{O}B \tag{DEX}$$

Note that standard deontic logic (**SDL**) is not conflict-tolerant. One reason is that it validates the principle (D), $\vdash \mathsf{O}A \supset \neg\mathsf{O}\neg A$. Thus, $\neg(\mathsf{O}A \wedge \mathsf{O}\neg A)$ is a theorem of **SDL** and all conflicts of the form $\mathsf{O}A \wedge \mathsf{O}\neg A$ lead to explosion.

There are various proposals for conflict-tolerant deontic logics. First, one could restrict or reject the *ex contradictione quodlibet* principle ($(A \wedge \neg A) \vdash B$, for any B), i.e., go paraconsistent (see e.g. [4, 5, 14]). Another approach is to restrict the aggregation principle (if $\mathsf{O}A$ and $\mathsf{O}B$, then $\mathsf{O}(A \wedge B)$) or to abandon it (see [6, 8, 13, 15]).

Yet another approach is given by Goble's logics **DPM** (see [7, 9, 10]). They prevent deontic explosion by restricting the inheritance principle

$$\text{If } \vdash A \supset B, \text{ then } \vdash \mathsf{O}A \supset \mathsf{O}B \tag{RM}$$

Note that any system that validates full aggregation, full inheritance as well as *ex contradictione quodlibet* leads to explosion when applied to conflicts of the form $\mathsf{O}A \wedge \mathsf{O}\neg A$. By aggregation $\mathsf{O}(A \wedge \neg A)$ is derivable from $\mathsf{O}A \wedge \mathsf{O}\neg A$ and, in view of *ex contradictione quodlibet*, $\mathsf{O}B$ follows from $\mathsf{O}(A \wedge \neg A)$ by (RM).

We will argue in this paper that, although Goble's **DPM** logics are sufficiently conflict-tolerant, they are suboptimal in other respects. In order to overcome this, we will present adaptive strenghtenings of the **DPM** logics. The idea behind adaptive logics (see [2, 3]) is to interpret a given premise set "as normally as possible". In our case obligations are interpreted as non-conflicting as possible. It will be demonstrated that the adaptive systems are significantly stronger than the **DPM** logics and approximate **SDL**. For instance, for premise sets that are conflict-free, the adaptive versions of the **DPM** systems lead to exactly the same consequence set as **SDL**.

Let us outline the structure of this paper. In Section 2, we introduce Goble's **DPM** systems and explain their semantics in Section 3. We show that the **DPM** systems have some shortcomings in Section 4. Motivated by the limitations of the **DPM** systems, we introduce the reader into adaptive logics in Section 5. In Sections 6 and 7 we present the adaptive strengthenings **ADPM.1** and **ADPM.2′**. We list some interesting properties of these systems in Section 8. In Section 9, we summarize the merits of the adaptive logics. We also provide a perspective on how the techniques presented in this paper may be transferred to other problems in the context of deontic logics. The Appendix features the proofs of our results.

2 Dealing with deontic conflicts by restricting inheritance

In the remainder we work with a propositional language enriched by a monadic obligation operator O. Where $\mathcal{S}$ is the set of sentential letters, our set of well-formed formulas $\mathcal{W}$ is given by the $\langle\neg, \wedge, \vee, \supset, \mathsf{O}\rangle$-closure of $\mathcal{S}$ with the usual rules for brackets. We define $A \equiv B$ by $(A \supset B) \wedge (B \supset A)$ and the permission operator $\mathsf{P}A$ by $\neg\mathsf{O}\neg A$.

The idea behind Goble's **DPM** systems is to restrict the inheritance principle via permission statements. The full inheritance principle (RM) is replaced by the following 'rule of permitted inheritance'

$$\text{If } \vdash A \supset B, \text{ then } \vdash \mathsf{P}A \supset (\mathsf{O}A \supset \mathsf{O}B) \tag{RPM}$$

What the rule (RPM) comes to is this: if A is obligatory and A entails B, then B is also obligatory *provided that* it is explicitly stated that A is permitted, or what comes to the same, that the obligation to bring about A is unconflicted.[1] Thus, $\mathsf{O}B$ follows neither from $\Gamma_1 = \{\mathsf{O}(A \wedge B), \mathsf{O}\neg(A \wedge B)\}$ nor from $\Gamma_2 = \{\mathsf{O}(A \wedge B)\}$, but it does follow from $\Gamma_3 = \{\mathsf{O}(A \wedge B), \mathsf{P}(A \wedge B)\}$.

Classical propositional logic enriched with the rules (RPM), and

$$\text{If } \vdash A \equiv B, \text{ then } \vdash \mathsf{O}A \equiv \mathsf{O}B \tag{RE}$$

and the axioms

$$\vdash \mathsf{O}\top \tag{N}$$

$$\vdash (\mathsf{O}A \wedge \mathsf{O}B) \supset \mathsf{O}(A \wedge B) \tag{AND}$$

defines the system **DPM.1**. More precisely, **DPM.1** is the least set of formulas containing all classical tautologies of formulas of $\mathcal{W}$, plus all instances of (N) and (AND), that is closed under Modus Ponens, (RE), and (RPM) with '$\vdash$' indicating membership in **DPM.1**. We define in a canonical way, $\vdash_{\mathbf{DPM.1}} A$ iff A is a member of **DPM.1**. Furthermore, where $\Gamma \subseteq \mathcal{W}$, $\Gamma \vdash_{\mathbf{DPM.1}} A$ iff for some $B_1, \ldots, B_n \in \Gamma$ we have $\vdash_{\mathbf{DPM.1}} (B_1 \wedge \cdots \wedge B_n) \supset A$.[2]

Besides **DPM.1** Goble presented another system, **DPM.2**, that also employs the restricted inheritance principle (RPM), but that moreover restricts aggregation. We have motivated the restriction of the inheritance principle and of the aggregation principle as a way to gain conflict-tolerant deontic logics. As will be stated in Theorem 1, **DPM.1** does not validate (DEX). Hence, since **DPM.1** is already a conflict-tolerant deontic logic, the question arises concerning the use of this further restriction. Let us give some reasons. First, it is not clear that aggregation should hold unrestrictedly. For instance, should aggregation be applied to conflicting obligations? Example: do we want to derive

---

[1] In view of the definition of $\mathsf{P}A$, $\mathsf{O}A \wedge \mathsf{P}A$ expresses that the obligation OA is unconflicted.

[2] See also [16] where the authors define consequence relations for rank-1 modal logics in this way and prove strong completeness.

$\mathsf{O}(A \wedge B)$ from $\{\mathsf{O}A,\ \mathsf{O}\neg A,\ \mathsf{O}B\}$? Analogously, should aggregation be applied in cases where it leads to (additional) deontic conflicts? For instance, should one allow that $\mathsf{O}(A \wedge B)$ is derivable from $\{\mathsf{O}A,\ \mathsf{O}B,\ \mathsf{O}\neg(A \wedge B)\}$, thus creating an additional conflict? A negative answer to these questions motivates the restriction of aggregation. Secondly, principle (P), $\neg\mathsf{O}\bot$, has quite some intuitive appeal. Obviously it is impossible to bring about $\bot$. The Kantian principle 'ought implies can' says that we are not obliged to bring about things that are impossible to realize. However, allowing for unrestricted aggregation in the presence of a conflict $\mathsf{O}A \wedge \mathsf{O}\neg A$ leads to $\mathsf{O}(A \wedge \neg A)$ and hence to $\mathsf{O}\bot$. Thus, adding (P) as an axiom to **DPM.1** leads to explosion when applied to deontic conflicts. This can be avoided by restricting aggregation.

