On the Two Faces of Deontics: Semantic Betterness and Syntactic Priority

Johan van Benthem[1 2], Davide Grossi[1], Fenrong Liu[3]

[1] Institute for Logic, Language and Computation, University of Amsterdam
Science Park 904, 1098 XH, Amsterdam, The Netherlands
{j.vanbenthem,d.grossi}@uva.nl
[2] Department of Philosophy, Stanford University
Stanford, California 94305, USA
[3] Department of Philosophy, Tsinghua University
100084, Beijing, China
fenrong@tsinghua.edu.cn

Abstract. This paper looks at deontic logic as resulting from both a betterness ordering on states (i.e., a 'deontic preference') and a priority ordering on properties (i.e., a 'law' explicitly representing a standard of behavior). The correspondence between these two orderings offers a rich perspective from which to look at deontic scenarios and puzzles, and in particular at contrary-to-duties. The framework naturally lends itself to describing dynamics involving both orderings, thereby providing a new analysis of norm change as 'betterness change'.

Key words: Deontic logic, preference logic, modal logic.

1 Introduction

The present paper re-explores established ideas developed for the preference-based semantics of deontic logic in the light of recent studies in the modal logic of preference (in particular, [35, 21]). It moves from the well-known semantics for dyadic obligation first introduced in [27], where dyadic obligations of the type "it is obligatory that φ under condition ψ" are interpreted by making use of an 'ideality relation' and the notion of maximality:

$$\mathcal{M}, s \models \mathbf{O}(\varphi \mid \psi) \iff Max(\llbracket \psi \rrbracket_{\mathcal{M}}) \subseteq \llbracket \varphi \rrbracket_{\mathcal{M}} \tag{1}$$

where $\llbracket . \rrbracket_{\mathcal{M}}$ denotes the truth-set function of $\mathcal{M}$ and $\mathcal{M}$ is a model built on a Kripke frame $\mathcal{F} = (S, \preceq)$. In this frame the states in S are ordered according to the ideality relation $\preceq$. Formula 1 builds on the observation—stemming from ethical theory—that the deontic notions of obligation, permission and prohibition can be naturally made sense of in terms of an "ideality" ordering $\preceq$ on possible worlds:

> " [...] to assert that a certain line of conduct is [...] absolutely right or obligatory, is obviously to assert that more good or less evil will exist in the world, if it is adopted, than if anything else be done instead." [39][4]

[4] Cited in [55, p. 6].

Depending on the properties of $\preceq$, different logics can be obtained. In particular, [27] starts with a $\preceq$ which is only reflexive, moving then to total pre-orders, i.e., binary relations that are reflexive, transitive and connected.[5]

From the end of the Sixties, the very same idea behind Formula 1 reappears in many branches of philosophical logic like, eminently, conditional logic and doxastic logic. In conditional logic, the expression $Max(\llbracket\psi\rrbracket_{\mathcal{M}}) \subseteq \llbracket\varphi\rrbracket_{\mathcal{M}}$ has been used to give semantics to counterfactual conditionals $\psi \Rightarrow \varphi$ [48, 33], and in doxastic logic, to give semantics to conditional beliefs $\mathbf{B}(\varphi \mid \psi)$.

Although Formula 1 was an object of many criticisms,[6] variations gave rise to several studies in the preference-based semantics of deontic logic, which enjoyed considerable attention up till the Nineties.[7] So, the present paper provides a "fresh look at an old idea" and offers an attempt at reviving such tradition by reinterpreting it in the light of recent developments in the modal logic of preference. More specifically, the insights given by Formula 1, will be extended by developing the following points: i) the ideality order—which we will call "betterness" relation—can be fruitfully viewed as generated by a set of explicitly prioritized criteria; ii) both the betterness relation and the priorities on criteria naturally support a dynamic point of view, giving rise to a richer system of deontic dynamics as based on preference logic. This latter point also brings deontic logic closer to the family of dynamic logics of belief and preference change [51].

Outline of the paper. The paper extends results presented in [52]. It is structured as follows. Section 2 introduces the machinery of priority sequences to provide an original account of CTDs. Section 3 presents a modal logic of "betterness" for talking about the sort of orders arising from such sequences. Section 4 capitalizes on the first sections and discusses the type of perspective on deontics obtained by juxtaposing the 'semantic' view of deontics yelded by betterness relations with the 'syntactic' view of deontics yielded by priority sequences (Theorem 2 and Corollary 1). Section 5 applies techniques introduced in [35] to study betterness dynamics and identifies a correspondence between 'syntactic' normative changes, and 'semantic' ones at the level of the betterness relation (Theorem 3 and Corollary 2). Section 6 pursues the insights of Section 5 further by providing a formal analysis of the problem of norm merging (Corollary 3) and generalizing the notion of priority sequence to the notion of priority graph. Finally, Section 7 concludes.

[5] For an overview on the early model-theoretic work on dyadic obligation we refer the reader to [34].

[6] One criticism is that Formula 1 makes conditional obligations lack the property of antecedent strengthening (see [50]). This, however, makes perfect sense in our view as it is precisely what needs to follow from the idea of "most ideal worlds".

[7] See [53] for an overview.

2 Priorities and betterness

The betterness relation between states involved in the preference-based semantics of obligations (Formula 1) is, often, a sort of betterness derived from some kind of explicit "coding" of what is better in terms of relevant properties, as the following quote from St. Paul illustrates in a lively manner:

> "It is good for a man not to touch a woman. But if they cannot contain, let them marry: for it is better to marry than to burn." [46, Ch. 7][8]

This is, in the terminology of deontic logic, a typical contrary-to-duty structure [43] expressing what states are best, what states are best among the non-best ones, and so on, up to a finite depth. In the terminology of conditional logic it is a Lewis system of spheres [33], in the terminology of doxastic logic it is a Grove model [25]. In the following sections we will briefly discuss this type of structures in the light of notions and results developed in [35, Ch. 3], and illustrate them by formalizing a classical example of CTD obligations.

2.1 Priority sequences

This section introduces the key notions we will work with in the paper.

Definition 1 (P-sequence). *Let $\mathcal{L}(\mathbf{P})$ be a propositional language built on the set of atoms $\mathbf{P}$, S a non-empty set of states and $\mathcal{I} : \mathbf{P} \longrightarrow 2^S$ a valuation function. A P-sequence for $\mathcal{I}$ is a tuple $\mathcal{B}^{\mathcal{I}} = \langle B, \prec \rangle$ where:*[9]

- *$B \subset \mathcal{L}(\mathbf{P})$ with $|B| < \omega$;*
- *$\prec$ is a strict linear order on B, i.e., an irreflexive, transitive and total binary relation;*
- *for all $\varphi, \psi \in B$, if $\varphi \prec \psi$ then $[\![\psi]\!]_{\mathcal{I}} \subset [\![\varphi]\!]_{\mathcal{I}}$*

where $[\![\varphi]\!]_{\mathcal{I}}$ denotes the truth-set of φ according to $\mathcal{I}$. [10] *The set of all P-sequences for $\mathcal{I}$ is denoted $\mathbb{B}^{\mathcal{I}}$. Given a P-sequence $\mathcal{B}^{\mathcal{I}}$ for $\mathcal{I}$ denote with $Max(\mathcal{B}^{\mathcal{I}})$ the maximum element of $\mathcal{B}^{\mathcal{I}}$. Also, we denote with $Max^{+}(\mathcal{B}^{\mathcal{I}})$ the maximum element of $\mathcal{B}^{\mathcal{I}}$ which has a non-empty denotation according to $\mathcal{I}$, if it exists, or $\top$ otherwise.*

In other words, a priority sequence is a finite chain of distinct propositional formulae $\varphi_n \prec \ldots \prec \varphi_1$ from a language $\mathcal{L}(\mathbf{P})$ whose denotations form a finite ascending chain of sets $[\![\varphi_1]\!]_{\mathcal{I}} \subset \ldots \subset [\![\varphi_n]\!]_{\mathcal{I}}$.[11] For this reason they can be

[8] This passage is cited in [55, p. 6].

