A Logic for Prioritized Normative Reasoning*

Frederik Van De Putte and Christian Straßer
Centre for Logic and Philosophy of Science
Ghent University
Blandijnberg 2, 9000 Gent, Belgium
{frvdeput.vandeputte,christian.strasser}@ugent.be

May 11, 2011

Abstract

In this paper we present the logic $\mathbf{MP}_{\sqsubset}$ that explicates reasoning on the basis of prioritized obligations. Although formal criteria to handle prioritized obligations have been formulated in the literature, little attention has been paid to the actual (non-monotonic) reasoning that makes use of these criteria. The dynamic proof theory of $\mathbf{MP}_{\sqsubset}$ fills this lacuna. This paper focuses on premise sets consisting of possibly conflicting prima facie obligations that have a modular order. $\mathbf{MP}_{\sqsubset}$ allows to derive –inter alia– the actual, all-things-considered obligations from such premise sets. It is an adaptive logic defined in the new generic format from [35], whence a rich meta-theory is immediately available (e.g. soundness and completeness, idempotence, reflexivity, etc.). In addition, we establish some meta-theoretic results that are specific to the context of prioritized obligations. With the aid of concrete examples, we illustrate properties of $\mathbf{MP}_{\sqsubset}$ which improve on other existing criteria for prioritized obligations. Finally, we show how this logic may also be applied to the context of prioritized belief bases.

keywords: prioritized obligations; adaptive logics; deontic conflicts

1 Introduction

1.1 Deontic Conflicts and Adaptive Logics

Deontic conflicts have been the subject of much debate in philosophical logic and computer science.[1] Roughly speaking, a deontic conflict occurs if two or more obligations[2] cannot be mutually realized – we will present a more precise characterization of deontic conflicts in Section 1.3. In the face of such conflicts, Standard Deontic Logic (henceforth **SDL**) leads to triviality in view of the rule (D): from OA, infer $\neg O \neg A$. Giving up (D) is necessary but not sufficient to

*We are indebted to Joke Meheus and Mathieu Beirlaen for their comments to previous drafts.

[1] See e.g. [37, 8, 9, 10, 32]. See [24] for an overview of the literature on deontic conflicts.

[2] In this paper, we use the generic term "obligations" to refer to duties, imperatives, rules, norms, (moral) laws, and so on.

allow for deontic conflicts: whenever the premises feature a deontic conflict, the other rules of **SDL** still cause *deontic explosion*, i.e. the conclusion that everything is obligatory.

To solve this problem, various conflict-tolerant deontic logics have been developed over the last few decades. As Lou Goble points out in his [12, pp. 462-465], there are basically three options to avoid deontic explosion in the face of conflicting obligations: (i) reject the modal inheritance rule: from OA and $A \vdash B$ infer OB (see e.g. [11, 12]); (ii) reject the axiom of aggregation (AND): $OA \wedge OB \supset O(A \wedge B)$ (see e.g. [36, 8, 32]); (iii) weaken one or more of the non-modal connectives. Examples of the third option can be found in [6, 29], where the classical negation is replaced by a paraconsistent one, and in [23, 7], where a relevant implication is used.

Implementing any of these options in terms of a monotonic logic falls prey to the objection that a number of intuitive inferences are no longer valid. There is a variety of non-monotonic formalizations that are conflict-tolerant and give rise to stronger consequence relations (e.g. Input/Output logics with constraints [22] and Horty's [16]). However, these approaches typically lack a proof theory that explicates the (dynamics of) reasoning on the basis of deontic conflicts.

Recently, adaptive deontic logics have been developed that are satisfactory in all the discussed respects: while allowing for genuine deontic conflicts, they offer a strong consequence relation and a dynamic proof theory (see [3, 25, 33, 34]). Every such adaptive logic is based on a monotonic conflict-tolerant deontic logic that is designed in terms of one of the three options (i)–(iii). The general idea of the adaptive logics is that the omitted rules of standard deontic logic are recuperated as much as possible – unless this leads to some form of explosion. In this way we obtain a significantly larger consequence set compared to the logic that defines the monotonic core of the adaptive logic. Moreover, given that the adaptive logics are defined in the standard format from [2], an intuitive dynamic proof theory that is sound and complete with respect to a static semantics, and a great number of metatheoretic results are immediately available.

Take for example the logic $\mathbf{P2.1}^r$ from [25]. This system is based on Goble's $\mathbf{SDL}_a\mathbf{P}_e$ from [8]. The semantics of $\mathbf{SDL}_a\mathbf{P}_e$ offers a way to interpret conflicts between prima facie obligations. The latter are expressed by the modal operator O_e. (AND) does not hold for these: $\nvdash_{\mathbf{SDL}_a\mathbf{P}_e} (O_eA \wedge O_eB) \supset O_e(A \wedge B)$. Goble's system also allows for the expression of actual, all-things-considered obligations: obligations that behave classically and are considered as guiding our actions. These are expressed with the aid of the operator O_a. Aggregation holds for such obligations: $\vdash_{\mathbf{SDL}_a\mathbf{P}_e} (O_aA \wedge O_aB) \supset O_a(A \wedge B)$, and also $\vdash_{\mathbf{SDL}_a\mathbf{P}_e} (O_aA \wedge O_eB) \supset O_e(A \wedge B)$.

The adaptive logic $\mathbf{P2.1}^r$ from [25] makes it possible to turn prima facie obligations into all-things-considered obligations, on the condition that they are not contradicted by other prima facie obligations. This is realized by allowing for conditional applications of the rule (O_eO_a): $O_eA \supset O_aA$. As a consequence, we regain the rule of aggregation for all those obligations that are not involved in a conflict. The rules of **SDL** are thus recuperated by making the detour via all-things-considered obligations.

1.2 Ordered Sets of Obligations

In this paper, the idea implemented in $\mathbf{P2.1}^r$ will be applied to prioritized obligations. Our logic $\mathbf{MP}_{\sqsubset}$ allows one to derive from a prima facie obligation to bring about A the all-things-considered obligation to bring about A, on the assumption that there is no conflicting obligation with at least the same priority. Note that in many cases, our norms come in different degrees of importance, specificity or urgency. For instance, in most countries, there is a fixed hierarchy between different kinds of traffic rules: those enforced by the signaling boards, by marks on the road, by traffic lights, or by a police officer's commands. In some specific situation, e.g. when we happen to be at the site of a car accident, the traffic rules may be overruled by more urgent and compelling obligations, such as taking an injured kid away from the site of the accident. When a conflict arises in these cases, the agents typically reason from their prima facie obligations and their respective degrees of priority, to find out what they ought to do.

The idea that prima facie obligations are to some extent ordered, and that this may help to resolve contradictions between them, was initiated by Ross in his [31]. Formal investigations of this idea started with [1], and are still ongoing, see e.g. [13, 14, 5]. Most authors in the field start from an ordered set of obligations, and provide a criterion to fix a set of all-things-considered obligations. However, they do not provide a proof theory that explicates the reasoning that could lead to such a selection. As we will argue, adaptive logics are a good candidate to fill this lacuna.

Our logic $\mathbf{MP}_{\sqsubset}$ allows us to solve the problem of prioritized conflicts within a new generic format of prioritized adaptive logics that was studied in [35]. As a result, many important meta-theorems, as well as a full-blown proof theory and semantics are immediately available.

To completely settle the working ground, we need to make two restrictions. First of all, we will focus on sets of obligations that are ordered in a modular way.[3] This means that the set of prima facie obligations can be represented by a tuple: $\mathbb{O} = \langle \mathcal{O}_1, \mathcal{O}_2, \ldots \rangle$, where $\mathcal{O}_1$ contains the most important obligations, $\mathcal{O}_2$ the obligations that are less important, $\mathcal{O}_3$ obligations that are still less important, and so on.We will sometimes say that an obligation A has *priority level* i, by which we mean that $A \in \mathcal{O}_i$. Note however that the same obligation may occur in different sets $\mathcal{O}_i$. We will use $\mathcal{O}$ to refer to the set $\mathcal{O}_1 \cup \mathcal{O}_2 \cup \ldots$, hence the set of all prima facie obligations (irrespective of their priority level).

We will also restrict ourselves to the framework of monadic deontic logic in this paper. Although it is possible to generalize our treatment of prioritized obligations to a dyadic setting, this would severely lengthen the paper since it leads to problems in its own right.[4] We will return to these restrictions in the concluding section, where we discuss possible extensions of the current system.

1.3 Some Examples

Before turning to the formal system, let us present some concrete examples. These will help us to clarify the logic we present below, and to compare it to other approaches in the literature. Case 1 is inspired by [13, pp. 6-7], Case 2 by the visiting daughters example from [17, p. 581], and Case 3 by the famous Smith example from [16, p. 37].

[3]For a general definition of the concept of a modular order, see e.g. [19].

[4]See [33, 34] for some adaptive logics based on dyadic deontic logics.

Case 1. Mary had a car accident, with some minor material damage as a result. She faces the obligation to stay at the site of the accident to fill in insurance papers (S). However, she also promised her mother to pick her up from the supermarket and take her home (M). Finally, her boss asked her to post an urgent letter this same morning (P). It is impossible to post the letter in time, and to fill in the entire insurance poll: S excludes P. However, she can call her mother to notify her she will be a bit later, and hence do both S and M without too much trouble. The obligation to do S has a higher priority than both the one to do P and the one to do M, while the latter two are equally important.

Case 2. Michael has promised his daughter to pay her a visit today (D). However, he recieves the news that his uncle is very ill, and now faces the problem that he should also visit him (U). The obligation to do U is more urgent than the one to do D. Moreover, U and D are practically incompatible. Finally, Michael also told his nephew he would drop by whenever he was around (N). The nephew lives close to the uncle, hence U and N can be combined easily, while N and D exclude each other. Visiting the nephew is less important than visiting the daughter.

Case 3. According to federal law, Smith has the obligation to pay taxes and either serve in the army or perform alternative service to his country: $T \wedge (F \vee S)$. However, according to his political convictions, he ought not to pay taxes ($\neg T$). Moreover, since he feels responsible for his family, he thinks he should not serve in the army ($\neg F$). This last obligation takes priority over all the other ones, while his political convictions are still more important to him than the laws of his country.

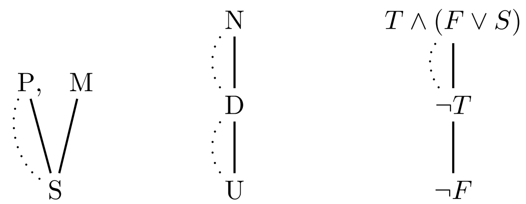

Figure 1: Illustrations for the priorities among the obligations in Cases 1–3. Dotted lines indicate incompatibilities while solid lines indicate priorities where obligations at the bottom of the graph have the highest priority.

Figure 1 represents the priority-relations and conflicts in the different cases. For now, we leave it to the reader to decide what the actual obligations of Mary, Michael and Smith are, given the above descriptions.

In the first two examples, we used a notion of impossibility, e.g. when we said in Case 1 that it is "impossible" for Mary to both stay at the site of the accident (S) and post the letter (P). Inspired by input-output logic [22], we will henceforth speak of "constraints" to refer to statements about practical or physical matters of fact that restrict the kind of actions we may perform. We will assume the set of constraints $\mathcal{C}$ to be consistent and closed under classical logic, i.e. $\mathcal{C} = Cn_{\mathbf{CL}}(\mathcal{C})$.