Due to the fact that there are various conflict-tolerant deontic logics that *only* restrict (or abandon) aggregation, the reader may still wonder why in **DPM.2** both principles are restricted. One reason is, as Goble pointed out in his [10], that many systems that restrict (but do not abandon) aggregation are not conflict-tolerant *enough*. In his critical analysis he elaborated various refined explosion principles. Besides the very strict notion of deontic explosion that underlies (DEX), namely situations in which all obligations are derivable, there are weaker notions. Take for instance the following explosion principle:[3]

$$\text{If } \nvdash \neg B \text{ then } \mathsf{O}A, \mathsf{O}\neg A \vdash \mathsf{O}B \tag{DEX-1}$$

Another notion of deontic explosion is given if, for every B, $\mathsf{O}B \vee \mathsf{O}\neg B$ is derivable. Semantically speaking this corresponds to the case where all models are such that for every B there is either the obligation to bring about B or there is the obligation to bring about not-B. Although weaker than (DEX) it is equally counter-intuitive that $\mathsf{O}B \vee \mathsf{O}\neg B$ is derivable from $\{\mathsf{O}A \wedge \mathsf{O}\neg A\}$. Hence, we expect from conflict-tolerant deontic logics that they do not validate the following explosion principle:

$$\mathsf{O}A, \mathsf{O}\neg A \vdash \mathsf{O}B \vee \mathsf{O}\neg B \tag{DEX-2}$$

This may be weakened further. Facing a deontic conflict, $\mathsf{O}A \wedge \mathsf{O}\neg A$ as well as an unconflicted obligation $\mathsf{O}C \wedge \neg\mathsf{O}\neg C$, it would be undesired that, for every B, the formula $\mathsf{O}B \vee \mathsf{O}\neg B$ would be derivable. This is expressed as follows:

$$\mathsf{O}A, \mathsf{O}\neg A, \mathsf{O}C, \neg\mathsf{O}\neg C \vdash \mathsf{O}B \vee \mathsf{O}\neg B \tag{DEX-3}$$

Validating (DEX-3) is counter-intuitive, since for some arbitrary B the conflict $\mathsf{O}A \wedge \mathsf{O}\neg A$ together with the other, otherwise unrelated and unproblematic obligation $\mathsf{O}C$ do not entail the obligation $\mathsf{O}B$.

By restricting aggregation along with inheritance the various advantages can be combined. In this way we gain systems that follow the Kantian intuition

---

[3] We slightly adjusted the criteria (DEX-1)–(DEX-3) (the latter two will be introduced in a moment) offered by Goble since his criteria were formulated in terms of theoremhood while we focus on the consequences of premise sets.

'ought implies can', that hence validate (P), and that are strongly conflict-tolerant such that they do not validate any of the explosion principles (DEX), (DEX-1)–(DEX-3).

In order to achieve such a conflict-tolerant logic, Goble uses the following permitted aggregation principle:

$$\vdash \big(\mathsf{O}A \wedge \mathsf{O}B \wedge \mathsf{P}(A \wedge B)\big) \supset \mathsf{O}(A \wedge B) \tag{PAND}$$

The idea is to apply aggregation to $\mathsf{O}A$ and $\mathsf{O}B$, *provided that* $A \wedge B$ is explicitly permitted. Goble's logic **DPM.2** is defined by (RPM), (RE), (N), (P), and (PAND). The consequence relation $\vdash_{\mathbf{DPM.2}}$ is defined analogous to $\vdash_{\mathbf{DPM.1}}$.

There is an alternative way of restricting aggregation that offers certain advantages over (PAND), namely:

$$\vdash \big(\mathsf{O}A \wedge \mathsf{O}B \wedge \mathsf{P}A \wedge \mathsf{P}B\big) \supset \mathsf{O}(A \wedge B) \tag{PAND'}$$

Here, the idea is to apply aggregation to $\mathsf{O}A$ and $\mathsf{O}B$ provided that *both* A and B are explicitly permitted.

The logic $\mathbf{DPM.2'}$ is defined by (RPM), (RE), (N), (P), and (PAND'). The consequence relation $\vdash_{\mathbf{DPM.2'}}$ is defined analogous to $\vdash_{\mathbf{DPM.1}}$.

Henceforth we will use **DPM** as a generic term for **DPM.1**, **DPM.2** and $\mathbf{DPM.2'}$.

As announced already, **DPM** is sufficiently conflict-tolerant not to validate any of the introduced explosion principles.

Theorem 1. *Where* $\mathbf{L} \in \{\mathbf{DPM.1}, \mathbf{DPM.2}, \mathbf{DPM.2'}\}$*,* $\mathbf{L}$ *does not validate any of the explosion principles* (DEX)*,* (DEX-1)*–*(DEX-3)*.*

Lou Goble argued in [7, 9] in favor of the following criterion of adequacy for conflict-tolerant deontic logics:

> ($\star$): A conflict-tolerant deontic logic should be such that the result of adding (D), namely $\vdash \mathsf{O}A \supset \neg\mathsf{O}\neg A$, as an axiom leads to the same consequence relation as **SDL**.[4]

Theorem 2. *Where* $\alpha \in \{1, 2'\}$*,* $\mathbf{DPM}.\alpha$ *satisfies* ($\star$)*.*

The logic $\mathbf{DPM.2'}$ has several advantages as compared to Goble's **DPM.2**. First, the restricted aggregation principle of $\mathbf{DPM.2'}$, i.e. (PAND'), coheres better with the idea underlying (RPM) than (PAND). Note that the idea underlying (RPM) was to restrict inheritance to those obligations that are explicitly permitted, or what comes to the same, are explicitly unconflicted. This idea is applied to the aggregation principle by (PAND')—aggregation can only be applied if both obligations are explicitly unconflicted. In contrast, the idea underlying (PAND) is to apply aggregation to $\mathsf{O}A$ and $\mathsf{O}B$ provided that

---

[4] There are various alternative axiomatizations of **SDL**. Goble for instance uses full propositional logic, (D), (N), (RE), (RM), and (AND). A consequence relation $\vdash_{\mathbf{SDL}}$ can be defined analogous to $\vdash_{\mathbf{DPM.1}}$. We slightly adjusted Goble's ($\star$) since Goble is mainly interested in theoremhood, while we focus on consequence relations.

the *outcome* of the aggregation, $A \wedge B$, is explicitly permitted. Thus in the case of (PAND), but not in the case of (PAND'), $\mathsf{O}(A \wedge B)$ is derivable from $\{\mathsf{O}A,\ \mathsf{O}\neg A,\ \mathsf{O}B, \mathsf{P}(A \wedge B)\}$.[5] Second, **DPM.2′** satisfies Goble's criterion ($\star$) while **DPM.2** does not. Third, by choosing **DPM.2′** instead of **DPM.2** as a basis for the adaptive strengthenings that are introduced in Section 5 we will avoid certain technical problems.[6]

Before we take a look at some of the shortcomings of the **DPM** logics, let us introduce the semantics.

3 The semantics of DPM

The semantics that we introduce in this section are very similar to Goble's neighborhood semantics for his **DPM** logics in [9, 7]. The only difference is that we employ an actual world. This makes the semantics philosophically more intuitive for our application, since we are not only interested in modeling theoremhood but rather in defining a semantic consequence relation.

One of the basic ideas for the neighborhood semantics is that propositions are interpreted in terms of sets of worlds. Moreover, each world has associated with it propositions, i.e., sets of worlds. The idea is that an obligation $\mathsf{O}A$ is true at a world w, in case A is one of its associated propositions.

Let $\wp(X)$ be the power-set of some set X. A neighborhood frame F is a tuple $\langle W, \mathcal{O}\rangle$ where W is a set of points and $\mathcal{O} : W \to \wp(\wp(W))$. We call elements of W worlds. Thus, $\mathcal{O}$ assigns to each world $w \in W$ a set of propositions, i.e., $\mathcal{O}(w) \subseteq \wp(W)$. We write from now on $\mathcal{O}_w$ instead of $\mathcal{O}(w)$. An F-model M on a frame F is a triple $\langle F, v, @\rangle$ where $@ \in W$ is called the actual world and $v : \mathcal{S} \to \wp(W)$. A propositional atom is mapped by v into the set of worlds in which it is supposed to hold. Where $w \in W$ and $|A|_M =_{\mathrm{df}} \{w \in W \mid M, w \models A\}$, we define:

(M-$\mathcal{P}$) $M, w \models A$ iff $w \in v(A)$, where $A \in \mathcal{S}$
(M-$\mathcal{O}$) $M, w \models \mathsf{O}A$ iff $|A|_M \in \mathcal{O}_w$
(M-$\neg$) $M, w \models \neg A$ iff $M, w \not\models A$
(M-$\vee$) $M, w \models A \vee B$ iff $M, w \models A$ or $M, w \models B$
(M-$\wedge$) $M, w \models A \wedge B$ iff $M, w \models A$ and $M, w \models B$

---

[5] A restricted inheritance principle following the intuition of (PAND) would be: If $\vdash A \supset B$, then $\vdash \mathsf{P}B \supset (\mathsf{O}A \supset \mathsf{O}B)$. Inheritance is applied to $\mathsf{O}A$ in order to derive $\mathsf{O}B$ if it does not result in a deontic conflict $\mathsf{O}B \wedge \mathsf{O}\neg B$.