[9] When no confusion arises we will often drop the superscript in $\mathcal{B}^{\mathcal{I}}$.

[10] Alternatively, we could explicitly define $\prec$ in terms of $\subset$ and impose totality, i.e.: $\prec \; := \; \supset|_B$ (where $\supset|_B$ denotes the restriction of $\supset$ to the elements in B) with, in addition, $\prec$ being total. This definition is sound as the third condition in Definition 1 already imposes that $\prec \; \subseteq \; \supset|_B$, and this, together with the totality of $\prec$, enforces also that $\prec \; \supseteq \; \supset|_B$.

[11] When no confusion arises we will often drop the subscript in $[\![\varphi]\!]_{\mathcal{I}}$.

referred to as lists $\varphi_1, \ldots, \varphi_n$ of elements with $|B| = n$.[12] It is worth stressing that a P-sequence is always relative to a valuation function. If $[\![\varphi_1]\!]_{\mathcal{I}}$ and $[\![\varphi_2]\!]_{\mathcal{I}}$ are incomparable with respect to set-theoretic inclusion, then they cannot be part of the same P-sequence for $\mathcal{I}$.[13]

We now define a simple way to order states according to a given P-sequence.

Definition 2 (Deriving betterness from P-sequences). *Let $\mathcal{B} = \langle B, \prec \rangle$ be a P-sequence, S a non-empty set of states and $\mathcal{I} : \mathbf{P} \longrightarrow 2^S$ a valuation function. The preference relation $\preceq_{\mathcal{B}}^{IM} \subseteq S^2$ is defined as follows:*

$$s \preceq_{\mathcal{B}}^{IM} s' := \forall \varphi \in B : s \in [\![\varphi]\!] \Rightarrow s' \in [\![\varphi]\!]. \tag{2}$$

where IM is just a mnemonics for 'implication'. Given a P-sequence $\mathcal{B}$ for a valuation $\mathcal{I}$, Formula 2 generates also a Kripke model $\mathcal{M}_{\mathcal{B}}^{IM} = \langle S, \preceq_{\mathcal{B}}^{IM}, \mathcal{I} \rangle$.

Intuitively, Definition 2 orders states in S according to which elements of the P-sequence they satisfy. If a state satisfies a property in the sequence, then it also satisfies, by Definition 1, all $\prec$-worse properties in the sequence. We therefore obtain an order on states with the following properties.

Fact 1 (Properties of $\preceq_{\mathcal{B}}^{IM}$) *Let $\mathcal{B} = \langle B, \prec \rangle = (\varphi_1, \ldots, \varphi_n)$ be a P-sequence for $\mathcal{I} : \mathbf{P} \longrightarrow 2^S$. It holds that:*

1. *Relation $\preceq_{\mathcal{B}}^{IM}$ is a total pre-order[14] whose strict part $\prec_{\mathcal{B}}^{IM}$ is conversely well-founded;*[15]
2. *If $\varphi_i \prec \varphi_j$ then for all $s \in [\![\varphi_i]\!], s' \in [\![\varphi_j]\!]$: $s \preceq_{\mathcal{B}}^{IM} s'$;*
3. *If $\varphi_i \prec \varphi_j$ then for all $s \in [\![\varphi_i \wedge \neg\varphi_j]\!], s' \in [\![\varphi_j]\!]$: $s \prec_{\mathcal{B}}^{IM} s'$.*

Proof. The first claim is straightforward as to reflexivity, transitivity and connectedness. As to converse well-foundedness, note that P-sequences are finite. So, if the cardinality of a P-sequence is n, it generates a total-preorder consisting of $n+1$ clusters of equally good states. Hence the strict part of the pre-order contains only bounded chains. The second and third claims follow directly from Definitions 1 and 2. □

The following is worth noticing. Given a P-sequence $\varphi_n \prec \ldots \prec \varphi_1$, the $\preceq_{\mathcal{B}}^{IM}$-minimal states in S are the one satisfying $\neg\varphi_n$, if such states exists. In fact, it can be the case that $[\![\varphi_n]\!] = S$, i.e., $\varphi_n = \top$. In such a case, the $\preceq_{\mathcal{B}}^{IM}$-minimal states are therefore the states satisfying $\neg\varphi_{n-1}$. This suggests that any

[12] It might be instructive to notice that Definition 1 could be restated by requiring the elements of the sequence to be disjoint in $\mathcal{I}$, instead of being ordered according to a finite $\subseteq$-chain (see [35] for further details).

[13] With respect to this it might be instructive to stress that the third condition in Definition 1 could in principle be lifted allowing for P-sequences among logically unrelated formulae. The presence of such condition is dictated by the applications to deontic logic which we are going to deal with in the next sections.

[14] That is, a relation which is reflexive, transitive and connected.

[15] That is, it does not contain infinitely ascending chains.

$\varphi_n \prec \ldots \prec \varphi_1$ P-sequence such that $[\![\varphi_n]\!] \neq S$ could be completed to a sequence $\varphi_{n+1} \prec \varphi_n \prec \ldots \prec \varphi_1$ where $\varphi_{n+1} = \top$.

A further technical discussion of Definitions 1 and 2 is provided in the appendix. Before moving to the deontic applications of P-sequences in the next section, however, we find it worth making a few remarks in order to place our priority-based study of deontics within a broader family of knowledge representation techniques. What Definitions 1 and 2 (or their variants given in the appendix) allow for is to use compact information of a syntactic kind—the P-sequence—to represent richer semantic information obtainable by means of a suitable recipe—the procedure to derive betterness relations. Representation strategies of this sort have a long tradition in knowledge representation, e.g., 'inheritance hierarchies' and 'semantic networks' with their relations to first-order logic [57] and default logic [18], or the entanglement relation in belief revision [40].

2.2 P-sequences and CTDs

The example below is a 'classic' of deontic logic and illustrates in a straightforward way the problem of CTDs [19]. We show how it can be easily represented by means of P-sequences.

Example 1 (Gentle murder). Here is the example:

> "Let us suppose a legal system which forbids all kinds of murder, but which considers murdering violently to be a worse crime than murdering gently. [...] The system then captures its views about murder by means of a number of rules, including these two:
> 1. It is obligatory under the law that Smith not murder Jones.
> 2. It is obligatory that, if Smith murders Jones, Smith murders Jones gently." [19, p. 194]

The scenario makes explicit two classes of ideality: a class (let us call it I_1) in which Smith does not murder Jones, i.e. $\mathsf{I}_1 := \neg m$; another one (let us call it I_2), in which either Smith does not murder Jones or he murders him gently, i.e., $\mathsf{I}_2 := \neg m \vee (m \wedge g)$. We thus have a P-sequence $\mathcal{B}$ such that $\mathsf{I}_2 \prec \mathsf{I}_1$. Such P-sequence is sufficient to order the states—according to the corresponding $\preceq^{IM}_{\mathcal{B}}$ relation—in three clusters such that the most ideal states are the ones satisfying I_1, the worse but not worst states are the ones that satisfy $\mathsf{V}_1 := \neg \mathsf{I}_1$ but at the same time I_2 and, finally, the worst states are the ones satisfying $\mathsf{V}_2 := \neg \mathsf{I}_2$ (and hence V_1 too). See Figure 1.

To sum up, the intuition behind a P-sequence $p_1, \ldots, p_n$ for a given interpretation function is that each atom p_i gives rise to a bipartition $\{\mathcal{I}(p_i), -\mathcal{I}(p_i)\}$ of the domain of discourse S, and the more we move towards the right-hand side (i.e., the bottom) of the sequence the more atoms p_i are falsified. As shown by Example 1, in a deontic reading this simply means that, the more we move towards the right-hand side of the sequence the more violations hold.

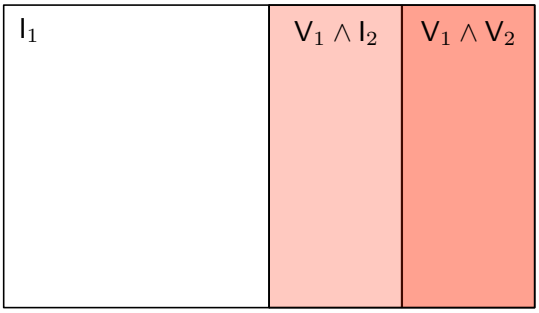

Fig. 1. Gentle Murder as an ordered, labelled tripartition of the state space. Th rectangle represents the state space and is partitioned in clusters of equally preferred states. Clusters are ordered from more to less preferred from left to right.