We may now express in a more precise way what is meant by a deontic conflict. A deontic conflict is always relative to the set of prima facie obligations $\mathcal{O}$ and a set of constraints $\mathcal{C}$. That our prima facie obligations conflict with our constraints, means that in view of the latter, we cannot mutually realize each of the former. More formally, it means that $\mathcal{O} \cup \mathcal{C}$ is inconsistent.[5] As we indicated before, our actual obligations should properly guide our actions, whence we want them to be conflict-free according to this general concept: we want it to be possible that each of them can be carried out, according to the given set of constraints.[6]

In Section 2, we will discuss how the various examples are translated into premise sets for the adaptive logic $\mathbf{MP}_{\sqsubset}$. The latter is presented in Section 3. Section 4 presents some meta-theoretic results that are specific to the context of prioritized obligations. We will return to the examples in Section 5 to illustrate the proof theory of $\mathbf{MP}_{\sqsubset}$ and to compare some existing criteria for preferred obligations to this new system. In Section 6, we briefly argue that our logic may also be applied to prioritized belief bases. We mention some prospects for further research and loose ends in the concluding section.

2 The Logic MP

As mentioned in Section 1, every adaptive logic is based on a monotonic logic, which it typically strengthens non-monotonically. In the language of adaptive logics, such an underlying logic is called the "lower limit logic" of the adaptive logic. Hence we will first define a lower limit logic **MP**, which can be seen as **SDL** extended with a multi-modal variant of the logic **P** from [8].

2.1 The Language of MP

The language of **MP** contains an infinite number of conflict-tolerant ought-operators: $O_1, O_2, O_3, \ldots$. The formula $O_i A$ should be read as: there is a prima facie obligation of priority level i that tells us to do or bring about A. It is important to note that the priority of the normative standard gets higher as the priority index gets *lower*. O_1-obligations are thus the strongest, most compelling prima facie obligations, O_2-obligations are weaker, and so on. Each O_i-operator behaves exactly like the O_e-operator from $\mathbf{SDL}_a\mathbf{P}_e$ – see Section 1.1. Actual obligations will be denoted by the O-operator without index. This O-operator behaves just as the ordinary O-operator from **SDL**.

Henceforth, let $\mathbb{N}$ denote the set of natural numbers without 0. $\mathcal{L}$ will refer to the standard language of classical propositional logic, built up from the connectives $\neg, \vee, \wedge, \supset, \equiv$ and a set $\mathcal{S}$ of sentential letters. We use $\mathcal{W}^l$ to refer to the set of literals (sentential letters and their negations). $\mathcal{L}^M$ is $\mathcal{L}$ extended with the modal operators O and $\langle O_i \rangle_{i \in \mathbb{N}}$. Although it is possible to define the respective

[5] Since constraints may also be of a purely logical kind, our concept includes more basic deontic conflicts, such as having the obligation to do A and the obligation to do $\neg A$.

[6] Although we do not discuss them here, one might also have deontic constraints on the set of actual obligations. Such constraints relate to the concept of "derogation" as studied in e.g. [1]. To derogate $\mathcal{O}$ in A means to interpret $\mathcal{O}$ in such a way that A is no longer entailed by it. For example, where $\neg A$ is a basic human right, and $\mathcal{O}$ is a set of laws that entail that A is obligatory, a judge may select a subset of $\mathcal{O}$ such that A is no longer entailed.

duals P and $\langle P_i \rangle_{i\in\mathbb{N}}$, we will not do so here, since we are only interested in reasoning about obligations.

Where $\mathcal{W}$ is the set of all well-formed formulas of $\mathcal{L}$, the set of well-formed formulas of $\mathcal{L}^M$ is defined as the smallest set $\mathcal{W}^M$ that satisfies the following conditions:[7]

(i) if $A \in \mathcal{W}$ then $A \in \mathcal{W}^M$
(ii) if $A \in \mathcal{W}$ then $OA \in \mathcal{W}^M$
(iii) if $A \in \mathcal{W}$ then $O_iA \in \mathcal{W}^M$ for every $i \in \mathbb{N}$
(iv) if $A \in \mathcal{W}^M$ then $\neg A \in \mathcal{W}^M$
(v) if $A, B \in \mathcal{W}^M$ then $A \vee B, A \wedge B, A \supset B, A \equiv B \in \mathcal{W}^M$

We will now explain how an ordered set of obligations, together with a set of constraints, is translated into a premise set. Physical and practical constraints will be translated as follows: where A is ruled out by such a constraint, $\neg OA$ is introduced as a premise. Thus, the impossibility to do both S and P will be expressed by $\neg O(P \wedge S)$. This is justified by the observation that we want our actual obligations to be obeyable in a practical sense, as required by the "ought implies can"-principle (OIC): $OA \supset \Diamond A$ (where $\Diamond A$ expresses that A is physically or practically possible). Note that (OIC) implies $\neg\Diamond A \supset \neg OA$ by contraposition. Of course, a more natural approach would make use of additional nomological modalities, but this would severely complicate the language of the system. Logical constraints will be dealt with solely by the logic **MP** itself, as we will explain in Section 2.2.

This means that the translation works as follows. Where $\mathbb{O} = \langle \mathcal{O}_1, \mathcal{O}_2, \ldots \rangle$ is a sequence of sets of propositions that represent obligatory states or actions, and $\mathcal{C}$ is a set of physical and practical constraints, we define:

$\Gamma_{\mathbb{O},\mathcal{C}} = \{O_iA \mid A \in \mathcal{O}_i, i \in \mathbb{N}\} \cup \{\neg O\neg A \mid A \in \mathcal{C}\}$

In the remainder, we will use Γ as a metavariable for sets that are obtained by such a translation. Let us illustrate this by the examples from Section 1.3. These will be translated as follows:

Case 1 : $\Gamma_1 = \{O_1S, O_2P, O_2M, \neg O\neg(\neg S \vee \neg P)\}$

Case 2 : $\Gamma_2 = \{O_1U, O_2D, O_3N, \neg O\neg(\neg U \vee \neg D) \neg O\neg(\neg N \vee \neg D)\}$

Case 3 : $\Gamma_3 = \{O_3(T \wedge (F \vee S)), O_2\neg T, O_1\neg F\}$

Since replacement of equivalents holds within the scope of the O-operator, we can rewrite the first two sets as $\Gamma'_1 = \{O_1S, O_2P, O_2M, \neg O(S \wedge P)\}$ and $\Gamma'_2 = \{O_1U, O_2D, O_3N, \neg O(U \wedge D) \neg O(N \wedge D)\}$.

2.2 The Logic MP

In order to allow for conflicting obligations, we will generalize Goble's multirelational semantics for the system **P** – see e.g. [8]. The latter is itself a generalization of the semantics of **SDL**: not one, but many accessibility relations are in play.Goble defines a model in terms of a set of accessibility relations: $\mathcal{R} = \{R_1, R_2, \ldots\}$. To handle prioritized sets of obligations, two basic changes to Goble's **P**-system will be made.

[7] As is clear from the definition of $\mathcal{W}^M$, we do not allow for nested obligations.

First of all, we use a *set* of sets of accessibility relations $\boldsymbol{R} = \{\mathcal{R}_1, \mathcal{R}_2, \ldots\}$. Where $i \in \mathbb{N}$, $\mathcal{R}_i$ refers to a set of prima facie obligations of priority level i, that all have the same priority level i. This way, two prima facie obligations with the same priority may still be conflicting. As may be expected, the O_i-operator is linked to the set $\mathcal{R}_i$.

The other change has to do with the additional O-operator. Since we want this operator to behave classically, it is stipulated that there is a single accessibility relation R that corresponds to the accessibility relation of **SDL**. This implies that the logic itself takes care of the logical constraints on the set of actual obligations: for every self-contradictory formula A, $\neg OA$ is an **MP**-theorem.

An **MP**-model M is a quintuple $\langle W, \boldsymbol{R}, R, v, w_0 \rangle$, where W is a set of possible worlds, $\boldsymbol{R} = \{\mathcal{R}_1, \mathcal{R}_2, \ldots\}$ is a set of non-empty sets of serial accessibility relations on W, R is a serial accessibility relation, $v : \mathcal{S} \times W \to \{0, 1\}$ is an assignment function and $w_0 \in W$ is the actual world.

The valuation v_M defined by the model M is characterized by:

C1 where $A \in \mathcal{S}$, $v_M(A, w) = v(A, w)$
C2 $v_M(\neg A, w) = 1$ iff $v_M(A, w) = 0$
C3 $v_M(A \vee B, w) = 1$ iff $v_M(A, w) = 1$ or $v_M(B, w) = 1$
C4 $v_M(A \wedge B, w) = 1$ iff $v_M(A, w) = 1$ and $v_M(B, w) = 1$
C5 $v_M(A \supset B, w) = 1$ iff $v_M(A, w) = 0$ or $v_M(B, w) = 1$
C6 $v_M(A \equiv B, w) = 1$ iff $v_M(A, w) = v_M(B, w)$
C7 $v_M(O_i A, w) = 1$ iff, for an $R_j \in \mathcal{R}_i$, $v_M(A, w') = 1$ for all w' such that $R_i ww'$
C8 $v_M(OA, w) = 1$ iff $v_M(A, w') = 1$ for all w' such that Rww'

An **MP**-model M verifies A, $M \Vdash A$ iff $v_M(A, w_0) = 1$. $\models_{\mathbf{MP}} A$ iff all **MP**-models verify A. We say that M is an **MP**-model of Γ, $M \Vdash \Gamma$ iff $M \Vdash A$ for every $A \in \Gamma$. We use $\mathcal{M}_{\mathbf{MP}}(\Gamma)$ to refer to the set of **MP**-models of Γ. Finally, $\Gamma \models_{\mathbf{MP}} A$ iff all **MP**-models of Γ verify A.

A syntax for **MP** is obtained as follows. We extend an axiomatization of classical propositional logic (henceforth **CL**) with the following axioms (where $A, B \in \mathcal{W}$):

K $O(A \supset B) \supset (OA \supset OB)$
D $OA \supset \neg O \neg A$

and close it under modus ponens (MP) and the following rules (where $A, B \in \mathcal{W}$):

RN if $\vdash A$, then $\vdash OA$
RM_i where $i \in \mathbb{N}$: if $\vdash A \supset B$, then $\vdash O_i A \supset O_i B$
P_i where $i \in \mathbb{N}$: if $\vdash A$, then $\vdash \neg O_i \neg A$
RN_i where $i \in \mathbb{N}$: if $\vdash A$, then $\vdash O_i A$

where $\vdash$ indicates membership in the set of **MP**-axioms. The two axioms (K) and (D) together with the rules (MP) and (RN) deliver **SDL** for O. The rules (RM_i), (P_i) and (RN_i) deliver Goble's **P** for all operators O_i. Note that there are no bridging rules that link the different ought-operators: OA does not imply that $O_i A$ or vice versa; $O_i A$ does not imply that $O_j A$ for any $j \neq i$. Note also that $O_i A \nvdash_{\mathbf{MP}} \neg O_i \neg A$ and $O_i A, O_j \neg A \nvdash O_k B$ for any $i, j, k \in \mathbb{N}$. In Section 4, we will discuss even more general results with regards to the conflict-tolerance of **MP**.