[6] In [19], van der Torre and Tan presented a sequential system which, in a first phase, disables the application of (RM) and allows for the application of a restricted aggregation rule. In a second phase, it disables this aggregation rule and allows for the application of (RM). Although this system overcomes some of the problems facing **DPM.2** and **DPM.2′** (see Section 4), it can do so only by introducing two different O-operators and by requiring that (RM) is never applied before the restricted aggregation rule. As the authors themselves admit, this is rather strange from an intuitive point of view (see also [10], pp. 470-471). The adaptive extension of **DPM.2′** (see Section 7) overcomes the problems discussed in Section 4 without having to introduce another O-operator and without requiring any order in applying its inference rules.

(M-$\supset$) $M, w \models A \supset B$ iff $M, w \models \neg A \vee B$
(M-$\top$) $M, w \models \top$
(M-$\bot$) $M, w \not\models \bot$

Furthermore, $M \models A$ iff $M, @ \models A$. Where $\Gamma \subseteq \mathcal{W}$, we say that M is an F-model of Γ iff M is an F-model and $M \models A$ for all $A \in \Gamma$.

We define the following requirements on frames $F = \langle W, \mathcal{O} \rangle$. For all $w \in W$:

a) $W \in \mathcal{O}_w$
b) If $X \in \mathcal{O}_w$ and $Y \in \mathcal{O}_w$, then $X \cap Y \in \mathcal{O}_w$
b') If $X \in \mathcal{O}_w$; $Y \in \mathcal{O}_w$; $W \setminus X \notin \mathcal{O}_w$; and $W \setminus Y \notin \mathcal{O}_w$, then $X \cap Y \in \mathcal{O}_w$
b") If $X \in \mathcal{O}_w$; $Y \in \mathcal{O}_w$; $W \setminus (X \cap Y) \notin \mathcal{O}_w$, then $X \cap Y \in \mathcal{O}_w$
c) If $X \subseteq Y$; $X \in \mathcal{O}_w$ and $W \setminus X \notin \mathcal{O}_w$, then $Y \in \mathcal{O}_w$
d) $\emptyset \notin \mathcal{O}_w$

Condition a) corresponds to (N), b) corresponds to (AND), b') corresponds to (PAND'), b") corresponds to (PAND), c) corresponds to (RPM), and d) corresponds to (P). We call the class of all frames that satisfy a), b) and c) the **DPM.1**-frames, the ones that satisfy a), b'), c) and d) the **DPM.2′**-frames, and the ones that satisfy a), b"), c) and d) the **DPM.2**-frames.

Let $\Gamma \subseteq \mathcal{W}$. A semantic consequence relation can be defined as follows. Where F is a frame, $\Gamma \Vdash_F A$ iff for all F-models M of Γ, $M \models A$. Moreover, where $\alpha \in \{1, 2, 2'\}$, $\Gamma \Vdash_{\mathbf{DPM}.\alpha} A$ iff $\Gamma \Vdash_F A$ for all **DPM**.α-frames F.

Theorem 3. *Where* $\alpha \in \{1, 2, 2'\}$ *and* $\Gamma \subseteq \mathcal{W}$*,* $\Gamma \vdash_{\mathbf{DPM}.\alpha} A$ *iff* $\Gamma \Vdash_{\mathbf{DPM}.\alpha} A$*.*

4 Some shortcomings of DPM

In order to apply the weakened inheritance principle (resp. also the weakened aggregation principle in the case of **DPM.2** and **DPM.2′**) the user has to "manually" add permission statements. For instance, in order to apply the restricted inheritance principle (RPM) to $\mathsf{O}A$ we also need $\mathsf{P}A$. In cases in which $\mathsf{P}A$ is not derivable from the premises by means of **DPM**, the user has to add manually $\mathsf{P}A$ to the premises. This is suboptimal for various reasons.

(1) For all interesting cases, determining which permission statements can safely be added to a set of premises (that is, in such a way that no explosion follows) requires *reasoning*. This kind of reasoning falls entirely outside the scope of the **DPM** systems and is therefore left to the user of the **DPM** systems. So, the **DPM** systems are inadequate to fully explicate the reasoning processes that are needed to apply the **DPM** systems in a sensible way (that is, in a way that modal inheritance is applied "as much as possible").

(2) The fact that permissions have to be added manually is especially problematic in cases where the relationship between the premises is interwoven. For instance in complicated setups it might not be obvious at all that $\mathsf{O}A \wedge \mathsf{O}\neg A$ is derivable. However, suppose that in this case the user naïvely added $\mathsf{P}A$ to the

premises in order to apply (RPM) to $\mathsf{O}A$. Since $\mathsf{P}A$ is equivalent to $\neg\mathsf{O}\neg A$ the user caused in this way an explosion.

(3) For most premise sets the **DPM** systems are rather weak. Recall that in order to achieve the deductive strength of **SDL** we had to add (D) to the axiomatization of **DPM.1** (resp. **DPM.2′**). Suppose we accept **SDL** as the normative standard for the modeling of non-conflicting obligations. It would be desirable then that conflict-tolerant logics apply all rules of **SDL** to non-conflicting obligations. For instance, given a premise set $\Gamma \subseteq \mathcal{W}$ that is conflict-free, we expect a deontic logic to lead to a consequence set that is the same as that of **SDL** without the need of strengthening the premise set manually by adding instances of (D), or by adding premises. Thus, if the premises are conflict-free, the logic should apply all the rules of **SDL** unrestrictedly.

The discussion above motivates the following strengthening of Goble's requirement ($\star$).

> ($\star\star$): For **SDL**-consistent premise sets[7] $\Gamma \subseteq \mathcal{W}$ a conflict-tolerant deontic logic should lead to the same consequence set as **SDL**.

Note that neither of the introduced **DPM** logics satisfies ($\star\star$). We will in the next section adaptively strengthen both logics so that they satisfy criterion ($\star\star$).

As should have become evident from the discussion above, there is a certain trade-off for monotonic deontic logics such as **DPM**. In order to offer conflict-tolerance certain **SDL**-principles such as (D) and inheritance have to be restricted or abandoned. In return, this weakens the logics even in cases in which it would be unproblematic to apply the principles in question.

The adaptive strengthenings introduced in the next sections will overcome this trade-off. On the one hand, applying the logic does not involve any user interference. On the other hand, by interpreting a premise set as non-conflicting as possible, principles such as inheritance, (D) and aggregation in the case of the adaptive strengthening of **DPM.2′** will be applied as much as possible.

5 Adaptive logics

The main feature of adaptive logics is that they interpret a given premise set "as normally as possible". The standard of normality depends on the application. For instance, there are inconsistency-adaptive logics that, while allowing for classical inconsistencies, interpret a given premise set as consistently as possible (see e.g. [1, 3]). In our case the idea is to interpret premise sets "as non-conflicting as possible". We will in the following give a precise meaning to this vague notion. Adaptive logics in standard format are defined as a triple:

1. the **lower limit logic**, i.e., a reflexive, transitive, monotonic, and compact logic[8] that has a characteristic semantics and contains classical logic,

---

[7] Γ is **SDL**-consistent iff $\Gamma \nvdash_{\mathbf{SDL}} \bot$.

[8] A logic $\mathbf{L}$ is reflexive iff $\Gamma \subseteq Cn_{\mathbf{L}}(\Gamma)$, it is monotonic iff $Cn_{\mathbf{L}}(\Gamma) \subseteq Cn_{\mathbf{L}}(\Gamma \cup \Gamma')$, it is transitive iff if $\Gamma' \subseteq Cn_{\mathbf{L}}(\Gamma)$ then $Cn_{\mathbf{L}}(\Gamma') \subseteq Cn_{\mathbf{L}}(\Gamma)$ and it is compact iff if $A \in Cn_{\mathbf{L}}(\Gamma)$ then there is a finite $\Gamma' \subseteq \Gamma$ such that $A \in Cn_{\mathbf{L}}(\Gamma')$.

2. the set of **abnormalities** Ω, characterized by a (possibly restricted) logical form,
3. the adaptive **strategy**, in this paper the minimal abnormality strategy.