This point of view is in line with several approaches in deontic logic (e.g., [31, 43]) which have stressed the importance of acknowledging a sub-ideality component in order to provide a satisfactory formalization of CTDs. In particular, in [31] models are considered where the standard deontic accessibility relation is bipartitioned into an 'ideal' and a 'sub-ideal' part. As a result, each states has ideals and sub-ideals alternatives. Modulo some secondary technical details (mainly concerning the properties the accessibility relation is taken to satisfy), such models could be seen as the ones resulting from the P-sequence $\langle \mathsf{I} \rangle$ consisting of one single ideality atom. So P-sequences push this intuition further allowing not only for the representation of ideality and sub-ideality relations, but also for the representation of sub-sub-ideality relations, at any finite depth.

One last group of approaches to CTDs worth mentioning here are the ones that have argued for the application of nonmonotonic logics (e.g., default logic like in [29, 38] or argumentation theory [42]) to deontic logic problems. The relation of our priority/betterness-based approach with these contributions is quite direct, and lies on the simple observation that the betterness relation works exactly like the plausibility relations in preference-based analyses of non-monotonic reasoning and default reasoning in particular (cf. [45]). As observed in [54], the sort of defeasibility involved in CTDs, which is there called *factual defeasibility*, consists in 'factual' assumptions determining a shift in the sphere of most ideal states compatible with the assumption. So, in the gentle murder case, the assumption "Smith murders Jones" shifts ideality from "Smith does not murder Jones" to "Smith murders Jones gently". We will come back to the nonmonotonicity of CTDs and its formalization in Section 3.

2.3 *"To make the best of sad circumstances"*

Although Example 1 has nicely illustrated how a CTD structure can be rendered by a P-sequence, it still remains to be shown how the basic CTD reasoning operates on such structures. The idea is to express how "to make the best out of

sad circumstances" [27]. On a P-sequence this means, intuitively, take the best states that survive the "sad circumstances". To make this precise we have to introduce the following refinement of Definition 1, which relativizes the notion of P-sequence to the occurrence of given circumstances.

Definition 3 (Restricted P-sequences). *Let $\mathcal{B} = \langle B, \prec \rangle$, with $|B| = n$ and $B \subset \mathcal{L}(\mathbf{P})$, be a P-sequence for $\mathcal{I} : \mathbf{P} \longrightarrow 2^S$. The restriction of $\mathcal{B}$ to a formula ψ of $\mathcal{L}(\mathbf{P})$ is a structure $\mathcal{B}^\psi = \langle B^\psi, \prec^\psi \rangle$ where:*

- $B^\psi := \{\varphi_i \wedge \psi \mid \varphi_i \in B\}$;
- $\prec^\psi := \{(\varphi_i \wedge \psi, \varphi_j \wedge \psi) \mid (\varphi_i, \varphi_j) \in \prec\}$.

Given a restricted P-sequence $\mathcal{B}^\psi$ for $\mathcal{I}$, we denote with $Max(\mathcal{B}^\psi)$ the maximum element of $\mathcal{B}^\psi$. Also, we denote with $Max^+(\mathcal{B}^\psi)$ the maximum element of $\mathcal{B}^\psi$ which has a non-empty denotation according to $\mathcal{I}$, if it exists, or ψ otherwise.

Intuitively, the restriction of a P-sequence with respect to (the interpretation) of a formula ψ simply intersects the elements of the original P-sequence with ψ and keeps the original linear order. The result of such operation bears effects for the Max^+ of the P-sequence. Typically, $Max^+(\mathcal{B})$ might differ from $Max^+(\mathcal{B}^\psi)$ as $[\![1_\mathcal{B} \wedge \psi]\!]$ (i.e., the intersection of the maximum element of $\mathcal{B}$ with ψ) might be empty. Notice that if all elements in $\mathcal{B}^\psi$ turn out to be empty for a given $\mathcal{I}$, $Max^+(\mathcal{B}^\psi)$ is taken to be ψ itself (cf. Definition 1).

Example 2 (Gentle murder (continued)). Consider the P-sequence of the Gentle murder given in Example 1: $\mathcal{B} = (\mathsf{I}_1, \mathsf{I}_2)$. The restricted P-sequence $\mathcal{B}^{\mathsf{V}_1}$ is then $(\mathsf{I}_1 \wedge \mathsf{V}_1, \mathsf{I}_2 \wedge \mathsf{V}_1)$. In such a sequence the top element has necessarily an empty denotation. This means that the best among the still available states are the states $Max^+(\mathcal{B}^{\mathsf{V}_1}) = \mathsf{I}_2 \wedge \mathsf{V}_1$ (see Definition 3). Another interesting restricted P-sequence in the Gentle murder context is $\mathcal{B}^{\mathsf{V}_2}$, which describes what the original P-sequence prescribes under the assumption that also the CTD obligation "kill gently" has been violated. In this case $Max^+(\mathcal{B}^{\mathsf{V}_2}) = \mathsf{V}_2$, that is to say, if also the last CTD obligation has been violated, then we end up in a set of all equally bad states. This illustrates a characteristic feature of all finite CTD structures.

Stated more positively, our approach, including the logical elaboration to follow, provides a simple perspective the robustness of norms and laws viewed as CTD structures: they can still function when transgressions have taken place.

Other major deontic puzzles such as the Chisholm Paradox can be given an analogous representation, which we will do in Section 4.3. Before moving to the next section, it is worth mentioning that representing CTD structures as finite chains of properties is not a new idea in the literature on deontic logic. To the best of our knowledge, this idea was first adumbrated informally in [55]. The first formal account of CTD structures as sequences of formulae is to be found in [22], where an elegant Gentzen calculus is developed for handling formulae of the type $\varphi_1 \otimes \ldots \otimes \varphi_n$ with $\otimes$ a connective representing a sort of "sub-ideality" relation in a CTD structure. Unlike this proof-theoretic approach, our approach is geared towards semantics and aims at connecting such CTD structures to

modal logics interpreted on orders, and ultimately to conditional obligations in the standard maximality-based semantics (Section 3).[16]

3 A modal logic of betterness

This section presents a very simple language for talking about total pre-orders.

3.1 Language, semantics, axiomatics

Language. Language $\mathcal{L}(\forall, \preceq)$ is built from a countable set $\mathbf{P}$ of atoms according to the following BNF:

$$\mathcal{L}(\forall, \preceq) : \varphi ::= p \mid \top \mid \neg\varphi \mid \varphi \wedge \varphi \mid [\preceq]\varphi \mid [\forall]\ \varphi$$

where $p \in \mathbf{P}$. Modal duals and Boolean operators are defined as usual. Intuitively, $[\preceq]$ quantifies over all states which are at least as good as the current one, and $[\forall]$ over all states (universal modality).

Semantics. We define now the models and the satisfaction relation for the language just introduced.

Definition 4 (Models). *A model for $\mathcal{L}(\forall, \preceq)$ on the set of atoms $\mathbf{P}$ is a tuple $\mathcal{M} = \langle S, \preceq, \mathcal{I}\rangle$ where:*

- *S is a non-empty set of states;*
- *$\preceq$ is a conversely well-founded total preorder over S;*[17]
- *$\mathcal{I} : \mathbf{P} \longrightarrow 2^S$.*

As usual, we define $s \prec s'$ as $s \preceq s'$ and $s' \not\preceq s$.

Definition 5 (Satisfaction). *Let $\mathcal{M} \in \mathbb{M}$. The satisfaction of a formula $\varphi \in \mathcal{L}(\forall, \preceq)$ by a pointed model $(\mathcal{M}, s)$ is inductively defined as follows:*

$$\begin{aligned}
\mathcal{M}, s \models p &\iff w \in \mathcal{I}(p)\\
\mathcal{M}, s \models [\preceq]\varphi &\iff \forall s' \in S \text{ s.t. } s \preceq s' : \mathcal{M}, s' \models \varphi\\
\mathcal{M}, s \models [\forall]\ \varphi &\iff \forall s' \in S : \mathcal{M}, s' \models \varphi
\end{aligned}$$

The standard Boolean clauses are omitted.