We define $\Gamma \vdash_{\mathbf{MP}} A$ iff there are $B_1, \ldots, B_n \in \Gamma$ such that $\vdash_{\mathbf{MP}} (B_1 \wedge \ldots \wedge B_n) \supset A$. Note that this definition immediately entails that the consequence relation is compact. The proof of the following theorem is outlined in the appendix:

Theorem 1 $\Gamma \vdash_{\mathbf{MP}} A$ *iff* $\Gamma \models_{\mathbf{MP}} A$ *(Soundness and Completeness for* **MP***)*

3 The Prioritized Adaptive Logic $\mathbf{MP}_{\sqsubset}$

3.1 The Adaptive Approach

Recall that the central aim of this paper is to capture how we can *reason* from an ordered set $\mathcal{O}$ of prima facie obligations and a set $\mathcal{C}$ of constraints towards a set of actual obligations. This will be realized in terms of a dynamic proof theory that explicates such reasoning. In Section 3.2, we will present this proof theory in detail. To facilitate the reading, we will introduce the basic concepts here in an informal way.

In our current formal framework, to derive actual obligations from prima facie obligations may be realized by the following rule (where $i \in \mathbb{N}$):

($\mathrm{O}_i\mathrm{O}$) if O_iA, then OA

Recall that this rule is not valid in **MP**. However, adding ($\mathrm{O}_i\mathrm{O}$) to the axioms of **MP** would result in plain triviality whenever our prima facie obligations are jointly incompatible.

The adaptive logic we will present uses the language and inference rules of **MP**, but enhances it with the *defeasible* application of ($\mathrm{O}_i\mathrm{O}$). As soon as such a particular application turns out to be harmful, the logic ensures that it is retracted. However, other applications of the same rule may still be allowed for by the logic.

The central motor behind adaptive logics is the assumption that certain formulas – the so-called abnormalities – are false "until and unless proven otherwise". In this case, abnormalities express that something is a prima facie obligation (of some priority level i), but not an actual obligation. Hence any formula of the form $O_iA \wedge \neg OA$ is an abnormality. Consider $\Gamma_1 = \{O_1S, O_2P, O_2M, \neg O(S \wedge P)\}$. The adaptive logic derives OM on the assumption that $O_2M \wedge \neg OM$ is false. Note that $\Gamma_1 \vdash_{\mathbf{MP}} OM \vee (O_2M \wedge \neg OM)$.

Three problems arise when implementing this idea. A complication that is well-known from the study of adaptive logics consists in the fact that we cannot always derive abnormalities separately, though we may be able to derive disjunctions of them. This is the case in our example: $\Gamma_1 \vdash_{\mathbf{MP}} (O_1S \wedge \neg OS) \vee (O_2P \wedge \neg OP)$, while $\Gamma_1 \nvdash_{\mathbf{MP}} O_1S \wedge \neg OS$ and $\Gamma_1 \nvdash_{\mathbf{MP}} O_2P \wedge \neg OP$. If we would assume both $O_1S \wedge \neg OS$ and $O_2P \wedge \neg OP$ to be false, we would end up with an **MP**-trivial consequence set. This is avoided by taking derivable *disjunctions* of abnormalities, or Dab-formulas in short, into account – see Section 3.2 where this is formally specified.

A second difficulty concerns the form of the abnormalities. Simply taking $O_iA \wedge \neg OA$ for any $A \in \mathcal{W}$ as the form of the abnormalities, leads to a so-called flip-flop-logic: a logic that considers every prima facie obligation as actual when there are no conflicts at all, but that reduces to **MP** as soon as a Dab-formula is derivable from the premise set. In our example, we can prove that $\Gamma_1 \vdash_{\mathbf{MP}} (O_2M \wedge \neg OM) \vee (O_1(S \vee \neg M) \wedge \neg O(S \vee \neg M)) \vee (O_2P \wedge \neg OP)$. Hence M would become problematic in view of Γ_1 and OM would not be derivable. This difficulty is overcome by using a slightly more complex form of the abnormalities, borrowed from [26].

The idea behind the new form of the abnormalities is this: let Θ be a finite and non-empty set of literals and let $\bigvee \Theta$ be the disjunction of the members of Θ (if $\Theta = \{A\}$, $\bigvee \Theta = A$). The obligation to do $\bigvee \Theta$ is said to behave abnormally, if for some $i \in \mathbb{N}$, $O_i \bigvee \Theta \wedge \neg O \bigvee \Theta$ is true, *or* if for some non-empty $\Theta' \subset \Theta$, $O_i \bigvee \Theta' \wedge \neg O \bigvee \Theta'$ is true. This warrants that if some part of the disjunction $\bigvee \Theta$ already behaves abnormally, then $\bigvee \Theta$ automatically becomes suspicious as well. The resulting abnormality will be abbreviated by $\sigma^i(\Theta)$. As we did for obligations, we will speak of the priority level of a given abnormality: $\sigma^i(\Theta)$ has priority level i.

The third problem is specific to this paper: how can the logic take the priorities of the obligations (vs. abnormalities) into account? Note that the lower the index of an abnormality, the more we should avoid this abnormality, since it corresponds to an obligation of a higher level. We will use the generic format from [35] to deal with the priority order on the obligations resp. abnormalities. The main idea is that if two prima facie obligations are conflicting and one of them has higher priority, then the latter will be actualized (unless there is another, yet more important obligation that conflicts with it).

3.2 Some Definitions

In this section, we will present the formal apparatus that characterizes the logic $\mathbf{MP}_{\sqsubset}$. We start with some general preliminary tools and briefly explain the $\mathbf{MP}_{\sqsubset}$-semantics. After that, we define the dynamic proof theory.

The General Framework. The logic $\mathbf{MP}_{\sqsubset}$ is defined by a triple: (i) the lower limit logic $\mathbf{MP}$, (ii) a sequence of sets of abnormalities $\langle \Omega_i \rangle_{i \in \mathbb{N}}$, and (iii) the strategy $\sqsubset$-minimal abnormality.[8] The lower limit logic provides all the consequences that hold unconditionally, i.e. without relying on the normal behavior of the premises.

In order to define the sequence of sets of abnormalities, we need some technical preparations. Where Θ is a non-empty and finite subset of $\mathcal{W}^l$, we will use the following abbreviation:

$$\sigma^i(\Theta) = \bigvee \{O_i \bigvee \Theta' \wedge \neg O \bigvee \Theta' \mid \emptyset \neq \Theta' \subseteq \Theta\} \qquad (1)$$

To avoid notational clutter, we will skip the set brackets for concrete sets of literals, e.g. we write $\sigma(p, \neg q)$ instead of $\sigma(\{p, \neg q\})$. To get better grip on the form of the abnormalities and their abbreviation, consider the following examples:

$$\sigma^2(p) = O_2 p \wedge \neg O p$$

$$\sigma^2(\neg q) = O_2 \neg q \wedge \neg O \neg q$$

$$\sigma^3(\neg p, q) = (O_3 \neg p \wedge \neg O \neg p) \vee (O_3 q \wedge \neg O q) \vee (O_3(\neg p \vee q) \wedge \neg O(\neg p \vee q))$$

[8] A weaker variant of $\mathbf{MP}_{\sqsubset}$ has the strategy $\sqsubset$-reliability. For reasons of space we will not discuss that system here – see [35] for more details.

$$\sigma^1(p,q,r) = (\Box p \wedge \neg p) \vee (\Box q \wedge \neg q) \vee (\Box r \wedge \neg r) \vee (\Box(p \vee q) \wedge \neg(p \vee q)) \vee (\Box(p \vee r) \wedge \neg(p \vee r)) \vee (\Box(q \vee r) \wedge \neg(q \vee r)) \vee (\Box(p \vee q \vee r) \wedge \neg(p \vee q \vee r))$$

Note that the number of disjuncts of an abnormality $\sigma^i(\Theta)$ grows exponentially with the number of literals in Θ. Also, where Θ' is a non-empty subset of Θ, we have that $\vdash_{\mathbf{MP}} \sigma^i(\Theta') \supset \sigma^i(\Theta)$.

Where $i \in \mathbb{N}$, the ith set of abnormalities is defined as $\Omega_i = \{\sigma^i(\Theta) \mid \emptyset \neq \Theta \subset \mathcal{W}^l\}$. Note that $\Omega_i \cap \Omega_j = \emptyset$ for every $i, j \in \mathbb{N}, i \neq j$. We thus obtain the sequence of abnormalities $\langle \Omega_i \rangle_{i \in \mathbb{N}}$. Let henceforth $\Omega = \bigcup_{i \in \mathbb{N}} \Omega_i$, and let $Dab(\Delta)$ denote a disjunction of members of Ω.

Semantics. The easiest way to explain how the order on the obligations is implemented in $\mathbf{MP}_{\sqsubset}$ is by considering its semantics. The set of $\mathbf{MP}_{\sqsubset}$-models of Γ is a subset of the set of $\mathbf{MP}$-models of Γ. This immediately entails that $\mathbf{MP}_{\sqsubset}$ strengthens $\mathbf{MP}$ – see also Theorem 7 in Section 3.3. To explain how the $\mathbf{MP}_{\sqsubset}$-models are selected, we need some more definitions.

For every $\mathbf{MP}$-model M, we define its abnormal part $Ab(M) = \{A \in \Omega \mid M \Vdash A\}$. Models are selected resp. deselected by the adaptive logic in view of their abnormal part. This abnormal part can contain abnormalities from the many different sets Ω_i, whence we may consider it as a tuple: $\langle Ab(M) \cap \Omega_1, Ab(M) \cap \Omega_2, \ldots \rangle$. This means that we can define a lexicographic order $\sqsubset$ on the abnormal parts, and more generally, on $\wp(\Omega)$ as a whole:

Definition 1 *Where $\Delta, \Delta' \subseteq \Omega$: $\langle \Delta \cap \Omega_i \rangle_{i \in \mathbb{N}} \sqsubset_{\mathsf{lex}} \langle \Delta' \cap \Omega_i \rangle_{i \in \mathbb{N}}$ iff there is an $i \in \mathbb{N}$ such that (1) for all $j < i$, $\Delta \cap \Omega_j = \Delta' \cap \Omega_j$, and (2) $\Delta \cap \Omega_i \subset \Delta' \cap \Omega_i$. We write $\Delta \sqsubset \Delta'$ iff $\langle \Delta \cap \Omega_i \rangle_{i \in \mathbb{N}} \sqsubset_{\mathsf{lex}} \langle \Delta' \cap \Omega_i \rangle_{i \in \mathbb{N}}$.*[9]

To illustrate the above definition, consider the following example:

$$\Delta = \{\sigma^1(p), \sigma^2(q, \neg r), \sigma^3(r), \sigma^4(\neg p, q)\}$$
$$\Delta' = \{\sigma^1(p), \sigma^2(q, \neg r), \sigma^2(s), \sigma^4(\neg p, q)\}$$
$$\Delta'' = \{\sigma^1(p), \sigma^2(q, \neg r), \sigma^2(s), \sigma^3(r)\}$$
$$\Delta''' = \{\sigma^1(\neg q), \sigma^2(q, \neg r), \sigma^2(s), \sigma^3(r)\}$$

According to Definition 1, $\Delta \sqsubset \Delta' \sqsubset \Delta''$. That is, Δ beats Δ' at level 2, and Δ' beats Δ'' at level 3. It follows immediately that $\Delta \sqsubset \Delta''$. However, $\Delta \not\sqsubset \Delta'''$, since the two are incomparable at level 1. All this becomes a lot more clear as soon as we represent these sets of abnormalities Θ in columns, where each separate column represents the intersection of Θ with an Ω_i – see Table 1.