The adaptive logic strengthens its lower limit logic. Since we are interested in interpreting premise sets as non-conflicting as possible, a minimal requirement for our lower limit is that it is conflict-tolerant. We will use **DPM.1** or **DPM.2′** for this purpose.[9] Let henceforth $\alpha \in \{1, 2'\}$. In order to avoid an unnecessary level of abstraction we introduce the reader into adaptive logics immediately by means of the systems that we are going to introduce in this paper.[10]

The adaptive strategy together with the abnormalities determine what it means to interpret the premises as normally as possible. The standard format currently is defined for two types of strategies: the reliability strategy and the minimal abnormality strategy. Choosing the latter results in a stronger consequence relation. However, the proof theory is more involving and it is computationally more complex compared to the one on the basis of the reliability strategy [20]. The reason why we prefer the minimal abnormality strategy over the reliability strategy for our application is that it allows for the derivation of $\mathsf{O}(A \vee B)$ from two incompatible obligations $\mathsf{O}A$ and $\mathsf{O}B$ while reliability does not. We will give an example of this later on (see Example 3).

Since our aim is to interpret a given premise set Γ as non-conflicting as possible, we define our abnormalities to be deontic conflicts, $\mathsf{O}A \wedge \mathsf{O}\neg A$. We denote the set of all abnormalities by Ω.

Now we have the three elements needed to define the adaptive logics that are explicated in this paper: (1) the lower limit logic is either **DPM.1** or **DPM.2′**, (2) the set of abnormalities is $\Omega = \{\mathsf{O}A \wedge \mathsf{O}\neg A \mid A \in \mathcal{W}\}$, and (3) the minimal abnormality strategy. We dub these systems **ADPM.1** and **ADPM.2′**.

Semantically the minimal abnormality strategy is realized by selecting a certain well-defined set of **DPM**.α-models of a given premise set Γ, namely the ones that are "minimally abnormal". The models are selected with respect to their abnormal part, i.e. the abnormalities they verify: where M is a **DPM**.α-model, $\mathrm{Ab}(M) = \{A \in \Omega \mid M \models A\}$. For a given logic **L**, we write $\mathcal{M}_{\mathbf{L}}(\Gamma)$ for the set of **L**-models verifying all members of Γ.

Definition 1. A **DPM**.α-model $M \in \mathcal{M}_{\mathbf{DPM}.\alpha}(\Gamma)$ is *minimally abnormal* iff there is no **DPM**.α-model $M' \in \mathcal{M}_{\mathbf{DPM}.\alpha}(\Gamma)$ such that $\mathrm{Ab}(M') \subset \mathrm{Ab}(M)$.[11] We write $\mathcal{M}_{\mathbf{ADPM}.\alpha}(\Gamma)$ for all the minimal abnormal **DPM**.α-models of Γ.

$\Gamma \Vdash_{\mathbf{ADPM}.\alpha} A$ (A is an **ADPM**.α*-semantic consequence of* Γ) iff A is verified by all $M \in \mathcal{M}_{\mathbf{ADPM}.\alpha}(\Gamma)$.

Hence all the selected models are such that they validate (in the set-theoretical sense) a minimal amount of conflicts. This justifies our claim that the adaptive

---

[9] We will discuss the case of **DPM.2** being a lower limit logic shortly in Section 7.

[10] For a more generic introduction see [2, 3].

[11] As has been shown in [2, 3], the lower limit logic models of a premise set Γ are smooth with respect to their abnormal part and set inclusion. That is to say, for every lower limit model M of Γ there is a minimally abnormal model M' of Γ such that $\mathrm{Ab}(M') \subseteq \mathrm{Ab}(M)$. This property is called strong reassurance by adaptive logicians.

logics interpret premises as non-conflicting as possible.

Let us proceed with the syntactic counter-part to the semantic selection. It is realized by dynamic adaptive proofs. While all the rules of **DPM**.α are valid, a key feature of adaptive proofs is that they allow for certain additional rules to be applied *conditionally.* In our case the idea is to apply the inheritance principle or principle (D) conditionally. Recall that **DPM**.α, in order to avoid deontic explosions, only validates a restricted version of (RM). This led to the problem (discussed in Section 4) that in many cases the user needs to add manually permission statements. Note that the following is valid in **DPM**.α:

$$\text{If } \vdash_{\mathbf{DPM}.\alpha} A \supset B, \text{ then } \mathsf{O}A \vdash_{\mathbf{DPM}.\alpha} \mathsf{O}B \vee (\mathsf{O}A \wedge \mathsf{O}\neg A).$$

The underlying idea of the restriction is that inheritance is applicable to $\mathsf{O}A$ if there is no deontic conflict concerning $\mathsf{O}A$. In the adaptive logic we make use of this: the inheritance principle is applied to $\mathsf{O}A$ on the condition $\{\mathsf{O}A \wedge \mathsf{O}\neg A\}$. That is to say, on the condition that there is no reason to suppose that there is a deontic conflict concerning $\mathsf{O}A$. This is still very vague, but we will make it more precise in a moment. Suppose $\mathsf{O}A$ is one of the premises. A fragment of an adaptive proof may look as follows:

1	$\mathsf{O}A$	PREM	$\emptyset$
2	$\mathsf{O}(A \vee B)$	1; RC	$\{\mathsf{O}A \wedge \mathsf{O}\neg A\}$

Line 2 contains a conditional application of the inheritance rule to $\mathsf{O}A$. This is indicated by the generic rule RC (that is discussed in more detail later) and by a fourth column in which the conditions of the respective lines are contained. Conditions are finite sets of abnormalities. The condition of line 1 is empty since it is the result of a premise introduction. The condition of line 2 is $\{\mathsf{O}A \wedge \mathsf{O}\neg A\}$. Now suppose we are able to derive the following disjunction of abnormalities from the given premises at a line l:

$$l \quad (\mathsf{O}A \wedge \mathsf{O}\neg A) \vee (\mathsf{O}C \wedge \mathsf{O}\neg C) \qquad \ldots; \text{RU} \quad \emptyset$$

By the generic rule RU we indicate the (unconditional) applications of the **DPM**.α rules. Note that the disjunction of abnormalities that has been derived at line l also features the condition of line 2. This gives us reasons to suspect that $\mathsf{O}A$ may after all be part of a deontic conflict. In this case line 2 is marked according to the marking definition. We will define the marking for the minimal abnormality strategy in a moment. Formulas that are the second element of marked lines are not considered to be consequences of the adaptive logic at that stage. Note however that markings may come and go. Assume for instance that $\mathsf{O}C \wedge \mathsf{O}\neg C$ has been derived at a later stage of the proof. In this case there is no reason anymore to suspect that $\mathsf{O}A$ is part of a deontic conflict and the application of inheritance at line 2 can again be considered as valid. The marking will thus be defined on basis of the minimal disjunctions of abnormalities that have been derived at a given stage of the proof.

We give now a precise account of the generic rules RU and RC and the marking definition.

We have already indicated that the rules of the lower limit logic $\mathbf{DPM}.\alpha$ are unrestrictedly valid in the adaptive logic. This is characterized by the generic rule RU:

$$\text{RU} \quad \text{If } A_1, \ldots, A_n \vdash_{\mathbf{DPM}.\alpha} B : \quad \begin{array}{ll} A_1 & \Delta_1 \\ \vdots & \vdots \\ A_n & \Delta_n \\ \hline B & \Delta_1 \cup \cdots \cup \Delta_n \end{array}$$

Note that the conditions of the used lines featuring the A_i's are carried forward, resulting in B being derived on the condition $\Delta_1 \cup \cdots \cup \Delta_n$. Moreover, premises are introduced by PREM on the empty condition.

The conditional rule allows for conditional applications of certain rules that are invalid in **DPM**. As is usually done by adaptive logic scholars we use the notation $\mathsf{Dab}(\Delta) =_{\text{df}} \bigvee_{A \in \Delta} A$ where Δ is a finite set of abnormalities ($\Delta \subset \Omega$). Conditional applications of rules are handled by the generic rule RC:

$$\text{RC} \quad \text{If } A_1, \ldots, A_n \vdash_{\mathbf{DPM}.\alpha} B \vee \mathsf{Dab}(\Theta) : \quad \begin{array}{ll} A_1 & \Delta_1 \\ \vdots & \vdots \\ A_n & \Delta_n \\ \hline B & \Delta_1 \cup \cdots \cup \Delta_n \cup \Theta \end{array}$$

In addition to the conditional rule, we also need a *marking definition.* This determines when a condition of a line is violated and hence has to be marked. The marking indicates that the formula that occurs on that line is no longer considered as derived in the proof. In order to formulate marking definition, we need a few more notions.