Intuitively, $[\forall]$ -formulae state properties that hold in all states in the model (global properties), and $[\preceq]$-formulae state properties that hold in all the states that are at least as good as the evaluation state (i.e., in order-theoretic terms, the upset of the evaluation state).

[16] With respect to this, the interesting question arises of whether the Gentzen calculus developed in [22] is complete for our semantics, or whether it could be embedded in the logic to be presented in Section 3.

[17] This means that $\preceq$ is a binary relation which is: reflexive, transitive, connected, and such that it has no infinite ascending chain.

Axiomatics. The logic is axiomatized as follows, where $i \in \{\preceq, \forall\}$:

(**Prop**)	propositional tautologies
(**K**)	$[i](\varphi_1 \to \varphi_2) \to ([i]\varphi_1 \to [i]\varphi_2)$
(**T**)	$[i]\varphi \to \varphi$
(**4**)	$[i]\varphi \to [i][i]\varphi$
(**5**)	$\neg[\forall]\varphi \to [\forall]\neg[\forall]\varphi$
(**.3**)	$(\langle\preceq\rangle\varphi \wedge \langle\preceq\rangle\psi) \to (\langle\preceq\rangle(\varphi \wedge \langle\preceq\rangle\psi) \vee \langle\preceq\rangle(\varphi \vee \psi) \vee \langle\preceq\rangle(\psi \wedge \langle\preceq\rangle\varphi))$
(**Incl**)	$[\forall]\varphi \to [\preceq]\varphi$
(**Dual**)	$\langle i\rangle\varphi \leftrightarrow \neg[i]\neg\varphi$

The logic is just an extension of **S4.3** with the universal modality.

Theorem 1 (Strong completeness). *The logic above is sound and strongly complete with respect to the class of total pre-orders.*

Proof (Sketch). Completeness can be proven via the standard canonical model technique by exploiting the fact that point-generated sub-frames preserve modal validity (Cf. [7]). This allows us to work on a point-generated $[\forall]$ -submodel of the canonical model which by **Incl** the generated $[\preceq]$-submodel, and where $[\forall]$ is the universal modality and $[\preceq]$ is reflexive, transitive and, by rootedness and **.3**, it is also connected. The rest of the proof is standard. □

Modal logics of order. It might be instructive to mention that the logic just presented, which is essentially an extension of systems studied in [35, 21], falls into the broader line of research, initiated by [12, 13], which uses modal languages to represent and reason about ordered structures as they arise in the analysis of the notions of preference, normality, plausibility and affine ones. The main macroscopic difference between the two approaches resides in the use of the universal modality. In [12, 13], this modality is defined via two modalities. The first one is equivalent to our $[\preceq]$. The second one—let us denote it $[\succ]$—quantifies over the downset of the current state (current state excluded). So we get that, $[\forall]\ \varphi := [\preceq]\varphi \wedge [\succ]\varphi$. This definition does not yield the universal modality itself, as one can easily notice, but an **S5** modality including a path-connected cluster of the total pre-order. Then, exactly like in our case (cf. proof of Theorem 1), taking pointed-generated subframes guarantees completeness.

3.2 Defining 'best'

Our logic is quite expressive. The very first semantics for dyadic deontic logic [27] interpreted formulae $\mathbf{O}(\varphi \mid \psi)$ as "all the best ψ-states are φ" (Formula 1). Within our logic, a maximality operator can be defined as follows:

$$[\mathsf{Best}(\psi)]\ \varphi := [\forall]\ (\psi \to \langle\preceq\rangle(\psi \wedge [\preceq](\psi \to \varphi)) \tag{3}$$

That is, the best ψ-states are φ if and only if, for all states, either they are not φ or there is a better φ-state such that all states above it are either not ψ or φ. This definition was first proposed, together with some variants, by the aforementioned [12, 13]. Here, we restated it by making use of the universal modality.[18]

A few comments are in order. For Definition 3 to capture the intended meaning of the maximality operator converse well-foundedness should be assumed in Definition 4, to avoid empty sets of maximals for some formulae. If there are no maximals of a formula ψ, then the statement "all the best ψ-states are $\bot$" would hold, which is not the case under the definition in Formula 3. In fact, in the absence of converse well-foundedness, as observed in [13], Formula 3 expresses something more than maximality, namely that each ψ state can always access a better ψ state, above which it is always the case that ψ implies φ.

Example 3 (Gentle murder (continued)). Consider the P-sequence for valuation $\mathcal{I}$ of the Gentle murder introduced in Example 1, and let $\preceq_{\mathcal{B}}^{IM}$ be the total preorder generated by that sequence. We have that, for any state s in the model $\mathcal{M}_{\mathcal{B}}^{IM} = (S, \preceq_{\mathcal{B}}^{IM}, \mathcal{I})$:

$$\mathcal{M}_{\mathcal{B}}^{IM}, s \models [\mathsf{Best}(\top)]\ \mathsf{I}_1$$
$$\mathcal{M}_{\mathcal{B}}^{IM}, s \models [\mathsf{Best}(\mathsf{V}_1)]\ \mathsf{I}_2$$
$$\mathcal{M}_{\mathcal{B}}^{IM}, s \models [\mathsf{Best}(\mathsf{V}_2)]\ \mathsf{V}_2.$$

It is easy to see that the maximality statements above correspond with the analysis via restricted P-sequences provided in Example 2. In a way, Examples 2 and 3 depict two complementary views of obligation. The next section investigates this bond in details.

4 Deontics as founded on classification and betterness

We briefly sum up the technical results presented thus far about our logical framework:

- CTD structures can be represented syntactically as P-sequences;
- P-sequences determine total-preorders with a conversely well-founded strict part, and we can reason about such structures within a suitablemodal logic.

Given a P-sequence $\mathcal{B}$, these two insights suggest two possible—but, we will see, equivalent—ways of defining dyadic obligation operators: 1) what ought to be the case is what the best non-empty property of the P-sequence logically implies; 2) what ought to be the case is what holds in the best states. In formulae:

$$\mathbf{O}(\varphi \mid \psi) := [\forall]\ (Max^{+}(\mathcal{B}^{\psi}) \rightarrow \varphi) \tag{4}$$
$$\mathbf{O}(\varphi \mid \psi) := [\mathsf{Best}(\psi)]\ \varphi \tag{5}$$

[18] Other variants of Formula 3 are possible. For instance, [21, Ch. 3] proposes $[\mathsf{Best}(\psi)]\ \varphi := [\forall]\ ((\psi \wedge \neg\langle\prec\rangle\psi) \rightarrow \varphi)$ for the case that models are finite.

where, recall that $Max^{+}(\mathcal{B}^{\psi})$ denotes the best non-empty element of $\mathcal{B}$ in the restriction of $\mathcal{B}$ to ψ (see Definition 3). The second one is the very first semantics of dyadic deontic logic to which we already referred in Section 3. The first one is somehow more original, although reminiscent of the Andersonian-Kangerian reduction of deontic logic.[19]

Formulae 4 and 5 are an interesting 'dichotomy' in the definition of "ought". While Formula 4 resorts to a "classification" stating that φ is what necessarily follows from a given (syntactic) standard of behavior—the P-sequence—under circumstances φ, Formula 5 simply resorts to a given (semantic) betterness relation on states and the notion of maximality. The present section shows how they coincide, thus illustrating the two faces of the notion of obligation. To say it with St. Thomas Aquinas:

> "Voluntas [...] bonorum consonat legi, a qua malorum voluntas discordat." [47, ia.2ae 96,5]

That is to say, the preferences of the good men are in accordance with the law, and those of the bad men in disaccordance.