Ω	Ω_1	Ω_2	Ω_3	Ω_4	...
$\langle \Delta \cap \Omega_i \rangle_{i \in \mathbb{N}} =$	$\{\sigma^1(p)\}$	$\{\sigma^2(q, \neg r)\}$	$\{\sigma^3(r)\}$	$\{\sigma^4(\neg p, q)\}$	...
$\langle \Delta' \cap \Omega_i \rangle_{i \in \mathbb{N}} =$	$\{\sigma^1(p)\}$	$\{\sigma^2(q, \neg r), \sigma^2(s)\}$	$\emptyset$	$\{\sigma^4(\neg p, q)\}$	...
$\langle \Delta'' \cap \Omega_i \rangle_{i \in \mathbb{N}} =$	$\{\sigma^1(p)\}$	$\{\sigma^2(q, \neg r), \sigma^2(s)\}$	$\{\sigma^3(r)\}$	$\emptyset$	...
$\langle \Delta''' \cap \Omega_i \rangle_{i \in \mathbb{N}} =$	$\{\sigma^1(\neg q)\}$	$\{\sigma^2(q, \neg r), \sigma^2(s)\}$	$\{\sigma^3(r)\}$	$\emptyset$	...

[9] Technically speaking, $\sqsubset$ is not a lexicographic order, but since it is obtained from the lexicographic order $\sqsubset_{\mathsf{lex}}$ in a rather straightforward way, we will often call $\sqsubset$ a lexicographic order as well. Lexicographic orders have been previously used for the formal explication of reasoning on the basis of prioritized information. Lehmann used them to deal with priorities among defaults [20], Nebel [27] in order to deal with prioritized theory bases and Hansen [13] applied Nebel's preference order to the context of prioritized imperatives.

Table 1: A representation of the sets $\Delta, \Delta', \Delta''$ and Δ''' as tuples of sets of abnormalities.

Once we have the definition of $\sqsubset$, the $\mathbf{MP}_{\sqsubset}$-semantics is fairly simple. The $\mathbf{MP}_{\sqsubset}$-models of Γ are those $\mathbf{MP}$-models of Γ whose abnormal part is $\sqsubset$-minimal. We use $\mathcal{M}_{\mathbf{MP}_{\sqsubset}}(\Gamma)$ to refer to the set of all such models, also called the $\sqsubset$-minimal abnormals models of Γ.

Definition 2 *$M \in \mathcal{M}_{\mathbf{MP}_{\sqsubset}}(\Gamma)$ iff $M \in \mathcal{M}_{\mathbf{MP}}(\Gamma)$ and there is no $M' \in \mathcal{M}_{\mathbf{MP}}(\Gamma)$ such that $Ab(M') \sqsubset Ab(M)$.*

Definition 3 *$\Gamma \models_{\mathbf{MP}_{\sqsubset}} A$ iff A is true in all $M \in \mathcal{M}_{\mathbf{MP}_{\sqsubset}}(\Gamma)$.*

Proof Theory. Every line in an $\mathbf{MP}_{\sqsubset}$-proof consists of four elements: a line number i, a formula A, a justification and a finite condition $\Delta \subset \Omega$. Where Γ is the set of premises, the inference rules are given by the following three tables (we omit the line numbers and justifications):

PREM If $A \in \Gamma$:

$$\begin{array}{ll} \vdots & \vdots \\ \hline A & \emptyset \end{array}$$

RU If $A_1, \ldots, A_n \vdash_{\mathbf{MP}} B$:

$$\begin{array}{ll} A_1 & \Delta_1 \\ \vdots & \vdots \\ A_n & \Delta_n \\ \hline B & \Delta_1 \cup \ldots \cup \Delta_n \end{array}$$

RC If $A_1, \ldots, A_n \vdash_{\mathbf{MP}} B \vee Dab(\Theta)$:

$$\begin{array}{ll} A_1 & \Delta_1 \\ \vdots & \vdots \\ A_n & \Delta_n \\ \hline B & \Delta_1 \cup \ldots \cup \Delta_n \cup \Theta \end{array}$$

A *stage* of a proof is a sequence of lines and a proof is a chain of such stages. Adding a line to a proof by applying one of the above rules brings the proof to its next stage, which is the sequence of all lines written so far.

Recall that we wanted to allow for the defeasible application of the rule ($\mathrm{O}_i\mathrm{O}$) from Section 3.1. The third item of the following lemma shows how this rule can be applied in an $\mathbf{MP}_{\sqsubset}$-proof, using the conditional rule RC:

Lemma 1 *Each of the following holds:*

(i) Where Θ is a finite set of literals, $O_i \bigvee \Theta \vdash_{\mathbf{MP}} O \bigvee \Theta \vee \sigma^i(\Theta)$.

(ii) Where $A \in \mathcal{W}$ and $\bigwedge_J \bigvee \Theta_j$ is a conjunctive normal form of A, $O_i A \vdash_{\mathbf{MP}} OA \vee \bigvee_J \sigma^i(\Theta_j)$.

(iii) Where $A \in \mathcal{W}$ and $O_i A$ is derived on a line l of an $\mathbf{MP}_{\sqsubset}$-proof from Γ, we can derive OA on a line l' in an extension of this proof.

Proof. Ad (i): Suppose $O_i \bigvee \Theta$. Then $O \bigvee \Theta \vee (O_i \bigvee \Theta \wedge \neg O \bigvee \Theta)$, whence $O \bigvee \Theta \vee \sigma^i(\Theta)$.

Ad (ii): Suppose O_iA. By (RM$_i$), for each $j \in J$, $O_i \bigvee \Theta_j$. By item (i), $O \bigvee \Theta_j \vee \sigma^i(\Theta_j)$ for each $j \in J$. Hence, $\bigwedge_J (O \bigvee \Theta_j) \vee \bigvee_J \sigma^i(\Theta_j)$. Since aggregation holds for O, $O \bigwedge_J \bigvee \Theta_j \vee \bigvee_J \sigma^i(\Theta_j)$. Since inheritance holds for O, $OA \vee \bigvee_J \sigma^i(\Theta_j)$.
Ad (iii): Suppose O_iA is derived on a line l of an $\mathbf{MP}_{\sqsubset}$-proof from Γ. Let the condition of line l be Δ. By the conditional rule (RC) and item (ii), we can derive OA on a line l' in an extension of the proof, on the condition $\Delta \cup \{\bigvee_J \sigma^i(\Theta_j)\}$, where $\bigwedge_J \bigvee \Theta_j$ is a conjunctive normal form of A. ■

Lemma 1(iii) indicates that it is possible to apply (O$_i$O) in an $\mathbf{MP}_{\sqsubset}$-proof. To explain in what sense this application is defeasible, we will need to introduce a few more concepts and definitions.

At every stage of a proof, a marking definition – see below – determines for each line in the proof whether it is marked or not. That a line which has as its second element A is marked at stage s, indicates that according to our best insights at this stage, A cannot be considered as a consequence on that line. If the line is unmarked at stage s, this indicates that A is a consequence, again, according to our best insights at this stage.

$Dab(\Delta)$ is a Dab-formula at stage s of a proof, iff it is the second element of a line in the proof with an empty condition. $Dab(\Delta)$ is a *minimal* Dab-formula at stage s iff there is no other Dab-formula $Dab(\Delta')$ at stage s for which $\Delta' \subset \Delta$. Where $Dab(\Delta_1), Dab(\Delta_2), \ldots$ are the minimal Dab-formulas at stage s of a proof, let $\Sigma_s(\Gamma) = \{\Delta_1, \Delta_2, \ldots\}$.

We say that $\varphi \subseteq \Omega$ is a *choice set* of $\Sigma_s(\Gamma)$ iff for every $\Delta \in \Sigma_s(\Gamma)$, $\varphi \cap \Delta \neq \emptyset$. For the border case where $\Sigma_s(\Gamma) = \emptyset$, this means that every set $\varphi \subseteq \Omega$ is a choice set of $\Sigma_s(\Gamma)$, including the empty set. We say that φ is a $\sqsubset$-minimal choice set of $\Sigma_s(\Gamma)$ iff there is no choice set ψ of $\Sigma_s(\Gamma)$ such that $\psi \sqsubset \varphi$. Let $\Phi_s^{\sqsubset}(\Gamma)$ be the set of $\sqsubset$-minimal choice sets of $\Sigma_s(\Gamma)$. It is proven in [35] that at every stage s of a proof from Γ, $\Phi_s^{\sqsubset}(\Gamma)$ is non-empty.

Let us try to give an intuitive characterization of the above concepts. Suppose we have derived a number of Dab-formulas at stage s. Hence we have to acknowledge that some abnormalities have to be true (since the Dab-formulas are derived on the empty condition). More specifically, we know that at least the abnormalities from some choice set φ of $\Sigma_s(\Gamma)$ have to be true. We then select those choice sets of $\Sigma_s(\Gamma)$ that are $\sqsubset$-minimal. $\Phi_s^{\sqsubset}(\Gamma)$ thus represents a stage-relative interpretation of how abnormal our premise set is. We now use this interpretation to mark lines in an $\mathbf{MP}_{\sqsubset}$-proof from Γ:

Definition 4 $\mathbf{MP}_{\sqsubset}$*-Marking: a line l with formula A is marked at stage s iff, where its condition is Δ: (i) no $\varphi \in \Phi_s^{\sqsubset}(\Gamma)$ is such that $\varphi \cap \Delta = \emptyset$, or (ii) for a $\varphi \in \Phi_s^{\sqsubset}(\Gamma)$, there is no line on which A is derived on a condition Θ for which $\Theta \cap \varphi = \emptyset$.*

Put differently: a line with formula A is *un*marked at stage s iff its condition has an empty intersection with at least one $\varphi \in \Phi_s^{\sqsubset}(\Gamma)$, and for every $\psi \in \Phi_s^{\sqsubset}(\Gamma)$, there is a line on which A is derived on a condition Δ such that $\Delta \cap \psi = \emptyset$.

This proof theory reflects a dynamic aspect of our reasoning about conflicting obligations: we may take some obligation as an actual obligation at some point, but only later on learn that this leads to a conflict on the level of the actual obligations. This may be due to the additional information, but it may also be the result of our reasoning about the same set of obligations. At that point, we have to retract some of our earlier conclusions. In Section 5, we will clarify this mechanism, using the canonical examples from Section 1.

As lines can become marked, unmarked and marked again throughout a proof, we also need a stable notion of derivability:

Definition 5 *A is* finally derived *from Γ on line l of a finite stage s iff (i) A is the second element of line l, (ii) line l is not marked at stage s, and (iii) every extension of the stage in which line l is marked may be further extended in such a way that line l is unmarked again.*

Definition 6 $\Gamma \vdash_{\mathbf{MP}_{\sqsubset}} A$ *(* A is finally $\mathbf{MP}_{\sqsubset}$-derivable from Γ*) iff A is finally derived on a line of an $\mathbf{MP}_{\sqsubset}$-proof from Γ.*

3.3 Some Metatheoretic Properties

As defined above, the logic $\mathbf{MP}_{\sqsubset}$ fits the format of a prioritized adaptive logic from [35]. This means that a rich meta-theory is immediately available. We will only mention some of these results here for illustrative purposes, and refer to [35] for more details.