A formula $\mathsf{Dab}(\Delta)$ is a *minimal* Dab*-formula* at a stage s of the proof iff it is the formula of a line with condition $\emptyset$ and no $\mathsf{Dab}(\Delta')$ with $\Delta' \subset \Delta$ is the formula of a line with condition $\emptyset$. A *choice set* of $\Sigma = \{\Delta_1, \Delta_2, \ldots\}$ is a set that contains an element out of each member of Σ. A *minimal choice set* of Σ is a choice set of Σ of which no proper subset is a choice set of Σ. Where $\mathsf{Dab}(\Delta_1), \ldots, \mathsf{Dab}(\Delta_n)$ are the minimal Dab-formulas that are derived on condition $\emptyset$ at stage s, $\Phi_s(\Gamma)$ is the set of minimal choice sets of $\{\Delta_1, \ldots, \Delta_n\}$ for a premise set Γ.

Definition 2 (Marking for minimal abnormality)**.** Line i is marked at stage s iff, where A is the second element and Δ is the condition of line i,

(i) there is no $\Delta' \in \Phi_s(\Gamma)$ such that $\Delta' \cap \Delta = \emptyset$, or

(ii) for some $\Delta' \in \Phi_s(\Gamma)$, there is no line at which A is derived on a condition Θ for which $\Delta' \cap \Theta = \emptyset$.

Given a set of abnormalities Ω, the marking definition determines whether lines are “in” or “out” of the proof at a certain stage, i.e., it governs the internal dynamics of the proof procedure.

In our introductory example it was illustrated that markings in an adaptive proof may come and go. While line 2 is marked as long as $(\mathsf{O}A \wedge \mathsf{O}\neg A) \vee (\mathsf{O}C \wedge \mathsf{O}\neg C)$ is a minimal Dab-formula, it is unmarked as soon as $\mathsf{O}C \wedge \mathsf{O}\neg C$ is derived.

In order to define the consequence set of an adaptive logic we are interested in a stable criterion for derivability.

Definition 3. A is *finally derived* from Γ on line i of a proof at stage s iff (i) A is the second element of line i, (ii) line i is not marked at stage s and (iii) every extension of the proof in which line i is marked may be further extended in such a way that line i is unmarked.

$\Gamma \vdash_{\mathbf{AL}} A$ (A is *finally* **ADPM**.α*-derivable* from Γ) iff A is finally derived on a line of a proof from Γ.

Let us state a central representational result for adaptive logics in standard format proven in [2]:

Theorem 4. *Where* $\Gamma \subseteq \mathcal{W}$, $\Gamma \vdash_{\mathbf{ADPM}.\alpha} A$ *iff* $\Gamma \Vdash_{\mathbf{ADPM}.\alpha} A$.

Besides the lower limit logic there is also an upper limit logic for each adaptive logic. It is the strengthening **UDPM**.α of **DPM**.α that trivializes abnormalities. It is defined by, $\Gamma \vdash_{\mathbf{UDPM}.\alpha} A$ iff $\Gamma \cup \{\neg B \mid B \in \Omega\} \vdash_{\mathbf{DPM}.\alpha} A$.

We will show that for the adaptive systems introduced in this paper the upper limit **UDPM**.α is **SDL**. For a proof of the following result see Theorem 12 in [2] (note that there premise sets that are consistent with respect to the upper limit logic are called "normal").

Theorem 5. *For* **UDPM**.α*-consistent premise sets* Γ, **ADPM**.α *leads to the same consequence set as* **UDPM**.α.

In the remainder we will sometimes speak about adding an axiom to **ADPM**.α. What we mean is the adaptive logic that is the result of adding the axiom to the lower limit logic, i.e., $\langle$**DPM**.$\alpha*, \Omega$, minimal abnormality$\rangle$ where **DPM**.$\alpha*$ is **DPM**.α strengthened by the axiom.

6 The adaptive logic ADPM.1

In this section we introduce a concrete adaptive system on the basis of Lou Goble's **DPM.1**. As discussed above, **ADPM.1** is defined by the triple

$$\langle \mathbf{DPM.1}, \Omega, \text{minimal abnormality}\rangle.$$

We have already mentioned the basic idea. In order to interpret a given premise set as non-conflicting as possible we define the abnormalities to be deontic conflicts, $\Omega = \{\mathsf{O}A \wedge \mathsf{O}\neg A \mid A \in \mathcal{W}\}$. This makes it possible to apply the inheritance principle to $\mathsf{O}A$ on the condition $\{\mathsf{O}A \wedge \mathsf{O}\neg A\}$ by the generic rule RC. Moreover, also (D) may be applied conditionally. Note that the following is valid:

$$\mathsf{O}A \vdash_{\mathbf{DPM}} \mathsf{P}A \vee (\mathsf{O}A \wedge \mathsf{O}\neg A)$$

This allows to derive $\mathsf{P}A$ from $\mathsf{O}A$ on the condition $\{\mathsf{O}A \wedge \mathsf{O}\neg A\}$ by the generic rule RC.

Example 1. Let us take a look at a first concrete example of a proof in **ADPM.1**:

1	$\mathsf{O}A$	PREM	$\emptyset$
2	$\mathsf{O}\neg A$	PREM	$\emptyset$
3	$\mathsf{O}(B \wedge C)$	PREM	$\emptyset$
4	$\mathsf{O}B$	3; RC	$\{\mathsf{O}(B \wedge C) \wedge \mathsf{O}\neg(B \wedge C)\}$
5	$\mathsf{P}B$	4; RC	$\{\mathsf{O}(B \wedge C) \wedge \mathsf{O}\neg(B \wedge C), \mathsf{O}B \wedge \mathsf{O}\neg B\}$

Note that $\mathsf{O}A, \mathsf{O}\neg A, \mathsf{O}(B \wedge C) \nvdash_{\mathbf{DPM.1}} \mathsf{O}B$ and $\mathsf{O}A, \mathsf{O}\neg A, \mathsf{O}(B \wedge C) \nvdash_{\mathbf{DPM.1}} \mathsf{P}B$. In Section 4 we have pointed out that this is suboptimal. A deontic logic should apply the rules of **SDL** to non-conflicting parts of the premise set and there should be no need for the user to add permission statements. In order to derive $\mathsf{O}B$ from $\mathsf{O}(B \wedge C)$ the user of **DPM.1** would have to manually add $\mathsf{P}(B \wedge C)$ to the premises. Moreover, there is no way of deriving $\mathsf{P}B$ from the given premises in **DPM.1**.

In contrast, the adaptive logic **ADPM.1** applies inheritance conditionally to $\mathsf{O}(B \wedge C)$ in order to derive $\mathsf{O}B$ at line 4. Moreover, (D) is applied conditionally to $\mathsf{O}B$ in order to derive $\mathsf{P}B$ at line 5. Note that $\mathsf{O}(B \wedge C)$ and $\mathsf{O}B$ are unrelated to the deontic conflict $\mathsf{O}A \wedge \mathsf{O}\neg A$, and hence, as discussed in Section 4, inheritance and (D) should be applied to them. It can easily be seen that lines 4 and 5 are finally derived.

Example 2. The following example of a **ADPM.1**-proof features a more complex setup. We use the abbreviation $!A$ for $\mathsf{O}A \wedge \mathsf{O}\neg A$.

1	$\mathsf{O}A$	PREM	$\emptyset$
2	$\mathsf{O}B$	PREM	$\emptyset$
3	$\mathsf{O}(C \vee D)$	PREM	$\emptyset$
4	$\mathsf{O}\neg(A \wedge C)$	PREM	$\emptyset$
5	$\mathsf{O}\neg(B \wedge D)$	PREM	$\emptyset$
6	$\mathsf{O}(A \wedge \neg C)$	1,4; RU	$\emptyset$
7	$\mathsf{O}(B \wedge \neg D)$	2,5; RU	$\emptyset$
^{16}8	$\mathsf{P}A \wedge \mathsf{P}B \wedge \mathsf{P}(C \vee D)$	1,2,3; RC	$\{!A, !B, !(C \vee D)\}$
9	$\mathsf{O}(A \vee \neg C) \vee !A$	1; RU	$\emptyset$
10	$\mathsf{O}(B \vee \neg D) \vee !B$	2; RU	$\emptyset$
11	$\mathsf{O}\neg C \vee !A$	4,9;RU	$\emptyset$
^{16}12	$\mathsf{O}\neg C$	11; RC	$\{!A\}$
13	$\mathsf{O}\neg D \vee !B$	5,10; RU	$\emptyset$
^{16}14	$\mathsf{O}\neg D$	13; RC	$\{!B\}$
15	$\mathsf{O}\neg(C \vee D) \vee !A \vee !B$	11,13; RU	$\emptyset$
16	$!A \vee !B \vee !(C \vee D)$	3,15; RU	$\emptyset$
17	$\neg\mathsf{P}A \vee \neg\mathsf{P}B \vee \neg\mathsf{P}(C \vee D)$	16; RU	$\emptyset$

Without engaging in more advanced reasoning processes it is for a user of **DPM.1** in no way clear which permission statements may be added without causing deontic conflicts or explosion. We have pointed out this problem in

Section 4 (point 2). Would (s)he, for instance, add $\mathsf{P}A$, $\mathsf{P}B$, and $\mathsf{P}(C \vee D)$ it would cause explosion, since via **DPM.1**, $\neg\mathsf{P}A \vee \neg\mathsf{P}B \vee \neg\mathsf{P}(C \vee D)$ is derivable at line 17. Note that in the adaptive proof line 8 is marked in view of line 16 and does therefore not cause any harm. Thus, the adaptive logic identifies given deontic conflicts and blocks undesired consequences from them. For instance, the unintuitive derivations of $\mathsf{O}\neg C$ and $\mathsf{O}\neg D$ at lines 12 and 14 are marked.