4.1 *Voluntas bonorum consonat legi*

Given a P-sequence, there is full accordance between defining what is obligatory as what is true in the best states of the order yielded by the P-sequence, or as what logically follows from the highest ranked property in the P-sequence. In other words, there is full correspondence between the 'letter' of the law (the P-sequence), and its content (the betterness ordering).

Theorem 2 ("Obligatory equals the consequences of best"). *Let $\varphi \in \mathcal{L}(\forall, \preceq)$ and $\mathcal{B} = \langle B, \prec \rangle$ be a P-sequence. For any model $\mathcal{M}_{\mathcal{B}}^{IM}$ derived from $\mathcal{B}$ as in Definition 2 and state s it holds that:*

$$\mathcal{M}_{\mathcal{B}}^{IM}, s \models [\mathsf{Best}(\psi)]\ \varphi \Longleftrightarrow \mathcal{M}_{\mathcal{B}}^{IM}, s \models [\forall]\ (Max^{+}(\mathcal{B}^{\psi}) \rightarrow \varphi).$$

where $\mathcal{B}^{\psi}$ is the restriction of $\mathcal{B}$ to ψ (Definition 3).

Proof. A proof can be obtained by the subsequent application, in this order, of Definition 5, Definition 2, Definition 3:

$$\begin{aligned}
\mathcal{M}_{\mathcal{B}}^{IM}, s \models [\mathsf{Best}(\psi)]\ \varphi &\Longleftrightarrow \forall s' \in Max(\langle \|\psi\|, \preceq_{\mathcal{B}}^{IM} \rangle) : \mathcal{M}_{\mathcal{B}}^{IM}, s' \models \varphi \\
&\Longleftrightarrow \forall s' \in \|\psi\| \text{ s.t. } [\forall s" \in \|\psi\| : s" \preceq_{\mathcal{B}}^{IM} s'] : \mathcal{M}_{\mathcal{B}}^{IM}, s' \models \varphi \\
&\Longleftrightarrow \forall s' \in \|\psi\| \text{ s.t. } [\forall s" \in \|\psi\|, \forall \mathsf{I}_i \in B : s'' \in \mathcal{I}(\mathsf{I}_i) \Rightarrow \\
&\qquad s' \in \mathcal{I}(\mathsf{I}_i)] : \mathcal{M}_{\mathcal{B}}^{IM}, s' \models \varphi \\
&\Longleftrightarrow \forall s' \in \|Max^{+}(\mathcal{B}^{\psi})\| : \mathcal{M}_{\mathcal{B}}^{IM}, s' \models \varphi \\
&\Longleftrightarrow \forall s' \text{ s.t. } \mathcal{M}_{\mathcal{B}}^{IM} \models Max^{+}(\mathcal{B}^{\psi}) : \mathcal{M}_{\mathcal{B}}^{IM}, s' \models \varphi \\
&\Longleftrightarrow \forall s' : \mathcal{M}_{\mathcal{B}}^{IM}, s' \models Max^{+}(\mathcal{B}^{\psi}) \rightarrow \varphi \\
&\Longleftrightarrow \mathcal{M}_{\mathcal{B}}^{IM}, s \models [\forall]\ (Max^{+}(\mathcal{B}^{\psi}) \rightarrow \varphi)
\end{aligned}$$

□

[19] We will come back to this aspect in Section 4.2.

In words, φ is the best that can hold given ψ if and only if φ is what is required by ideality under the circumstances ψ. Or, to put it yet otherwise, φ is what the law says is a primary obligation under circumstances ψ. We might suggestively say that Theorem 2 captures the ethical truism that 'what ought to be the case is what is best under the given circumstances'.

4.2 Anderson's reduction & P-sequences

Anderson's [3] and Kanger's [32] reduction of deontic logic consists in a definition of **O**-formulae to alethic modal logic $\Box$-formulae containing a designated violation or ideality constant:

$$\mathbf{O}\varphi := \Box(\neg\varphi \to \mathsf{V}) \tag{6}$$
$$\mathbf{O}\varphi := \Box(\mathsf{I} \to \varphi) \tag{7}$$

which are obviously equivalent under the assumption that $\mathsf{V} \leftrightarrow \neg\mathsf{I}$ is a valid principle. It is well-known that this reductionist view of deontic logic inherits a number of the weaknesses of deontic logic—among which the impossibility of representing CTDs satisfactorily. In this section, we briefly show how P-sequences offer a natural extension to Anderson's and Kanger's reduction.

We call a Kangerian-Andersonian P-sequence any P-sequence consisting of ideality atoms (or an inverse sequence of violation atoms). We have the following result. The following corollary further illustrates Theorem 2 in the light of Kangerian-Andersonian P-sequences.

Corollary 1 (Obligations from better to worse). *Let $\varphi \in \mathcal{L}(\forall, \preceq)$ and $\mathcal{B} = \langle B, \prec\rangle = (\mathsf{I}_1, \ldots, \mathsf{I}_n)$ be a Kangerian-Andersonian P-sequence for $\mathcal{I}$ of n non-empty elements. For any model $\mathcal{M}^{IM}_{\mathcal{B}}$ and state s it holds that:*

$$\mathcal{M}^{IM}_{\mathcal{B}}, s \models \mathbf{O}(\varphi \mid \top) \Longleftrightarrow \mathcal{M}^{IM}_{\mathcal{B}}, s \models [\forall]\ (\mathsf{I}_1 \to \varphi) \tag{8}$$
$$\mathcal{M}^{IM}_{\mathcal{B}}, s \models \mathbf{O}(\varphi \mid \mathsf{I}_1) \Longleftrightarrow \mathcal{M}^{IM}_{\mathcal{B}}, s \models [\forall]\ (\mathsf{I}_1 \to \varphi) \tag{9}$$
$$\mathcal{M}^{IM}_{\mathcal{B}}, s \models \mathbf{O}(\varphi \mid \mathsf{V}_i) \Longleftrightarrow \mathcal{M}^{IM}_{\mathcal{B}}, s \models [\forall]\ (\mathsf{I}_{i+1} \to \varphi) \quad \textit{for } 1 \leq i < n \tag{10}$$
$$\mathcal{M}^{IM}_{\mathcal{B}}, s \models \mathbf{O}(\varphi \mid \mathsf{V}_n) \Longleftrightarrow \mathcal{M}^{IM}_{\mathcal{B}}, s \models [\forall]\ (\mathsf{V}_n \to \varphi) \tag{11}$$

where $\mathbf{O}(\psi \mid \varphi)$ is defined by Formula 4 or 5.

Proof. Formulae 9-10 are instances of Theorem 2.

Formula 8 establishes that an unconditional obligation $\mathbf{O}(\varphi \mid \top)$ corresponds to what the most ideal states dictate. The corollary also shows how the content of obligations changes as we move from most ideal to least ideal circumstances. If we are in most ideal states, where I_1 holds, than what ought to be the case is in fact what already is the case (Formula 9). Formula 10 states that, if it is the case that the i^{th} ideality has been violated, where the i^{th} is not the last one in the sequence, then what ought to be is what follows from the $(i+1)^{th}$ ideality. In other words, if a norm is violated we look at the one telling us what to do if

that is the case, that is, we move downwards in the P-sequence. Finally, if we are in least ideal states, where I_n is false, then what ought to be the case is again what is already the case (Formula 11).

All in all, the corollary offers a reinterpretation and generalization of the Andersonian-Kangerian reduction, where statements of the type $\Box(\mathsf{I}_i \to \varphi)$ are interpreted as assertions concerning the relative inclusions of state labels (e.g., "all I_i-states are φ-states"), or of the ideality and sub-ideality relations proposed in [31], where such relations are taken to be the total-preorders generated by a P-sequence.

4.3 One more example: Chisholm's paradox [16]

We follow the presentation of the paradox given in [5].

1. It ought to be that Smith refrains from robbing Jones.
2. Smith robs Jones.
3. If Smith robs Jones, he ought to be punished for robbery.
4. It ought to be that if Smith refrains from robbing Jones he is not punished for robbery.