Theorem 2 $\Gamma \vdash_{\mathbf{MP}_{\sqsubset}} A$ *iff* $\Gamma \models_{\mathbf{MP}_{\sqsubset}} A$ *(Soundness and Completeness)*

Theorem 3 *If $M \in \mathcal{M}_{\mathbf{MP}}(\Gamma) - \mathcal{M}_{\mathbf{MP}_{\sqsubset}}(\Gamma)$, then there is an $M' \in \mathcal{M}_{\mathbf{MP}_{\sqsubset}}(\Gamma)$ such that $Ab(M') \sqsubset Ab(M)$. (Strong Reassurance)*

Corollary 1 *If Γ has* **MP***-models, then Γ has $\mathbf{MP}_{\sqsubset}$-models. (Reassurance)*

Theorem 4 *If $\Gamma' \subseteq Cn_{\mathbf{MP}_{\sqsubset}}(\Gamma)$, then $Cn_{\mathbf{MP}_{\sqsubset}}(\Gamma) = Cn_{\mathbf{MP}_{\sqsubset}}(\Gamma \cup \Gamma')$. (Cautious Indifference)*

Theorem 5 *$Cn_{\mathbf{MP}_{\sqsubset}}(\Gamma) = Cn_{\mathbf{MP}_{\sqsubset}}(Cn_{\mathbf{MP}_{\sqsubset}}(\Gamma))$. (Fixed Point)*

Theorem 6 *$\Gamma \subseteq Cn_{\mathbf{MP}_{\sqsubset}}(\Gamma)$. (Reflexivity)*

Theorem 7 *$Cn_{\mathbf{MP}}(\Gamma) \subseteq Cn_{\mathbf{MP}_{\sqsubset}}(\Gamma)$. ($\mathbf{MP}_{\sqsubset}$ strengthens* **MP***)*

Theorem 8 *If $\Gamma \cup \{A\} \vdash_{\mathbf{MP}_{\sqsubset}} B$, then $\Gamma \vdash_{\mathbf{MP}_{\sqsubset}} A \supset B$. (Deduction Theorem)*

4 Some Specific Properties of $\mathbf{MP}_{\sqsubset}$

In this section, we present some properties that are more specific to the logic $\mathbf{MP}_{\sqsubset}$ and its application to premise sets of the form $\Gamma_{\mathbb{O},\mathcal{C}}$. The first theorem and corollary below indicate to what extent **MP** and $\mathbf{MP}_{\sqsubset}$ are conflict-tolerant. The other theorem and the subsequent corollaries express a lower bound on the set of actual obligations that are $\mathbf{MP}_{\sqsubset}$-derivable from any set $\Gamma_{\mathbb{O},\mathcal{C}}$. We refer to the appendix for the proofs of Theorems 9 and 10.

Theorem 9 *$\Gamma_{\mathbb{O},\mathcal{C}}$ has* **MP***-models iff (every $A \in \mathcal{O}$ is* **CL***-satisfiable and $\mathcal{C}$ is* **CL***-satisfiable).*

So $\Gamma_{\mathbb{O},\mathcal{C}}$ may contain any conflict, as long as the set of constraints is internally consistent and there are no prima facie obligations that are contradictory in themselves. A nice property of deontic adaptive logics in general is that they are just as conflict-tolerant as their lower limit logic. This follows immediately from the property of Reassurance (see Corollary 1). Hence we obtain:

Corollary 2 *If every* $A \in \mathcal{O}$ *is* **CL**-*satisfiable and* $\mathcal{C}$ *is* **CL**-*satisfiable, then* $\Gamma_{\mathbb{O},\mathcal{C}}$ *has* $\mathbf{MP}_{\sqsubset}$-*models.*

Note that since the O-operator from $\mathbf{MP}_{\sqsubset}$ behaves according to the O-operator from **SDL**, it follows that whenever $\Gamma_{\mathbb{O},\mathcal{C}}$ is **MP**-satisfiable, then there is a $B \in \mathcal{W}$ such that $\Gamma_{\mathbb{O},\mathcal{C}} \nvdash_{\mathbf{MP}_{\sqsubset}} OB$. Hence the logic $\mathbf{MP}_{\sqsubset}$ also avoids deontic explosion, as long as the antecedent of Corollary 2 holds.

To spell out a lower bound on the set of actual obligations that are $\mathbf{MP}_{\sqsubset}$-derivable from a set $\Gamma_{\mathbb{O},\mathcal{C}}$, we first introduce two more concepts:[10]

Definition 7 *We call* $\Gamma_{\mathbb{O},\mathcal{C}}$

- conflict-free up to level n *iff the set* $\{A \in \mathcal{O}_i \mid i \leq n\} \cup \mathcal{C}$ *is* **CL**-*satisfiable.*
- conflict-free *iff* $\mathcal{O} \cup \mathcal{C}$ *is* **CL**-*satisfiable.*

Note that $\Gamma_{\mathbb{O},\mathcal{C}}$ is conflict-free whenever the prima facie obligations are not in conflict with the set of constraints – see Section 1.3 where we explained our notion of a deontic conflict relative to a set of constraints.

Theorem 10 *If* $\Gamma_{\mathbb{O},\mathcal{C}}$ *is conflict-free up to level n, then the following holds for all* $i \leq n$: *if* $\Gamma_{\mathbb{O},\mathcal{C}} \vdash_{\mathbf{MP}_{\sqsubset}} O_iA$, *then* $\Gamma_{\mathbb{O},\mathcal{C}} \vdash_{\mathbf{MP}_{\sqsubset}} OA$.

The following properties can easily be obtained from Theorem 10 and previously mentioned results. Corollary 3(i) follows from Theorem 10 in view of Theorem 7, Corollary 3(ii) follows from Theorem 10 in view of Theorem 6. Corollary 4 follows from Theorem 10 and Corollary 3 in view of Definition 7.

Corollary 3 *If* $\Gamma_{\mathbb{O},\mathcal{C}}$ *is conflict-free up to level n, then each of the following holds for all* $i \leq n$:

(i) *if* $\Gamma_{\mathbb{O},\mathcal{C}} \vdash_{\mathbf{MP}} O_iA$, *then* $\Gamma_{\mathbb{O},\mathcal{C}} \vdash_{\mathbf{MP}_{\sqsubset}} OA$
(ii) *if* $O_iA \in \Gamma_{\mathbb{O},\mathcal{C}}$, *then* $\Gamma_{\mathbb{O},\mathcal{C}} \vdash_{\mathbf{MP}_{\sqsubset}} OA$

Corollary 4 *If* $\Gamma_{\mathbb{O},\mathcal{C}}$ *is conflict-free, then each of the following holds for all* $i \in \mathbb{N}$:
(i) *if* $\Gamma_{\mathbb{O},\mathcal{C}} \vdash_{\mathbf{MP}_{\sqsubset}} O_iA$, *then* $\Gamma_{\mathbb{O},\mathcal{C}} \vdash_{\mathbf{MP}_{\sqsubset}} OA$
(ii) *if* $\Gamma_{\mathbb{O},\mathcal{C}} \vdash_{\mathbf{MP}} O_iA$, *then* $\Gamma_{\mathbb{O},\mathcal{C}} \vdash_{\mathbf{MP}_{\sqsubset}} OA$
(iii) *if* $O_iA \in \Gamma_{\mathbb{O},\mathcal{C}}$, *then* $\Gamma_{\mathbb{O},\mathcal{C}} \vdash_{\mathbf{MP}_{\sqsubset}} OA$

Corollary 4 implies that if there is no conflict between any of the prima facie obligations in view of the constraints, then all prima facie obligations and their **CL**-consequences will be considered as actual obligations (irrespective of their priority). Note that Theorem 10 only serves as a lower bound on $Cn_{\mathbf{MP}_{\sqsubset}}(\Gamma)$; as we will see in the next section, $\mathbf{MP}_{\sqsubset}$ is actually a lot stronger than this lower bound.

[10] Our notion of "conflict-freeness" is equivalent to the notion of "coherence" used in [15, p. 7], if we restrict the latter to unconditional obligations.

5 The Examples Reconsidered

5.1 Illustration of the Proof Theory

We will use the first example to illustrate the proof theory of $\mathbf{MP}_{\sqsubset}$. Recall that $\Gamma'_1 = \{O_1S, O_2M, O_2P, \neg O(S \wedge P)\}$. We start an $\mathbf{MP}_{\sqsubset}$-proof from Γ'_1 with the use of the rule PREM:

1	O_1S	PREM	$\emptyset$
2	O_2M	PREM	$\emptyset$
3	O_2P	PREM	$\emptyset$
4	$\neg O(S \wedge P)$	PREM	$\emptyset$

At this stage, there is no marking at all: since no Dab-formulas have been derived, $\Phi^{\sqsubset}_4(\Gamma'_1) = \{\emptyset\}$. Mary may infer from this that M is an actual obligation, on the conditional assumption that $O_2M \wedge \neg OM$ is false. Hence the proof continues like this:

5	$OM \vee (O_2M \wedge \neg OM)$	2; RU	$\emptyset$
6	OM	5; RC	$\{O_2M \wedge \neg OM\}$

The crucial move is made between stage 5 and 6: we can see that by the application of the conditional rule RC, the abnormality $O_2M \wedge \neg OM$ is pushed to the condition. Below, we will skip the intermediary step represented at line 5 – note that this is perfectly in line with the definition of the conditional rule RC. Since $\Phi^{\sqsubset}_6(\Gamma'_1) = \{\emptyset\}$, line 6 is unmarked at this stage, which indicates that at this stage, OM is considered as a consequence of the premise set. To illustrate why Definition 5 makes use of the notion of an extension of a proof, consider the following continuation of the proof (we restate line 6):

6	OM	5; RC	$\{O_2M \wedge \neg OM\}\checkmark^8$
7	$\neg O(S \wedge P \wedge M)$	4; RU	$\emptyset$
8	$(O_1S \wedge \neg OS) \vee (O_2P \wedge \neg OP) \vee (O_2M \wedge \neg OM)$	1,2,3,7; RU	$\emptyset$

At stage 8 of the proof, the second element of line 8 is the only Dab-formula in the proof, hence it is a minimal Dab-formula. According to the definitions from Section 3.2, $\Sigma_8(\Gamma'_1) = \{\{O_1S \wedge \neg OS, O_2P \wedge \neg OP, O_2M \wedge \neg OM\}\}$. There are two $\sqsubset$-minimal choice sets of $\Sigma_8(\Gamma'_1)$: $\Phi^{\sqsubset}_8(\Gamma'_1) = \{\{O_2P \wedge \neg OP\}, \{O_2M \wedge \neg OM\}\}$. This implies that line 6 is marked at stage 8, which is indicated by the $\checkmark^8$-sign.

However, we can extend the proof such that line 6 becomes unmarked at a later stage:

6	OM	5; RC	$\{O_2M \wedge \neg OM\}$
7	$\neg O(S \wedge P \wedge M)$	4; RU	$\emptyset$
8	$(O_1S \wedge \neg OS) \vee (O_2P \wedge \neg OP) \vee (O_2M \wedge \neg OM)$	1,2,3,7; RU	$\emptyset$
9	$(O_1S \wedge \neg OS) \vee (O_2P \wedge \neg OP)$	1,3,4; RU	$\emptyset$

As a result, the formula on line 8 is not a minimal Dab-formula anymore. We get that $\Phi^{\sqsubset}_9(\Gamma'_1) = \{\{O_2P \wedge \neg OP\}\}$. In this particular case, line 6 will remain unmarked in every extension of the proof. Along the same lines, we can extend the proof to finally derive OS:

10	OS	1; RC	$\{O_1S \wedge \neg OS\}$

The formula at line 9 does no harm to this derivation, since $O_1S \wedge \neg OS$ is freed from suspicion, so to speak, by the abnormality $O_2P \wedge \neg OP$ – note that $\{O_2P \wedge \neg OP\} \sqsubset \{O_1S \wedge \neg OS\}$.