7 The adaptive logic ADPM.2′

We have already pointed out various advantages of our **DPM.2′** logic over Goble's **DPM.2**. Besides these points **DPM.2′** is also more apt as a lower limit logic.

One idea to define an adaptive logic on the basis of **DPM.2** would be in terms of the triple ⟨**DPM.2**, Ω, minimal abnormality⟩. However, this has a severe shortcoming. Suppose our premises are $\Gamma = \{\mathsf{O}A, \mathsf{O}B\}$. Since these premises do not give rise to any deontic conflicts we expect from the logic to apply aggregation to them. However, $\Gamma \nvdash_{\mathbf{ADPM.2}} \mathsf{O}(A \wedge B)$. Since **DPM.2** also restricts aggregation beside inheritance it is desirable that the logic is able to apply aggregation conditionally in a similar way as **ADPM.1** applies inheritance and (D) conditionally. The way aggregation is restricted in **DPM.2**, namely by (PAND), does not allow for utilizing the same set of abnormalities as for **ADPM.1**. A way around this problem is to define the abnormalities in a different way, for instance by $\Omega' = \{\mathsf{O}A \wedge \mathsf{O}B \wedge \mathsf{O}\neg(A \wedge B) \mid A, B \in \mathcal{W}\}$. This would allow to apply the aggregation principle conditionally. However, this logic is very weak. For instance given two incompatible obligations $\mathsf{O}A, \mathsf{O}B$, where $\mathsf{O}\neg(A \wedge B)$ expresses their incompatibility, we are not able to derive $\mathsf{O}(A \vee B)$. The upshot is that, with **DPM.2** as the lower limit logic, we were not able to define a set of abnormalities in such a way that the resulting adaptive logic is sufficiently strong and aggregation is conditionally applicable.

The situation is different if we employ **DPM.2′** as lower limit logic. In that case, we can use the set of abnormalities Ω of **ADPM.1**. This contributes to a more unifying adaptive framework. Let **ADPM.2′** be defined by the triple ⟨**DPM.2′**, Ω, minimal abnormality⟩. Note that in **DPM.2′** the following is a consequence of (PAND'):

$$\vdash_{\mathbf{DPM.2'}} \big(\mathsf{O}A \wedge \mathsf{O}B\big) \supset \big(\mathsf{O}(A \wedge B) \vee \big((\mathsf{O}A \wedge \mathsf{O}\neg A) \vee (\mathsf{O}B \wedge \mathsf{O}\neg B)\big)\big)$$

This makes it possible to apply aggregation to $\mathsf{O}A \wedge \mathsf{O}B$ on the condition $\{\mathsf{O}A \wedge \mathsf{O}\neg A, \mathsf{O}B \wedge \mathsf{O}\neg B\}$ by the rule RC. Again, inheritance and (D) are applied conditionally as demonstrated already for **ADPM.1**.

In order to demonstrate the modus operandi of **ADPM.2′** we take a look a concrete example.

Example 3. Let us come back to Sartre's unfortunate student. On the one hand, he has the obligation to stay with his sick mother ($\mathsf{O}M$). However, on the other hand he has the obligation to fight at the front against Nazi Germany

($\mathsf{O}F$). Due to the lack of alethic modalities we have to find a way to model the fact that both obligations are not mutually realizable. This can be realized by $\mathsf{O}\neg(M \wedge F)$. We add another premise which is independent of the first two: as a good citizen he is obliged to pay taxes and to vote, $\mathsf{O}(T \wedge V)$. In the following proof for **ADPM.2′** we use the abbreviation $!A$ for $\mathsf{O}A \wedge \mathsf{O}\neg A$.[12]

1	$\mathsf{O}M$	PREM	$\emptyset$
2	$\mathsf{O}F$	PREM	$\emptyset$
3	$\mathsf{O}\neg(M \wedge F)$	PREM	$\emptyset$
4	$\mathsf{O}(T \wedge V)$	PREM	$\emptyset$
5	$\mathsf{O}T$	4; RC	$\{!(T \wedge V)\}$
^{7}6	$\mathsf{O}(M \wedge F)$	1,2; RC	$\{!M, !F\}$
7	$!(M \wedge F) \vee !M \vee !F$	1,2,3; RU	$\emptyset$
8	$\mathsf{O}(M \vee F)$	1; RC	$\{!M\}$
9	$\mathsf{O}(M \vee F)$	2; RC	$\{!F\}$

As discussed in Section 4, we expect that all rules of **SDL** can be applied to non-conflicting obligations such as $\mathsf{O}(T \wedge V)$. Indeed, after having introduced the premises at lines 1–4, inheritance is applied to $\mathsf{O}(T \wedge V)$ in order to derive $\mathsf{O}T$, his duty to pay taxes. Furthermore, the application of aggregation to $\mathsf{O}M$ and $\mathsf{O}F$ at line 6 gets marked. This accords with the fact that (PAND') is a rule that isolates deontic conflicts: i.e. it is not applicable to conflicting obligations.

ADPM.2′ realizes our design requirements. On the one hand, it blocks rules from being applied to conflicting obligations (such as aggregation at line 6). On the other hand it allows for applications of rules to non-conflicting premises (such as inheritance at line 5 to $\mathsf{O}(T \wedge V)$) without requiring the manual addition of permission statements.

Moreover, note that it is desired to derive $\mathsf{O}(M \vee F)$, the student's obligation to either stay with his mother or to fight the Nazis. Indeed, since $\mathsf{O}(M \vee F)$ is derivable on both conditions, $\{\mathsf{O}M \wedge \mathsf{O}\neg M\}$ and $\{\mathsf{O}F \wedge \mathsf{O}\neg F\}$, lines 8 and 9 are not marked. Similarly for instance $\mathsf{O}(T \wedge (M \vee F))$ and $\mathsf{O}(V \wedge (M \vee F))$ are derivable.

In contrast, using one of the **DPM** logics, the user would have to add manually $\mathsf{P}(T \wedge V)$ to the premises in order to derive $\mathsf{O}T$. Furthermore, it is unclear how to derive $\mathsf{O}(M \vee F)$ in **DPM**. Only by adding either $\mathsf{P}M$ or $\mathsf{P}F$ permitted inheritance is applicable to $\mathsf{O}M$ or resp. $\mathsf{O}F$. However, there is no reason to prefer $\mathsf{P}M$ over $\mathsf{P}F$ or vice versa. Moreover, would the user add both of them, it may lead to explosion. Consider the case where also $\mathsf{P}\neg(M \wedge F)$ is a premise. It can be argued that, since M and F cannot be mutually realized an agent is allowed either to bring about not-M or to bring about not-F. However, in view of the final derivability of the formula at line 7, this leads to the derivability of $(\mathsf{O}M \wedge \mathsf{O}\neg M) \vee (\mathsf{O}F \wedge \mathsf{O}\neg F)$. In case the user would have added both, $\mathsf{P}M$ and $\mathsf{P}F$, explosion would result.

---

[12] The proofs of $\mathsf{O}(M \vee F)$ and $\mathsf{O}T$ for **ADPM.1** are left to the reader.

8 Some nice properties

This section features some nice meta-theoretic properties of the introduced logics. The meta-theory of adaptive logics in standard format equips our systems immediately with soundness and completeness (see Theorem 4). As desired, the adaptive strengthenings are conflict-tolerant.