Once "Smith robbing Jones" is represented by r and "Smith refraining from robbing Jones" by $\neg r$ and, similarly, "Smith being punished" by p while "Smith not being punished" by $\neg p$, this set of ordinary language sentences—also called the Chisholm's set—can receive the following formalizations within SDL, which differ in the way they symbolize the conditional statements at points 3 and 4.

i)	*ii)*	*iii)*	*iv)*
$\mathbf{O}\neg r$	$\mathbf{O}\neg r$	$\mathbf{O}\neg r$	$\mathbf{O}\neg r$
r	r	r	r
$r \to \mathbf{O}p$	$\mathbf{O}(r \to p)$	$r \to \mathbf{O}p$	$\mathbf{O}(r \to p)$
$\neg r \to \mathbf{O}\neg p$	$\mathbf{O}(\neg r \to \neg p)$	$\mathbf{O}(\neg r \to \neg p)$	$\neg r \to \mathbf{O}\neg p$

It becomes then evident that SDL falls short in properly representing the Chisholm's set since: in i) the 4th statement $\neg r \to \mathbf{O}\neg p$ is a logical consequence of the 2nd r, which is not the case in the ordinary language formulation; in ii) the 3rd statement $\mathbf{O}(r \to p)$ is a logical consequence of the 1st $\mathbf{O}\neg r$, which is also not the case in the ordinary language formulation; finally, in iii) and iv) both $\mathbf{O}p$ and $\mathbf{O}\neg p$ are logical consequences of the set and hence $\mathbf{O}\bot$, which is clearly counter-intuitive.

The Chisholm's set displays the same structure of the gentle murder scenario. So let us see how it can be dealt with in our framework. The first essential step, is to have a model-theoretic look at the "paradox" like we did in Example 1.

Example 4 (Chisholm's models). Let $\mathbf{P} := \{p, r, \mathsf{I}_1, \mathsf{I}_2, \mathsf{V}_1, \mathsf{V}_2\}$ and assume $\mathsf{V}_1 := \neg\mathsf{I}_1$ and $\mathsf{V}_2 := \neg\mathsf{I}_2$. Consider then the following P-sequence for valuation $\mathcal{I}$:

$$\mathcal{B} = \langle\{\mathsf{I}_1, \mathsf{I}_2\}, \{(\mathsf{I}_2, \mathsf{I}_1), (\mathsf{I}_1, \mathsf{I}_1), (\mathsf{I}_2, \mathsf{I}_2)\}$$

where type I_1 is strictly preferred to type I_2 and the resulting model $\mathcal{M}_{\mathcal{B}}^{IM} = \langle W, \preceq_{\mathcal{B}}^{*}, \mathcal{I} \rangle$. A model $\mathcal{M}_{\mathcal{B}}^{IM} = \langle W, \preceq_{\mathcal{B}}^{IM}, \mathcal{I} \rangle$ is a Chisholm's model if the following formulae are valid:

$$\mathbf{O}(\neg r \mid \top) \tag{12}$$
$$\mathbf{O}(p \mid r) \tag{13}$$
$$\mathbf{O}(\neg p \mid \neg r) \tag{14}$$

or, equivalently by Theorem 2, if the following formulae are valid:

$$[\forall]\ (\mathsf{I}_1 \rightarrow \neg r) \tag{15}$$
$$[\forall]\ (r \rightarrow (\mathsf{I}_2 \rightarrow p)) \tag{16}$$
$$[\forall]\ (\mathsf{I}_1 \rightarrow \neg p) \tag{17}$$

Recall that the P-sequence for $\mathcal{I}$ requires $\mathcal{I}(\mathsf{V}_2) \subset \mathcal{I}(\mathsf{V}_1)$, and hence that $[\forall]\ ((\mathsf{I}_1 \rightarrow \mathsf{I}_2) \wedge \neg(\mathsf{I}_2 \rightarrow \mathsf{I}_1))$ is a validity. The Chisholm's scenario is thus naturally modeled by the r-states of a Chisholm's model, and in such states the $\preceq_{\mathcal{B}}^{IM}$-maximal states are p-states.

The above ones are all equivalent representations of the deontic statements in the Chisholm's set. In this case, no paradox arises. Instead, the formalization helps making explicit a semantically precise interpretation of the ordinary language formulation of the Chisholm's set. Formulae 12 and 15—which are equivalent by Corollary 1—all state that the most ideal states are $\neg r$-states. Formulae 13 and 16 represent the CTD obligation of the Chisholm's scenario: under the circumstance that r then it is most ideal that p. Finally, and probably most interestingly, Formulae 14 and 17 make clear that, in fact, the most ideal states are states in which no punishment occurs, since, by Corollary 1, it follows that they are all equivalent to $\mathbf{O}(\neg p \mid \top)$.

To get back to ordinary language, Example 4 shows that a natural and consistent interpretation of the Chisholm's scenario in terms of classification and preference goes as follows:

1. It is most ideal that Smith refrains to rob Jones;
2. Smith robs Jones;
3. The most ideal states under the assumption that Smith robs Jones are states in which Smith is punished;
4. It is most ideal that Smith is not punished.

This set-theoretic reading of the scenario gives rise to the sort of set-theoretic structure depicted in Figure 1. On the ground of these considerations we can also observe that the Chisholm's scenario does not state any CTD obligation in case, for instance, both r and $\neg p$ hold. On the other hand, it follows from our representation of the scenario that, in Chisholm's models, the most ideal among p-states are states in which r holds, in symbols, $\mathbf{O}(r \mid p)$.

5 Betterness Dynamics and Deontics

So far we have proposed a rich structured model for deontics under static circumstances. But deontic reasoning is crucially also about change, as events happen that change our evaluation of states-of-affairs in the world. Our concern in this section is *deontic dynamics*. In a way, we have already addressed a notion of dynamics by dealing with conditional obligations (Formula 1), that is, the simplest type of deontic dynamics where changes in the condition determine changes in the normative consequences. This type of dynamics does not modify the betterness relation and is achieved via a maximality-based definition. In the remaining of this section we will address 'genuine' changes in the betterness relation, as well as in the explicit codifications of betterness represented by the P-sequences.

5.1 Two level dynamics

In the current framework we can handle dynamical changes that are located both at the level of P-sequences and at the level of states. The operations that were considered in [35] for those two kinds of dynamics apply naturally here.

Before getting started with the formal definitions, in order to illustrate the intuitions backing this section we add a dynamic twist to the 'classic' example used in the presentation of the static framework: the gentle murder.

Example 5 (Gentle murder dynamified). Let us start with the P-sequence consisting of $\mathcal{B} = (\neg m)$. By Definition 2, this generates a dychotomous total preorder where all $\neg m$ states are above all m states: "It is obligatory under the law that Smith not murder Jones". Suppose this is the given deontic state-of-affairs. Now, how can we refine it in order to introduce the sub-ideal obligation to kill gently: "it is obligatory that, if Smith murders Jones, Smith murders Jones gently"? Or, in other words, how can we model the introduction, in the legal code, of this contrary-to-duty? Intuitively, this can happen in two ways:

1. We refine the given betterness ordering 'on the go' by requesting a further bipartition of the violation states, putting the $m \wedge g$-states above the $m \wedge \neg g$-states. This can be seen as the successful execution of a command of the sort "if you murder then murder gently!".
2. We introduce a new law 'from scratch', where $m \rightarrow g$ is explicitly stated as a class of possibly sub-ideal states. This can be seen as the enactment of a new P-sequence alltogether: $(\neg m, m \rightarrow g)$, which is the P-sequence we have already encountered in Example 1.[20]

The example illustrates how a CTD sequence can be dynamically created either by the utterance of a sequence of commands each stating what ought to be the case in a sub-ideal situation, or by the direct enactment of a whole P-sequence. These two points of view on the creation of obligations of a CTD type are the dynamic counterpart of the two-level deontic logic studied in the preceding sections. Their formal definitions follow.

[20] Notice that $m \rightarrow g$ is equivalent to $\neg m \vee (m \wedge g)$.