5.2 Mediocrity Does Not Rule

Consider the following $\mathbf{MP}_{\sqsubset}$-proof from $\Gamma_2' = \{O_1U, O_2D, O_3N, \neg O(U \wedge D), \neg O(N \wedge D)\}$:

1	O_1U	PREM	$\emptyset$
2	O_2D	PREM	$\emptyset$
3	O_3N	PREM	$\emptyset$
4	$\neg O(U \wedge D)$	PREM	$\emptyset$
5	$\neg O(N \wedge D)$	PREM	$\emptyset$
6	OU	1; RC	$\{O_1U \wedge \neg OU\}$
7	OD	2; RC	$\{O_2D \wedge \neg OD\}$ ✓[10]
8	ON	3; RC	$\{O_3N \wedge \neg ON\}$
9	$(O_1U \wedge \neg OU) \vee (O_2D \wedge \neg OD)$	1,2,4; RU	$\emptyset$
10	$(O_2D \wedge \neg OD) \vee (O_3N \wedge \neg ON)$	2,3,5; RU	$\emptyset$

Note that $\Sigma_{10}(\Gamma_2') = \{\{\sigma^1(U), \sigma^2(D)\}, \{\sigma^2(D), \sigma^3(N)\}\}$. We get that $\Phi_{10}^{\sqsubset}(\Gamma_2') = \{\{\sigma^2(D)\}\}$. Note that $\varphi = \{\sigma^1(U), \sigma^3(N)\}$ is not a $\sqsubset$-minimal choice set of $\Sigma_{10}(\Gamma_2')$, since $\{\sigma^2(D)\} \sqsubset \{\sigma^1(U), \sigma^3(N)\}$. As a result, only line 7 is marked. Line 6 will remain unmarked in every extension of this proof, whence OU is finally derivable: Michael has the obligation to visit his sick uncle. Since this removes the obligation to visit his daughter, we can also derive that Michael has to pass by his nephew (at line 8) – note that both U and N can be fulfilled, and all that kept Michael from visiting his nephew was the (now overridden) obligation to visit his daughter.

This example is instructive in that it shows a clear difference between $\mathbf{MP}_{\sqsubset}$ and the criterion of "Least Exposure" from Alchourrón and Makinsons [1]. According to this criterion, an obligation A is preferred if and only if it is an element of a maximal consistent subset Θ of the prima facie obligations, and every other maximal consistent subset Δ that does not contain A is more exposed. In our terminology, that Δ is more exposed than Θ means that Δ contains obligations with a higher priority index than any of the obligations in Θ.

Now consider the example. There are two maximal consistent subsets: $\{U, N\}$ and $\{D\}$. Since the latter is less exposed than the former, Alchourrón and Makinson's criterion yields D as a preferred obligation: Michael ought to visit his daughter, which also means that he cannot visit his uncle and his nephew. Hansen refers to this as the "Mediocrity Rules"-problem in his [13], and sees it as a severe drawback of Alchourrón and Makinson's criterion. This problem is avoided by $\mathbf{MP}_{\sqsubset}$: an obligation can only be suspended if it is involved in a deontic conflict, with obligations of the same or a higher priority level.

5.3 Relevance-Sensitivity

Consider Case 3, and its translation into $\Gamma_3 = \{O_3(T \wedge (F \vee S)), O_2\neg T, O_1\neg F\}$. Clearly, there is a conflict between the first and the second obligation. However, a specific property of the logic $\mathbf{MP}_{\sqsubset}$ is that it reduces such conflicts to conflicts between minimal disjunctions of literals. This is due to the fact that (i) **MP** validates modal inheritance (RM$_i$) and (ii) $\mathbf{MP}_{\sqsubset}$ is a strengthening of **MP** (see Theorem 7).

Let us explain what is going on, for this particular example. Note that $\Gamma_3 \vdash_{\mathbf{MP}} O_3T$ and $\Gamma_3 \vdash_{\mathbf{MP}} O_3(F \vee S)$. As a result, the obligations to do $F \vee S$ and the one to do $\neg F$ will not be considered problematic, since the real problem lies with T: $\Gamma_3 \vdash_{\mathbf{MP}} O_2\neg T \wedge O_3T$, hence also $\Gamma_3 \vdash_{\mathbf{MP}} (O_2\neg T \wedge \neg O\neg T) \vee (O_3T \wedge \neg OT)$. As a result, $\mathbf{MP}_{\sqsubset}$ considers only the atom T to be involved in the conflict, whereas $F \vee S$ and $\neg F$ are actualized.

We call this property relevance-sensitivity, echoing the concept of relevance in belief revision. A full comparison with relevant belief revision lies beyond the scope of the present paper, but let us briefly spell out the core intuitions.[11] A proto-typical example to explain the idea behind relevant belief revision is the revision of a belief base $\Delta = \{p \wedge q\}$ by the new information $A = \neg p$. A relevant belief revision $\Delta \dot{+} A$ is one that yields q. The underlying idea is that, notwithstanding the formulation of Δ, q is not relevant to the new information, whence it can be upheld without any problems. In a similar vain, a logic that deals with conflicting sets of obligations should be able to trace down those obligations that are relevant to the conflict and consider the other obligations as unproblematic, irrespective of the initial formulation of the whole set of obligations.

$\mathbf{MP}_{\sqsubset}$ has this property in common with the logic $\mathbf{P2.1}^r$ from [25]. Most of the existing criteria in the literature on prioritized information (beliefs, obligations, defaults) rely on (a selection among) the maximal consistent subsets from a (possibly inconsistent) base, and therefore depend quite heavily on the way this base is formulated. Examples are again Alchourrón and Makinson's Least Exposure, but also Brewka's preferred remainders, Nebel's prioritized removals, Prakken's criterion for hierarchic rebuttal and Sartor's "prevailing" relation.[12] More recently, the same basic idea was applied in Input/Output-logic by Boella and Van Der Torre, see [5]. Notwithstanding all the subtle differences between these systems, they have one thing in common: if some obligation is stated as a conjunction (such as e.g. $O_3(T \wedge (F \vee S))$), and one of the conjuncts is involved in a conflict, this renders the whole obligation useless.

In the case of the logic $\mathbf{MP}_{\sqsubset}$, we can derive the actual obligations to do $F \vee S$ and $\neg F$, and by these, we can derive the actual obligation to do S:[13]

1	$O_3(T \wedge (F \vee S))$	PREM	$\emptyset$
2	$O_2\neg T$	PREM	$\emptyset$
3	$O_1\neg F$	PREM	$\emptyset$
4	$O_3(F \vee S)$	1;RU	$\emptyset$
5	$O(F \vee S)$	4; RC	$\{\sigma^3(F, S)\}$

[11] See [28] where the concept of relevance was introduced, and [21, 18] for some fundamental results in this area. In [30], an adaptive logic for relevant belief revision is presented.

[12] See [13] for an overview of these consequence relations.

[13] According to the notational conventions from Section 3.2, $\sigma^3(F, S)$ abbreviates $(O_3(F \vee S) \wedge \neg O(F \vee S)) \vee (O_3F \wedge \neg OF) \vee (O_3S \wedge \neg OS)$.

6	$O\neg F$	3; RC	$\{O_1\neg F \wedge \neg O\neg F\}$
7	OS	5,6; RU	$\{\sigma^3(F,S), O_1\neg F \wedge \neg O\neg F\}$

As line 6 and 7 are not marked in any extension of this proof, $O\neg F$ and OS are finally $\mathbf{MP}_{\sqsubset}$-derivable from Γ_3. Likewise, we can finally derive $O\neg T$ in a $\mathbf{MP}_{\sqsubset}$-proof from Γ_3.

6 A Logic for Prioritized Beliefs

In this section, we will briefly show how $\mathbf{MP}_{\sqsubset}$ can be turned into an adaptive logic for (prioritized) belief base revision.[14] Prioritized belief bases are sequences of the form $\Psi = \langle \Theta_0, \Theta_1, \Theta_2, \ldots \rangle$, where each Θ_i is a set of formulas, and the index of the sets denotes their plausibility degree: Θ_0 is the set of most plausible beliefs, Θ_1 the set of second most plausible beliefs, and so on. In the face of such a belief base, we may ask ourselves which are the "all-things-considered, actual" beliefs. Roughly speaking, it is the aim of a prioritized revision operation to maintain consistency at the level of actual beliefs, but nevertheless keep as much of the initial information in Ψ as possible.

It is fairly easy to adapt $\mathbf{MP}_{\sqsubset}$ for this context. All we have to do is replace the operators O and $\langle O_i \rangle_{i\in\mathbb{N}}$ by their doxastic counterparts B and $\langle B_i \rangle_{i\in\mathbb{N}}$, adjust the respective interpretation of these operators, and provide a generic way to translate a prioritized belief base into a premise set.

The semantic structure of **MP** can be kept as it is: we have a set of sets of accessibility relations $\mathcal{R}_1, \mathcal{R}_2, \ldots$ – each of them corresponding to a B_i-operator – and a single accessibility relation R that corresponds to the B-operator. B_iA expresses that A has plausibility degree i.[15] As expected, B_iA does not entail $\neg B_i\neg A$. Hence the sets Θ_i from the belief base are not assumed to be internally consistent. The B-operator refers to the "actual, all-things-considered" beliefs of a rational agent. These beliefs behave classically, just as the actual obligations did in **MP**. Let us call the resulting monotonic logic **MPb**, which stands for "the **MP**-variant for belief bases". The preceding makes it quite obvious how the translation should proceed. Where Ψ is as before, we simply define $\Gamma_\Psi = \{B_iA \mid A \in \Theta_i, \Theta_i \in \Psi\}$.

Let us use Ω_i^b to denote the set obtained by replacing every occurrence of O_i by B_i, and every occurrence of O by B in each member of the set Ω_i. The adaptive logic $\mathbf{MPb}_{\sqsubset}$ has exactly the same structure as $\mathbf{MP}_{\sqsubset}$, but takes **MPb** as its lower limit logic and $\langle \Omega_i^b \rangle_{i\in\mathbb{N}}$ as its sequence of sets of abnormalities. One peculiar aspect of $\mathbf{MPb}_{\sqsubset}$, compared to other systems from the literature, is its dynamic proof theory. Another important property that is immediately carried over from $\mathbf{MP}_{\sqsubset}$, is that $\mathbf{MPb}_{\sqsubset}$ reduces contradictions between beliefs to the level of disjunctions of atoms – recall Section 5.3 where this was illustrated for the deontic variant. Finally, all the meta-theoretic results from Section 3.3 obviously hold for $\mathbf{MPb}_{\sqsubset}$ as well.

[14] See [38] for some other examples of adaptive logics that deal with prioritized belief bases.
[15] See [4, p. 474] for a discussion of how such degrees of plausibility may be justified.

7 Conclusion

Let us briefly summarize the main results of this paper. We have developed a logic $\mathbf{MP}_{\sqsubset}$ that deals with unconditional prioritized obligations and has a dynamic proof theory. We have described its main features and have established a number of intuitive properties of it. Concrete examples were presented, which illustrate the proof theory and highlight some differences with other approaches in the literature. Finally, we have provided an interpretation of a strucurally identical system $\mathbf{MPb}_{\sqsubset}$, which can be applied to prioritized belief bases.

We promised to say a bit more about the restrictions we made in the introduction, as removing one or more of these restrictions should be a task for future research. First of all, it might still be interesting to see if we can develop systems that give up the restriction that the order has to be modular. To do so would imply that we cross the safe boundaries of the existing generic formats for adaptive logics – all adaptive logics developed so far are based on a modular order of the abnormalities. This work will hence require thorough investigations on the metatheoretical level.

We also restricted ourselves to a monadic framework. The extension to a dyadic deontic logic might lead to some problems, e.g. should obligations with more specific condition receive a higher priority rank, or should contrary-to-duty obligations overrule conflicting unconditional obligations whenever the former are cancelled? We refer to [33, 34, 14] for discussions of the various problems and paradoxes relating to conditional obligations. An extension of $\mathbf{MP}_{\sqsubset}$ to the dyadic case should be able to cope with these issues to some extent.