Theorem 6. *Where* $\alpha \in \{1, 2'\}$*, none of Goble's explosion principles* (DEX), (DEX-1)–(DEX-3) *is valid in* **ADPM**.α*.*

Theorem 7. *Where* $\alpha \in \{1, 2'\}$*, the upper limit logic of* **ADPM**.α *is* **SDL***.*

Corollary 1. *Where* $\alpha \in \{1, 2'\}$*,* **ADPM**.α *satisfies* $(\star)$*.*

We introduced another, in a sense more demanding criterion, $(\star\star)$. For **SDL**-consistent premise sets the given logics should have the same derivative power as **SDL**. This criterion is not fulfilled by the **DPM** logics. However, as the following Corollary shows, it is fulfilled by our adaptive strengthenings. The Corollary is a direct consequence of Theorem 5 and Theorem 7.

Corollary 2. *Where* $\alpha \in \{1, 2'\}$*,* **ADPM**.α *satisfies* $(\star\star)$*.*

9 Outlook and Conclusion

In this paper we introduced the adaptive strengthenings **ADPM**.**1** and **ADPM**.**2**$'$ of Goble's conflict-tolerant logic **DPM**.**1** and of our variant **DPM**.**2**$'$ of Goble's **DPM**.**2**. We have demonstrated various advantages of the adaptive systems.

– **ADPM**.α (where henceforth $\alpha \in \{1, 2'\}$) is significantly stronger than **DPM**.α. It is not just the case that adding (D) to the logic leads to equivalent systems to **SDL**. For any **SDL**-consistent premise set, **ADPM**.α proves to be equivalent to **SDL**. Moreover, **ADPM**.α applies restricted inheritance "as much as possible". In contrast, in Goble's system many permission statements have to be added by the user in order to apply the inheritance principle. The needed permission statements are generated in the adaptive systems automatically. This brings us to another point.

– The adaptive systems **ADPM**.α have the design virtue that the logics model all the reasoning for the user. In contrast, in **DPM**.α the user not just has to interfere in order to derive as much consequences as possible (by adding permission statements). Moreover, finding out which permission statements are harmless and may be added to a given premise set involves advanced reasoning by the user. This is especially the case for complicated setups.

– The meta-theory of adaptive logics in standard format is well-established. Many key-features do not have to be proven (anymore) for the **ADPM**.α logics since they have been shown to be valid for all adaptive logics in standard format. For instance, the completeness and soundness of **ADPM**.α follows immediately from the completeness and soundness of **DPM**.α.

– The dynamic adaptive proofs mirror actual reasoning processes. While the insight in a given premise set grows, some lines of the proof may get marked, others unmarked due to the fact that their conditions turn out to be (not) trustworthy. Furthermore, the adaptive proofs are able to deal with new information in the form of new premises and thus to handle the defeasibility that comes with it.

While we presented in this paper ways to apply inheritance, (D), and aggregation conditionally by adaptive logics, the approach may be fruitfully applied in other contexts of deontic logics. Let us give some examples.

– Lou Goble presented in [9] conditional versions of his **DPM** systems. Also there, he uses a restricted version of the inheritance principle. The adaptive handling of inheritance and aggregation introduced in this paper can be applied in the conditional context straightforwardly.

– One problem for conditional deontic logics is related to the detachment of conditional obligations. Given a conditional obligation to bring about A if B is the case, written $\mathsf{O}(A \mid B)$, and given that the condition B is fulfilled, one may want to derive the 'actual obligation' to bring about A. However, detachment is not universally valid. Note that being served a meal and given that the meal is asparagus, we do not want to detach the obligation not to eat with our fingers, although its condition, that a meal is served, is fulfilled. This is due to the fact that being served asparagus we are in exceptional circumstances. Again, applying adaptively detachment to $\mathsf{O}(A \mid B)$ and B "as much as possible" may lead to interesting solutions to this problem. Furthermore, the semantics with an actual world introduced in this paper can easily be generalized to the conditional case. This way factual premises can be handled semantically. Similarly an adaptive approach can be used in order to apply strengthening the antecedent (if $\mathsf{O}(A \mid B)$, then $\mathsf{O}(A \mid B \wedge C)$) "as much as possible", i.e., to apply it whenever the factual premises do no describe an exceptional situation (see [18]).

Altogether, the work presented in this paper is inspiring and transferable to the tackling of other important problems within the context of deontic logics. The authors are going to investigate in these fields in future research.

APPENDIX

In order to prove soundness and completeness with respect to our semantics for **DPM**, we will show that it is equivalent to Goble's original **DPM** semantics. Since the authors in [16] have proven soundness and strong completeness for Goble's semantics this is sufficient.

Goble's original neighborhood semantics is very similar to the one presented here: the key difference is that we employ an actual world. Where frames are defined as before, an F-G-model M is a pair $\langle F, v \rangle$ where F is a frame and $v : \mathcal{S} \to \wp(W)$ as before. The essential difference concerns the definition of model-validity. While in our semantics it is defined in terms of validity with respect to the actual world, in Goble's semantics it is defined in terms of validity with respect to all given worlds: $M \models^G A$ iff $M, w \models A$ for all $w \in W$. All other definitions concerning validity are analogous. For a given frame $F = \langle W, \mathcal{O} \rangle$ and $\Gamma \subseteq \mathcal{W}$, $\Gamma \Vdash_F^G A$ iff for all F-G-models M and for all $w \in W$, if $M, w \models B$ for all $B \in \Gamma$, then $M, w \models A$. Moreover, where $\alpha \in \{1, 2'\}$, $\Gamma \Vdash^G_{\mathbf{DPM}.\alpha} A$ iff $\Gamma \Vdash_F^G A$ for all $\mathbf{DPM}.\alpha$-frames F. Schröder and Pattinson have shown the following strong completeness and soundness result in [16]:

Theorem 8. *Where* $\alpha \in \{1, 2'\}$ *and* $\Gamma \subseteq \mathcal{W}$, $\Gamma \Vdash^{G}_{\mathbf{DPM}.\alpha} A$ *iff* $\Gamma \vdash_{\mathbf{DPM}.\alpha} A$.

Theorem 9. *Where* $\alpha \in \{1, 2'\}$ *and* $\Gamma \subseteq \mathcal{W}$, $\Gamma \Vdash^{G}_{\mathbf{DPM}.\alpha} A$ *iff* $\Gamma \Vdash_{\mathbf{DPM}.\alpha} A$.

Proof. Let $\mathcal{F}$ be the class of $\mathbf{DPM}.\alpha$-frames. "$\Leftarrow$": Let $\Gamma \Vdash_{\mathbf{DPM}.\alpha} A$ and $F = \langle W, \mathcal{O}\rangle \in \mathcal{F}$. Suppose there is an F-G-model $M = \langle F, v\rangle$ and a world $w \in W$ for which $M, w \not\models A$ and $M, w \models B$ for all $B \in \Gamma$. Note that $M' = \langle F, v, w\rangle$ is a $\mathbf{DPM}.\alpha$-model of Γ for which $M' \nVdash A$,—a contradiction.

"$\Rightarrow$": Let $\Gamma \Vdash^{G}_{\mathbf{DPM}.\alpha} A$. Suppose for some frame $F = \langle W, \mathcal{O}\rangle \in \mathcal{F}$ there is an F-model $M_w = \langle F, v, w\rangle$ of Γ for which $M_w \not\models A$. Let $M = \langle F, v\rangle$. Note that $M, w \models B$ for all $B \in \Gamma$ and $M, w \nVdash A$,—a contradiction. ☐

Proof of Theorem 3 (Soundness and Completeness). Follows immediately by Theorem 8 and Theorem 9. ☐

Proof of Theorem 2. For **DPM.1** this has already been shown by Goble in [9]. Note that (D) together with (PAND) results in (AND). Since **DPM.1** strengthened by (D) has the same corresponding consequence relation as **SDL**, it also validates (N). Thus, **DPM.1** strengthened by (D) and **DPM.2′** strengthened by (D) have the same corresponding consequence relation. Thus, **DPM.2′** strengthened by (D) has the same corresponding consequence relation as **SDL**. ☐