Definition 6 (Betterness upgrade). *Let $\mathcal{L}(\mathbf{P})$ be a propositional language, let $TP(S)$ denote the set of all total pre-orders over S, and let $\preceq \in TP(S)$. The upgrade operation $\Uparrow \varphi$ for $\preceq$ is defined as follows:*

$$\Uparrow \varphi(\preceq) \;:=\; (?\varphi; \preceq; ?\varphi) \cup (?\neg\varphi; \preceq; ?\neg\varphi) \cup (?\neg\varphi; \top; ?\varphi). \tag{18}$$

where $?$ and $;$ are the standard relational operations of test and sequencing, and $\top$ denotes the universal relation. We will also refer to the betterness upgrade $\Uparrow \varphi(\mathcal{M})$ of a model $\mathcal{M} = \langle S, \preceq, \mathcal{I}\rangle$ as the model $\mathcal{M}' = \langle S, \Uparrow \varphi(\preceq), \mathcal{I}\rangle$.

Intuitively, the (betterness) upgrade operation gives instructions on how to modify a given total pre-order. After the new proposition φ has been incorporated, the upgrade places all φ-worlds on top of all $\neg\varphi$-worlds, keeping all other comparisons the same. Here, besides cutting links between the φ-worlds and $\neg\varphi$-worlds, new betterness links may be added by the disjunct $(?\neg\varphi; \top; ?\varphi)$. The operation $\Uparrow\varphi$ preserves the structural properties of our models (reflexivity, transitivity and totality, as well as converse well-foundedness as the strict order clearly remains bounded).

Example 6 (Betterness upgrade in the dynamified gentle murder). Following up on Example 5, consider the P-sequence $\mathcal{B} = (\neg m)$ and fix a valuation $\mathcal{I}$ on a set of states S. The total pre-order $\preceq^{IM}_{\mathcal{B}}$ on S yielded by $\mathcal{B}$ and $\mathcal{I}$ is a dichotomous relation clustering all $\neg m$-states above all m-states. Now let us proceed with the betterness upgrade of $\preceq^{IM}_{\mathcal{B}}$ by formula $m \to g$. By instantiating Definition 6 we obtain the following relation:

$$\begin{aligned}\Uparrow m \to g(\preceq^{IM}_{\mathcal{B}}) = (?m \to g; \preceq^{IM}_{\mathcal{B}}; ?m \to g) \cup (?m \wedge \neg g; \preceq^{IM}_{\mathcal{B}}; ?m \wedge \neg g)\\ \cup (?m \wedge \neg g; \top; ?m \to g).\end{aligned}$$

The last disjunct imposes that the set of states satisfying $\neg m \vee g$ are strictly better than the states satisfying $m \wedge \neg g$. At the same time, the first and second disjuncts guarantee that the $\neg m$-states are above the $m \wedge g$-ones (and hence are the best states) and, respectively, that the $m \wedge \neg g$-states all belong to a same indifference class. As a result, the dichotomous relation $\preceq^{IM}_{\mathcal{B}}$ is turned into a trichotomous one. This update is depicted in Figure 2.

However, in what follows our focus is not on changes at the level of the betterness order alone, but its close ties to changes at the level of P-sequences. To formalize such connections, we need a new technical notion.

Definition 7 (Priority changes inducing betterness changes). *Let $\mathcal{L}(\mathbf{P})$ be a propositional language built on $\mathbf{P}$, and $\mathcal{I}$ a valuation function on the set of states S. Let then $F : \mathbb{B}^{\mathcal{I}} \times \mathcal{L}(\mathbf{P}) \longrightarrow \mathbb{B}^{\mathcal{I}}$ be a function assigning a P-sequence to a P-sequence and a formula φ (recall that $\mathbb{B}^{\mathcal{I}}$ denotes the set of all P-sequences for $\mathcal{I}$). Finally, let $TP(S)$ denote the set of all total pre-orders over S and let $\sigma : TP(S) \times \mathcal{L}(\mathbf{P}) \longrightarrow TP(S)$ be a function assigning a total pre-order over the set of states S, given a total pre-order and a formula. We say that F induces the map σ, given a definition for deriving betterness relations from P-sequence (e.g., Definition 2), if and only if, for any P-sequence $\mathcal{B}$ and new formula φ, $\sigma(\preceq_{\mathcal{B}}, \varphi) = \preceq_{F(\mathcal{B},\varphi)}$.*

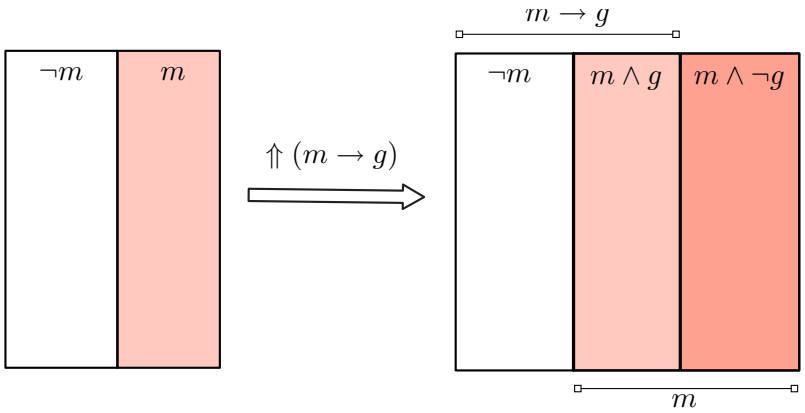

Fig. 2. Betterness upgrade in the gentle murder. The two rectangles are partitioned in clusters of equally preferred states. Clusters are ordered from more to less preferred from left to right.

In a nutshell, Definition 7 states that a priority change F induces a betterness change σ if and only if the new total pre-order obtained via σ is the same as the one derived from the P-sequence after the changes dictated by F.

Now, let σ be the betterness upgrade operation defined in Definition 6, and let F be the operation of postfixing a P-sequence $\mathcal{B}$ with a new formula φ (in symbols $\mathcal{B};\varphi$), that is, of adding a least propositional formula in the P-sequence. It turns out that the latter operation induces—in the sense of Definition 7—the former under Definition 2.

Theorem 3 (Correspondence of the two-level dynamics in CTDs). *Let $\mathcal{B} = \langle B, \prec\rangle = (\mathsf{l}_1, \ldots, \mathsf{l}_n)$ be a P-sequence, and let φ be of the form $\neg\mathsf{l}_n \to \psi$, with ψ propositional and such that $[\![\psi]\!] \supset \emptyset$. The following diagram commutes:*

$$\begin{array}{ccc} \mathcal{B} & \xrightarrow{;\varphi} & \mathcal{B};\langle\varphi\rangle \\ {\scriptstyle IM}\downarrow & & \downarrow{\scriptstyle IM} \\ \langle S, \preceq^{IM}_{\mathcal{B}}\rangle & \xrightarrow{\Uparrow\varphi} & \langle S, \Uparrow \varphi(\preceq^{IM}_{\mathcal{B}})\rangle \end{array}$$

Proof. First it should be observed that the postfixing of $\mathcal{B}$ with φ is always possible by the syntactic form of $\varphi := \neg\mathsf{l}_n \to \psi$ and the assumption that ψ is not empty. This guarantees that $[\![\mathsf{l}_n \vee \psi]\!] \supset [\![\mathsf{l}_n]\!]$ as required by Definition 1. This having been said, we need to prove the following equivalence: $s \preceq^{IM}_{\mathcal{B};\varphi} s' \iff s \Uparrow \varphi(\preceq^{IM}_{\mathcal{B}})\, s'$. [Right to left] Take $s \Uparrow \varphi(\preceq^{IM}_{\mathcal{B}})\, s'$. By Definition 6 this means that $s((?\varphi; \preceq^{IM}_{\mathcal{B}}; ?\varphi) \cup (?\neg\varphi; \preceq^{IM}_{\mathcal{B}}; ?\neg\varphi) \cup (?\neg\varphi; \top; ?\varphi))s'$. By the syntactic form of φ we obtain in turn that $\forall\psi \in \mathcal{B}$: $s \in [\![\psi]\!]$ implies $s' \in [\![\psi]\!]$ and that if $s \in [\![\varphi]\!]$ then $s' \in [\![\varphi]\!]$, from which it follows $s \preceq^{IM}_{\mathcal{B};\varphi} s'$ by Definition 1. [Left to right]