References

[1] Carlos Alchourrón and David Makinson. Hierarchies of regulations and their logic. In *New Studies in Deontic Logic*. Risto Hilpinen, 1981.

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

[3] Mathieu Beirlaen, Joke Meheus, and Christian Straßer. An inconsistency-adaptive deontic logic for normative conflicts. Forthcoming., 201x.

[4] Salem Benferhat, Didier Dubois, and Henri Prade. Some syntactic approaches to the handling of inconsistent knowledge bases: A comparative study. Part 2: The prioritized case. In Ewa Orłowska, editor, *Logic at Work. Essays Dedicated to the Memory of Helena Rasiowa*, pages 473–511. Physica Verlag (Springer), Heidelberg, New York, 1999.

[5] Guido Boella and Leendert van der Torre. Institutions with a hierarchy of authorities in distributed dynamic environments. *Artifical Intelligence and Law*, 16:53–71, 2008.

[6] Newton C.A. da Costa and Walter Carnielli. On paraconsistent deontic logic. *Philosophia*, 16:293–305, 1986.

[7] Lou Goble. Deontic logic with relevance. In Paul McNamara and Henri Prakken, editors, *Norms, Logics and Information Systems. New Studies on Deontic Logic and Computer Science*, pages 15–25. IOS Press, 1999.

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

[9] Lou Goble. Preference semantics for deontic logic. Part I: Simple models. *Logique et Analyse*, 183–184:383–418, 2003.

[10] Lou Goble. Preference semantics for deontic logic. Part II: Multiplex models. *Logique et Analyse*, 185–188:335–363, 2004.

[11] 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.

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

[13] Jörg Hansen. Deontic logics for prioritized imperatives. *Artifical Intelligence and Law*, 14:1–34, 2006.

[14] Jörg Hansen. Prioritized conditional imperatives: problems and a new proposal. *Autonomous Agent Multi-Agent Systems*, 17:11–35, 2008.

[15] Jörg Hansen, Gabriella Pigozzi, and Leendert Van Der Torre. Ten philosophical problems in deontic logic. In *Normative Multi-agent Systems*, 2007.

[16] John F. Horty. Moral dilemmas and nonmonotonic logic. *Journal of Philosophical Logic*, 23:35–65, 1994.

[17] John F. Horty. Reasoning with moral conflicts. *Noûs*, 37:557–605, 2003.

[18] George Kourousias and David Makinson. Parallel interpolation, splitting, and relevance in belief change. *The Journal of Symbolic Logic*, 72:994–1002, 2007.

[19] Daniel Lehmann and Menachem Magidor. What does a conditional knowledge base entail? *Artificial Intelligence*, 55(1):1–60, 1992.

[20] Daniel J. Lehmann. Another perspective on default reasoning. *Annals of Mathematics and Artificial Intelligence*, 15(1):61–82, 1995.

[21] David Makinson. Propositional relevance through letter-sharing. *Journal of Applied Logic*, 7:377–387, 2009.

[22] David Makinson and Leendert Van Der Torre. Constraints for input-output logic. *Journal of Philosophical Logic*, 30:155–185, 2001.

[23] Ed Mares. Andersonian deontic logic. *Theoria*, 58:3–20, 1992.

[24] Terrance McConnell. Moral dilemmas. 2006. Online access `http://plato.stanford.edu/entries/moral-dilemmas/`.

[25] Joke Meheus, Mathieu Beirlaen, and Frederik Van De Putte. Avoiding deontic explosion by contextually restricting aggregation. In Guido Governatori and Giovanni Sartor, editors, *Deontic Logic in Computer Science*, pages 148 – 165. Springer, 2010.

[26] Joke Meheus, Mathieu Beirlaen, Frederik Van De Putte, and Christian Straßer. Non-adjunctive deontic logics that validate aggregation as much as possible. Submitted.

[27] Bernhard Nebel. Syntax-based approaches to belief revision. In Peter Gärdenfors, editor, *Belief Revision*, pages 52–88. Cambridge University Press, 1992.

[28] Rohit Parikh. Beliefs, belief revision, and splitting languages. *Logic, language, and computation*, 2:266–278, 1999.

[29] Graham Priest. *In Contradiction. A Study of the Transconsistent*. Oxford University Press, Oxford, 2006. Second expanded edition (first edition 1987).

[30] Frederik Van De Putte. The dynamics of relevance: Adaptive belief revision. Unpublished paper.

[31] W. David Ross. *The Right and the Good*. Clarendon Press, 1963 [1930].

[32] Peter K. Schotch and Raymond E. Jennings. Non-kripkean deontic logic. In Risto Hilpinen, editor, *New Studies in Deontic logic*, pages 149–162. Reidel, Dordrecht, 1981.

[33] Christian Straßer. An adaptive logic framework for conditional obligations and deontic dilemmas. *Logic and Logical Philosophy*, 200x. Forthcoming.

[34] Christian Straßer. A deontic logic framework allowing for factual detachment. *Journal of Applied Logic*, 2010. Online access `http://dx.doi.org/10.1016/j.jal.2010.11.001`.

[35] Frederik Van De Putte and Christian Straßer. Extending the standard format of adaptive logics to the prioritized case. Submitted.

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

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

[38] Liza Verhoeven. Proof theories for some prioritized consequence relations. *Logique et Analyse*, 183–184:325–344, 2003.

APPENDIX

A Proof of Theorem 1

Soundness is proven by the usual inductive procedure and is safely left to the reader.

A canonical model $M_c^\Delta = \langle W_c, \mathbf{R}, R, v_c, \Delta\rangle$ is defined as follows:[16]

(i) W_c is the set of maximal consistent extensions of **MP**;
(ii) $\mathbf{R} = \{\mathcal{R}_1, \mathcal{R}_2, \ldots\}$, where $\mathcal{R}_i = \{R_i^A \mid A \in \mathcal{W}\}$ and $R_i^A = \{(w, w') \in W_c \times W_c \mid O_iA \notin w \text{ or } A \in w'\}$;
(iii) $R = \{(w, w') \in W_c \times W_c \mid \text{ either } OA \notin w \text{ or } A \in w'\}$ for an $A \in \mathcal{W}$;
(iv) for every $w \in W_c, p \in \mathcal{S}$: $v_c(p, w) = 1$ iff $p \in w$;
(v) and Δ is an arbitrary element of W_c.

Note that by the definition $\mathcal{R}_i$ is non-empty for each i. Moreover, each $R_i^A \in \mathcal{R}_i$ is serial. Suppose that $O_iA \notin w$. Then $R_i^A ww'$ for all $w' \in W_c$. Suppose that $O_iA \in w$. Assume there is no $w' \in W_c$ such that $A \in w'$. Then, $\vdash_{\mathbf{MP}} \neg A$. By (P$_i$), $\vdash_{\mathbf{MP}} \neg O_iA$. Hence, $O_iA \notin w$, — a contradiction. Hence there is a $w' \in W_c$ for which $A \in w'$. Thus, $R_i^A ww'$. By the same argument R is serial. Altogether this shows that M_c^Δ is an **MP**-model.

Lemma 2 *For all $A \in \mathcal{W}^M$ and all $w \in W_c$, $v_{M_c^\Delta}(A, w) = 1$ iff $A \in w$.*

Proof. This is shown by an induction on the complexity of A. The argument is straightforward. We only show the induction step for $A = O_iB$, the rest is left to the reader. Suppose $O_iB \in w$. Note that $R_i^B \in \mathcal{R}_i$ is defined in such a way that if $R_i^B ww'$ then $B \in w'$ and whence by the induction hypothesis $v_{M_c^\Delta}(B, w') = 1$. Hence, $v_{M_c^\Delta}(O_iB, w) = 1$.

Suppose now that $v_{M_c^\Delta}(O_iB, w) = 1$. There is a $R_i^C \in \mathcal{R}_i$ such that if $R_i^C ww'$ then $v_{M_c^\Delta}(B, w') = 1$. Assume that $\{C, \neg B\}$ is consistent. Then there is a $w'' \in W_c$ for which $C, \neg B \in w''$. Hence $B \notin w''$. By the induction hypothesis $v_{M_c^\Delta}(B, w'') = 0$. Thus, $(w, w'') \notin R_i^C$. Whence, $O_iC \in w$ and $C \notin w''$,—a contradiction. Hence, $\vdash_{\mathbf{MP}} C \supset B$ and by (RM$_i$), $\vdash_{\mathbf{MP}} O_iC \supset O_iB$.

Suppose $O_iC \in w$, then $O_iB \in w$ by (MP). Suppose (†) $O_iC \notin w$. Assume $\{\neg B\}$ is consistent. Hence, there is a maximal consistent extension w''' of $\{\neg B\}$. Since by (†) $R_i^C ww'''$, also $v_{M_c^\Delta}(B, w''') = 1$. Thus, by the induction hypothesis, $B \in w'''$,—a contradiction. Hence, $\vdash_{\mathbf{MP}} B$. By (RN$_i$), $\vdash_{\mathbf{MP}} O_iB$. Hence $O_iB \in w$. ■

Lemma 3 *Every* **MP***-consistent set of sentences Γ is satisfiable in* **MP***.*

Proof. By Lindenbaum's Lemma there is a maximal consistent extension Γ' of Γ. By Lemma 2 and since $\Gamma \subseteq \Gamma'$, $M_c^{\Gamma'} \Vdash \Gamma$. ■

Theorem 11 *If $\Gamma \Vdash_{\mathbf{MP}} A$ then $\Gamma \vdash_{\mathbf{MP}} A$.*

Proof. Suppose $\Gamma \Vdash_{\mathbf{MP}} A$. Hence, $\Gamma \cup \{\neg A\}$ is not satisfiable in **MP**. Hence by Lemma 3, $\Gamma \cup \{\neg A\} \vdash_{\mathbf{MP}} A$. Whence $\Gamma \vdash_{\mathbf{MP}} A$. ■

[16] This is very much inspired by the canonical model defined by Lou Goble in [8, p. 126].

B Proof of Theorem 9

We prove the left-right direction and the right-left direction of Theorem 9 as two separate lemmas.

Lemma 4 *If $\Gamma_{\mathbb{O},\mathcal{C}}$ has* **MP***-models, then every $A \in \mathcal{O}$ is* **CL***-satisfiable and $\mathcal{C}$ is* **CL***-satisfiable.*

Proof. Suppose that for some $A \in \mathcal{O}$, A is not **CL**-satisfiable. Let $M = \langle W, \mathbf{R}, R, v, w_0 \rangle$ be an **MP**-model of $\Gamma_{\mathbb{O},\mathcal{C}}$ — we derive a contradiction. Note that $O_i A \in \Gamma_{\mathbb{O},\mathcal{C}}$, whence there is a $R_j \in \mathcal{R}_i$ such that, for all $w \in W$ such that $R_j w_0 w$, $v_M(A, w) = 1$. However, since A is not **CL**-satisfiable, $v_M(A, w') = 0$ for all $w' \in W$.