Proof of Theorem 6. Let us first consider the case for **ADPM.2′**. Let $W = \wp(\mathcal{S})$ and p_1 and p_2 are sentential letters. We define $W_a = \{w \in W \mid p_1 \notin w, p_2 \notin w\}$, $W_b = \{w \in W \mid p_1 \notin w, p_2 \in w\}$, $W_c = \{w \in W \mid p_1 \in w, p_2 \notin w\}$, and $W_d = \{w \in W \mid p_1 \in w, p_2 \in w\}$. We define a frame $F = \langle W, \mathcal{O}\rangle$ where $\mathcal{O}_w = \{W_a \cup W_b, W_c \cup W_d, W\}$ for all $w \in W$. Note that F is a **DPM.2′**-frame. Let $M = \langle F, v, @\rangle$ where $v(p_i) = \{w \in W \mid p_i \in w\}$ and @ is any world in W. Note first that $M \models \mathsf{O}p_1, \mathsf{O}\neg p_1, \mathsf{O}\top, \mathsf{P}\top, \mathsf{P}(p_1 \wedge p_2)$ and $M \not\models \mathsf{O}(p_1 \wedge p_2)$. Thus, M models a counter-instance to (DEX), (DEX-1)–(DEX-3). Note furthermore that M is a minimally abnormal model of $\{\mathsf{O}p_1, \mathsf{O}\neg p_1\}$ (and thus also of $\{\mathsf{O}p_1, \mathsf{O}\neg p_1, \mathsf{P}\top\}$ and $\{\mathsf{O}p_1, \mathsf{O}\neg p_1, \mathsf{O}\top, \mathsf{P}\top, \mathsf{P}(p_1 \wedge p_2)\}$) since $\mathrm{Ab}(M) = \{\mathsf{O}A \wedge \mathsf{O}\neg A \mid\ \vdash A \equiv p_1\}$ and $\{\mathsf{O}p_1, \mathsf{O}\neg p_1\} \vdash_{\mathbf{DPM}.\mathbf{2'}} A$ for all $A \in \mathrm{Ab}(M)$.

The proof for **ADPM.1** is similar. Where $W = \wp(\mathcal{S})$ and $v : \mathcal{S} \to \wp(W)$, $p_i \mapsto \{w \in W \mid p_i \in w\}$, we define a frame $F = \langle W, \mathcal{O}\rangle$ where for all $w \in W$, $\mathcal{O}_w = \{W' \mid W' \supseteq v(p_1)\} \cup \{\emptyset\}$. Note that F is a **DPM.1**-frame. Let $M = \langle F, v, @\rangle$ where @ is any world in W. Evidently, $M \models \mathsf{O}p_1, \mathsf{O}\top, \mathsf{O}\bot, \mathsf{P}p_1, \mathsf{P}p_2$ and $M \not\models \mathsf{O}p_2$. Thus, M models a counter-instance to (DEX), (DEX-1)—(DEX-3). Also, M is a minimally abnormal model of $\{\mathsf{O}\top, \mathsf{O}\bot\}$ (and hence also of $\{\mathsf{O}\top, \mathsf{O}\bot, \mathsf{P}p_2\}$ and of $\{\mathsf{O}\top, \mathsf{O}\bot, \mathsf{O}p_1, \mathsf{P}p_1, \mathsf{P}p_2\}$) since $\mathrm{Ab}(M) = \{\mathsf{O}A \wedge \mathsf{O}\neg A \mid\ \vdash A \equiv \top\}$ and $\{\mathsf{O}\top, \mathsf{O}\bot\} \vdash_{\mathbf{DPM}.\mathbf{1}} A$ for all $A \in \mathrm{Ab}(M)$. ☐

Proof of Theorem 1. Due to the fact that all adaptively selected models are models of the lower limit logic, the counter-models to (DEX) and (DEX-1)—(DEX-3) constructed for **ADPM.1** and **ADPM.2′** in the proof of Theorem 6 are also counter-models for **DPM.1** and **DPM.2′**. Theorem 1 was proven for **DPM.2** by Goble in [9]. ☐

Proof of Theorem 7. Given a $\Gamma \subseteq \mathcal{W}$ we have to show that $\Gamma \cup \{\neg(\mathsf{O}A \wedge \mathsf{O}\neg A) \mid A \in \mathcal{W}\} \vdash_{\mathbf{DPM}.\alpha} B$ iff $\Gamma \vdash_{\mathbf{SDL}} B$. Note that $\vdash_{\mathbf{DPM}.\alpha} \neg(\mathsf{O}A \wedge \mathsf{O}\neg A) \equiv (\mathsf{O}A \supset \mathsf{P}A)$. Thus, $\Gamma \cup \{\neg(\mathsf{O}A \wedge \mathsf{O}\neg A)\} \vdash_{\mathbf{DPM}.\alpha} B$ iff $\Gamma \vdash_{\mathbf{DDPM}.\alpha} B$ where **DDPM**.α is **DPM**.α enriched by (D). However, since by Theorem 2 **DDPM**.α has the same corresponding consequence relation as **SDL**, we are finished. ☐

Proof of Corollary 1. For **SDL**-consistent premise sets Γ this is an immediate consequence of Corollary 2. Let $\Gamma \subseteq \mathcal{W}$ be **SDL**-inconsistent. Where **DADPM**.α (resp. **DDPM**.α) is **ADPM**.α (resp. **DPM**.α) enriched by (D) and $\Gamma' = \Gamma \cup \{\mathsf{O}A \supset \mathsf{P}A \mid A \in \mathcal{W}\}$, note that $\mathcal{M}_{\mathbf{SDL}}(\Gamma) = \emptyset = \mathcal{M}_{\mathbf{DDPM}.\alpha}(\Gamma) = \mathcal{M}_{\mathbf{DPM}.\alpha}(\Gamma') \supseteq \mathcal{M}_{\mathbf{ADPM}.\alpha}(\Gamma') = \mathcal{M}_{\mathbf{DADPM}.\alpha}(\Gamma)$. $\square$

References

[1] Diderik Batens. Towards the unification of inconsistency handling mechanisms. *Logic and Logical Philosophy*, 8:5–31, 2000. Appeared 2002.

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

[3] Diderik Batens. *Adaptive Logics and Dynamic Proofs. Mastering the Dynamics of Reasoning, with Special Attention to Handling Inconsistency.* Forthcoming.

[4] Mathieu Beirlaen, Joke Meheus, and Christian Straßer. A paraconsistent adaptive deontic logic for normative conflicts. Forthcoming.

[5] Newton Da Costa and Walter Carnielli. On paraconsistent deontic logic. *Philosophia*, 16(3):293–305, December 1986.

[6] Bas C. Van Fraassen. Values and the heart's command. *Journal of Philosophy*, 70(1):5–19, 1973.

[7] Lou Goble. Dilemmas in deontic logic. To appear.

[8] Lou Goble. Multiplex semantics for deontic logic. *Nordic Journal of Philosophical Logic*, 5:113–134, 2000.

[9] Lou Goble. A proposal for dealing with deontic dilemmas. In Alessio Lomuscio and Donald Nute, editors, *DEON*, volume 3065 of *Lecture Notes in Computer Science*, pages 74–113. Springer, 2004.

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

[11] Christopher W. Gowans, editor. *Moral Dilemmas.* Oxford University Press, New York, 1987.

[12] John F. Horty. Reasoning with moral conflicts. *Noûs*, 37(4):557–605, 2003.

[13] Joke Meheus, Mathieu Beirlaen, and Frederik Van De Putte. Avoiding deontic explosion by contextually restricting aggregation. In Guido Governatori and Giovanni Sartor, editors, *DEON (10th International Conference on Deontic Logic in Computer Science)*, volume 6181 of *Lecture Notes in Artificial Intelligence*, pages 148–165. Springer, 2010.

[14] Graham Priest. *In Contradiction: A Study of the Transconsistent.* Oxford University Press, 2006.

[15] P. K. Schotch and R. E. Jennings. Non-kripkean deontic logic. In R. Hilpinen, editor, *New Studies in Deontic Logic: Norms, Actions, and the Foundations of Ethics*, pages 149–162. Reidel Pub Co., 1981.

[16] Lutz Schröder and Dirk Pattinson. Rank-1 modal logics are coalgebraic. *Journal of Logic and Computation*, December 2008. Forthcoming. Online access `http://dx.doi.org/10.1093/logcom/exn096`.

[17] Walter Sinnott-Armstrong. *Moral Dilemmas.* Basil Blackwell, Oxford, 1988.

[18] Christian Straßer. A deontic logic framework allowing for factual detachment. *Journal of Applied Logic*, 2010. Forthcoming.

[19] Leendert van der Torre and Yao Hua Tan. Two-phase deontic logic. *Logique et Analyse*, (171-172):411–456, 2000.

[20] Peter Verdée. Adaptive logics using the minimal abnormality strategy are π_1^1-complex. *Synthese*, 167:93–104, 2009.