Suppose that $\mathcal{C}$ is not **CL**-satisfiable. Hence there are $A_1, \ldots, A_n \in \mathcal{C}$ such that $A_1 \wedge \ldots \wedge A_n$ is not **CL**-satisfiable. Since $\mathcal{C} = Cn_{\mathbf{CL}}(\mathcal{C})$, $A_1 \wedge \ldots \wedge A_n \in \mathcal{C}$. Let $B = A_1 \wedge \ldots \wedge A_n$. Let $M = \langle W, \mathbf{R}, R, v, w_0 \rangle$ be an **MP**-model of $\Gamma_{\mathbb{O},\mathcal{C}}$ — we derive a contradiction. Note that $\neg O \neg B \in \Gamma_{\mathbb{O},\mathcal{C}}$, whence there is a $w \in W$ such that Rw_0w and $v_M(B, w) = 1$. However, for every $w' \in W$, $v_M(B, w') = 0$ since B is not **CL**-satisfiable. ■

Lemma 5 *If every $A \in \mathcal{O}$ is* **CL***-satisfiable and $\mathcal{C}$ is* **CL***-satisfiable, then $\Gamma_{\mathbb{O},\mathcal{C}}$ has* **MP***-models.*

Proof. Suppose (1) every $A \in \mathcal{O}$ is **CL**-satisfiable and (2) $\mathcal{C}$ is **CL**-satisfiable. Let $M = \langle W, \mathbf{R}, R, v, w_0 \rangle$ be defined as follows:

(i) $W = \{w_A \mid A \in \mathcal{O}\} \cup \{w_{\mathcal{C}}, w_0\}$. For each $A \in \mathcal{O}$, w_A is a maximal consistent set in $\mathcal{L}$ that contains A – by (1), there is such a w_A for every $A \in \mathcal{O}$. $w_{\mathcal{C}}, w_0$ are maximal consistent sets such that $\mathcal{C} \subseteq w_{\mathcal{C}}, w_0$ – by (2), there are such $w_{\mathcal{C}}, w_0$.
(ii) $\mathcal{R}_i = \{R_A \mid A \in \mathcal{O}_i\}$, where $R_A = \{(w_0, w_A), (w_A, w_A)\}$.
(iii) $R = \{(w_0, w_{\mathcal{C}}), (w_{\mathcal{C}}, w_{\mathcal{C}})\}$.
(iv) For every $w \in W, p \in \mathcal{S}$: $v(p, w) = 1$ iff $p \in w$.

It is easy to see that M is a model of $\Gamma_{\mathbb{O},\mathcal{C}}$. Let $O_i A \in \Gamma_{\mathbb{O},\mathcal{C}}$. Then $A \in \mathcal{O}_i$. We have $M \Vdash O_i A$ iff $v_M(O_i A, w_0) = 1$ iff for some $R_i \in \mathcal{R}_i$, $v_M(A, w') = 1$ for all $w' \in W$ for which $R_i w_0 w'$. Note that $R_A = \{(w_0, w_A), (w_A, w_A)\} \in \mathcal{R}_i$ and that by the construction and the induction hypothesis $v_M(A, w_A) = 1$.

Let $\neg O \neg A \in \Gamma_{\mathbb{O},\mathcal{C}}$. Then $A \in \mathcal{C}$. We have $M \Vdash \neg O \neg A$ iff $v_M(O \neg A, w_0) = 0$ iff (there is a $w' \in W$ for which Rw_0w' and $v_M(\neg A, w') = 0$) iff $v_M(\neg A, w_{\mathcal{C}}) = 0$ iff $v_M(A, w_{\mathcal{C}}) = 1$. The latter holds by the construction. ■

C Proof of Theorem 10

We first prove three lemmas:

Lemma 6 *If $\Gamma_{\mathbb{O},\mathcal{C}}$ is* **MP***-satisfiable and conflict-free up to level n, then there is an* **MP***-model M' of $\Gamma_{\mathbb{O},\mathcal{C}}$ such that $Ab(M') \cap (\Omega_1 \cup \ldots \cup \Omega_n) = \emptyset$.*

Proof. Suppose $\Gamma_{\mathbb{O},\mathcal{C}}$ is **MP**-satisfiable and conflict-free up to level n. By Lemma 4, every $A \in \mathcal{O}$ is **CL**-satisfiable and $\mathcal{C}$ is **CL**-satisfiable. Let M be the model of $\Gamma_{\mathbb{O},\mathcal{C}}$ constructed in the proof of Lemma 5. We construct $M' = \langle W', \mathbf{R}', R', v', w_0 \rangle$ from M in the following way:

(i) $W' = \{w_A \mid A \in \mathcal{O}_i \text{ where } i > n\} \cup \{w'_{\mathcal{C}}, w_0\}$. $w'_{\mathcal{C}}$ is a maximal consistent extension (with respect to **CL**) of $\mathcal{C} \cup \mathcal{O}_1 \cup \ldots \cup \mathcal{O}_n$ – since $\Gamma_{\mathbb{O},\mathcal{C}}$ is conflict-free up to level n, there is such a $w'_{\mathcal{C}}$.

(ii) $R' = \{(w_0, w'_{\mathcal{C}}), (w'_{\mathcal{C}}, w'_{\mathcal{C}})\}$.

(iii) $\mathbf{R}' = \{\mathcal{R}_i \mid i > n\} \cup \{\mathcal{R}'_i \mid i \leq n\}$ where $\mathcal{R}'_i = \{R'\}$ for each $i \leq n$.

(iv) For every $w \in W', p \in \mathcal{S}$: $v'(p, w) = 1$ iff $p \in w$.

By the construction, M' (1) is an **MP**-model of $\Gamma_{\mathbb{O},\mathcal{C}}$ for which (2) $M' \Vdash O_iA$ iff $M' \Vdash OA$ for all $i \leq n$. The proof of (1) is analogous to the one above. Let $O_iA \in \Gamma_{\mathbb{O},\mathcal{C}}$. Then $A \in \mathcal{O}_i$. We have $M \Vdash O_iA$ iff $v_M(O_iA, w_0) = 1$ iff for some $R_i \in \mathcal{R}_i$, $v_M(A, w') = 1$ for all $w' \in W$ for which $R_i w_0 w'$.

Suppose first that $i > n$. Note that $R_A = \{(w_0, w_A), (w_A, w_A)\} \in \mathcal{R}_i$ and that by the construction $v_M(A, w_A) = 1$. Suppose now that $i \leq n$. Note that $R' = \{(w_0, w'_{\mathcal{C}}), (w'_{\mathcal{C}}, w'_{\mathcal{C}})\} \in \mathcal{R}_i$ and that by the construction $v_M(A, w'_{\mathcal{C}}) = 1$. Let $\neg O \neg A \in \Gamma_{\mathbb{O},\mathcal{C}}$. Then $A \in \mathcal{C}$. We have $M \Vdash \neg O \neg A$ iff $v_M(O\neg A, w_0) = 0$ iff (there is a $w' \in W$ for which Rw_0w' and $v_M(\neg A, w') = 0$) iff $v_M(\neg A, w'_{\mathcal{C}}) = 0$ iff $v_M(A, w_{\mathcal{C}}) = 1$. The latter holds by the construction.

Ad (2): By the construction, $M \Vdash OA$ iff $v_M(OA, w_0) = 1$ iff $v_M(A, w'_{\mathcal{C}}) = 1$ iff $v_M(O_iA, w_0) = 1$ for all $i \leq n$. The latter holds since $\mathcal{R}_i = \{R'\}$ for all $i \leq n$. ■

Lemma 7 *If $\Gamma_{\mathbb{O},\mathcal{C}}$ is **MP**-satisfiable and conflict-free up to level n, then for every $\mathbf{MP}_{\sqsubset}$-model M of $\Gamma_{\mathbb{O},\mathcal{C}}$: $Ab(M) \cap (\Omega_1 \cup \ldots \cup \Omega_n) = \emptyset$.*

Proof. Suppose $\Gamma_{\mathbb{O},\mathcal{C}}$ is **MP**-satisfiable and conflict-free up to level n. Let $M \in \mathcal{M}_{\mathbf{MP}}(\Gamma_{\mathbb{O},\mathcal{C}})$ be such that $Ab(M) \cap (\Omega_1 \cup \ldots \cup \Omega_n) \neq \emptyset$. Let M' be the model constructed in Lemma 6. Then in view of Definition 1, $Ab(M') \sqsubset Ab(M)$, whence $M \notin \mathcal{M}_{\mathbf{MP}}(\Gamma_{\mathbb{O},\mathcal{C}})$. ■

Lemma 8 *If $\Gamma_{\mathbb{O},\mathcal{C}} \vdash_{\mathbf{MP}_{\sqsubset}} O_iA$, then $\Gamma_{\mathbb{O},\mathcal{C}} \vdash_{\mathbf{MP}_{\sqsubset}} OA \vee Dab(\Delta)$ for a $\Delta \subset \Omega_i$.*

Proof. Let $\Gamma = \Gamma_{\mathbb{O},\mathcal{C}}$. Suppose $\Gamma \vdash_{\mathbf{MP}_{\sqsubset}} O_iA$. By the reflexivity of **MP**, $Cn_{\mathbf{MP}_{\sqsubset}}(\Gamma) \vdash_{\mathbf{MP}} O_iA$. Let $\bigwedge_J \bigvee \Theta_j$ be a conjunctive normal form of A. By Lemma 1(ii) we get that $Cn_{\mathbf{MP}_{\sqsubset}}(\Gamma) \vdash_{\mathbf{MP}} OA \vee \bigvee_J \sigma^i(\Theta_j)$ and by Theorem 7, $Cn_{\mathbf{MP}_{\sqsubset}}(\Gamma) \vdash_{\mathbf{MP}_{\sqsubset}} OA \vee \bigvee_J \sigma^i(\Theta_j)$. By Theorem 5, $\Gamma \vdash_{\mathbf{MP}_{\sqsubset}} OA \vee \bigvee_J \sigma^i(\Theta_j)$. Since $\{\sigma^i(\Theta_j) \mid j \in J\} \subset \Omega_i$, the lemma follows immediately. ■

The last two lemmas make the proof of Theorem 10 rather short. Suppose $\Gamma_{\mathbb{O},\mathcal{C}}$ is conflict-free up to level n and $\Gamma_{\mathbb{O},\mathcal{C}} \vdash_{\mathbf{MP}_{\sqsubset}} O_iA$, where $i \leq n$. Then by Lemma 8, $\Gamma_{\mathbb{O},\mathcal{C}} \vdash_{\mathbf{MP}_{\sqsubset}} OA \vee Dab(\Delta)$ for a $\Delta \subset \Omega_i$. By the soundness of $\mathbf{MP}_{\sqsubset}$ (Theorem 2), for every $\mathbf{MP}_{\sqsubset}$-model M of $\Gamma_{\mathbb{O},\mathcal{C}}$: (†) $M \Vdash OA \vee Dab(\Delta)$. There are two cases to consider:

Case 1. $\Gamma_{\mathbb{O},\mathcal{C}}$ is not **MP**-satisfiable. Then it immediately follows that $\Gamma_{\mathbb{O},\mathcal{C}}$ is $\mathbf{MP}_{\sqsubset}$-trivial, whence $\Gamma_{\mathbb{O},\mathcal{C}} \vdash_{\mathbf{MP}_{\sqsubset}} OA$ for all $A \in \mathcal{W}$.

Case 2. $\Gamma_{\mathbb{O},\mathcal{C}}$ is **MP**-satisfiable. Then for every $M' \in \mathcal{M}_{\mathbf{MP}_{\sqsubset}}(\Gamma)$, $M' \Vdash \neg Dab(\Delta)$ in view of Lemma 7. Hence $M \Vdash \neg Dab(\Delta)$, which implies by (†): $M \Vdash OA$. By the completeness of $\mathbf{MP}_{\sqsubset}$ (Theorem 2), $\Gamma_{\mathbb{O},\mathcal{C}} \vdash_{\mathbf{MP}_{\sqsubset}} OA$.