Assume that $s \preceq^{IM}_{\varphi;\mathcal{B}} s'$. By Definition 1 this means that $\forall \xi \in B \cup \{\varphi\} : s \in [\![\xi]\!]$ implies $s' \in [\![\xi]\!]$. In particular, $\forall \xi \in B : s \in [\![\xi]\!]$ implies $s' \in [\![\xi]\!]$ and if $s \in [\![\mathsf{l}_n \vee \psi]\!]$ then $s' \in [\![\mathsf{l}_n \vee \psi]\!]$. Spelling this out, by Definition 2, we have that (s, s') belongs to: either $(?\mathsf{l}_n \vee \psi; \preceq^{IM}_{\mathcal{B}}; ?\mathsf{l}_n \vee \psi)$; or $(?\neg\mathsf{l}_n \wedge \neg\psi; \preceq^{IM}_{\mathcal{B}}; ?\neg\mathsf{l}_n \wedge \neg\psi)$; or $(?\neg\mathsf{l}_n \wedge \neg\psi; \top; ?\mathsf{l}_n \vee \psi)$. It follows that $s \Uparrow \varphi(\preceq^{IM}_{\mathcal{B}})s'$ by Definition 6. □

This theorem is, in a way, a dynamic variant of the kind of static correspondence shown in Theorem 2. It makes the two faces of deontics explicit, also from a dynamic point of view.

Example 7 (Two level dunamics in the gentle murder). Let $\mathsf{l} := \neg m$. The two types of dynamics introduced in Example 5 and their correspondence can be captured by the following instantiation of Theorem 3 with $\varphi := m \to g$.

$$\begin{array}{ccc} \langle \neg m \rangle & \xrightarrow{;m\to g} & \langle \neg m \rangle; \langle m \to g \rangle \\ {\scriptstyle IM}\downarrow & & \downarrow{\scriptstyle IM} \\ \langle S, \preceq^{IM}_{\langle \neg m \rangle} \rangle & \xrightarrow{\Uparrow m\to g} & \langle S, \Uparrow m \to g(\preceq^{IM}_{\langle \neg m \rangle}) \rangle \end{array}$$

In words, the same deontic change can be obtained both by 'refining' the order dictated by the given law, via subsequent orders (of a certain syntactic form), as well as by enacting a new 'law' which correspondingly extends the given one.

5.2 *Voluntas bonorum consonat legi* ... in two level dynamics

In this section we put together Theorems 2 and 3 showing that the sort of correspondence proven by Theorem 2 is preserved by two-level betterness dynamics.

Corollary 2 (Obligation definitions under two level dyanamics). *Let $\varphi \in \mathcal{L}(\forall, \preceq)$, $\mathcal{B} = \langle B, \prec \rangle = (\mathsf{l}_1, \ldots, \mathsf{l}_n)$ be a P-sequence, and let ξ be of the form $\neg\mathsf{l}_n \to \rho$, with ρ propositional and such that $[\![\rho]\!] \supset \emptyset$. For any model $\mathcal{M}^{IM}_{\mathcal{B}}$ derived from $\mathcal{B}$ as in Definition 2 and state s, the following statements are equivalent:*

$$\Uparrow \xi(\mathcal{M}^{IM}_{\mathcal{B}}), s \models [\mathsf{Best}(\psi)]\ \varphi \tag{19}$$

$$\mathcal{M}^{IM}_{\mathcal{B};\langle\xi\rangle}, s \models [\mathsf{Best}(\psi)]\ \varphi \tag{20}$$

$$\mathcal{M}^{IM}_{\mathcal{B};\langle\xi\rangle}, s \models [\forall]\ (Max^{+}((\mathcal{B}; \langle\xi\rangle)^{\psi}) \to \varphi) \tag{21}$$

$$\Uparrow \xi(\mathcal{M}^{IM}_{\mathcal{B}}), s \models [\forall]\ (Max^{+}((\mathcal{B}; \langle\xi\rangle)^{\psi}) \to \varphi) \tag{22}$$

where $\mathcal{B}^{\psi}$ is the restriction of $\mathcal{B}$ to ψ (Definition 3).

Proof. The equivalence between Formula 19 and Formula 20 is a direct consequence of Theorem 3. The equivalence between Formula 20 and Formula 21 is a direct consequence of Theorem 2. Finally, the equivalence between Formula 21 and Formula 22 is again a direct consequence of Theorem 3. □

In short, Corollary 2 establishes that the definition of $\mathbf{O}(\varphi \mid \psi)$ as $[\mathsf{Best}(\psi)]\ \varphi$ remains equivalent to its definition as $[\forall]\ (Max^{+}(\mathcal{B}^{\psi}) \to \varphi)$ after both betterness upgrade and priority change of the P-sequence.

5.3 Betterness dynamics, CTDs and *strong permission*

On the ground of what presented in the two previous sections we put forth here an interpretation of the so-called *strong permission* in terms of CTDs. The claim we are going to defend is that strong permission can be soundly viewed as a 'speech act' or, more technically, an update, which brings about a contrary-to-duty.

In deontic logic, the distinction between weak and strong permission is an old one and dates back to [56]:

> "An act will be said to be permitted in the weak sense if it is not forbidden; and it will be said to be permitted in the strong sense if it is not forbidden but subject to norm. [...] Weak permission is not an independent norm-character. Weak permissions are not prescriptions or norms at all. Strong permission only is a norm-character." [56, p. 86]

To put it otherwise, the weak permission "it is permitted that φ" amounts simply to the mere absence of the prohibition "it is forbidden that φ", that is, it is equivalent to the dual of the (unconditional) obligation "it ought to be the case that φ", in symbols: $\neg\mathbf{O}(\neg\varphi \mid \top)$. So, weak permission is definable in terms of obligation but—and that is the quote's claim—that is not the case for strong permission. From the formal point of view this distinction poses the challenge of finding a representation of the notion of permission which does justice to this distinction.

As the literature in deontic logic witnesses, the formal analysis of the distinction between weak and strong permission has been object of quite some attention, although never occupying a predominant position on the research agenda in the field. Recently, however, the importance of this issue has been stressed again in [26] as one of ten 'philosophical problems' in deontic logic. This section intends to contribute to that literature from the point of view of betterness dynamics.

We start by viewing strong permission as the introduction of an exception to an already existing obligation. This is in line with interpretations put forth within legal theory.

> "Telling me what I am permitted to do provides no guide to conduct unless the permission is taken as an exception to a norm of obligation [...]. Norms of permission have the normative function only of indicating, within some system, what are the exceptions from the norms of the obligation of the system." [44, p. 120]

If we view strong permission as an operation of exception-introduction, it still remains to be understood what "exception" precisely means. We will look into this by means of an example.

Example 8 (Killing in self-defense). Let us start again with the primary obligation of the gentle murder scenario:

It is obligatory that Smith not kill Jones.[21]

Like in example 5, suppose we want now to introduce a change in this norm. However, suppose the change we want to introduce to be expressed by the following strong permission:

Smith is allowed to kill Jones in self-defense.

We abbreviate "killing" with k and "killing in self-defense" with d.

There seem to be only two possible ways to interpret the effects of the strong permission in the example.

1. The effect of the strong permission amounts to giving up the validity of $\mathbf{O}(\neg k \mid \top)$ for the validity of the weaker obligation $\mathbf{O}(k \rightarrow d \mid \top)$. Syntactically this would correspond to the removal of $\neg r$ from the P-sequence and its substitution with $k \rightarrow d$. In this case, a strong permission consists essentially in the expansion of the field of permissibility [28]. Such an expansion can be obtained by, in the first place, *rejecting* [2] or, more properly, *annulling* [23] an obligation and then by introducing a less restrictive one, or, equivalently, by just *derogating* part of the obligation.[22]
2. The effect of the strong permission does not modify the validity of $\mathbf{O}(\neg k \mid \top)$, instead, it introduces a CTD stating that, in case Smith kills Jones, he should better do that in self-defense: $\mathbf{O}(d \mid k)$. We will see in Example 9 how this can be rendered formally.

To put it shortly, which one of the two formulae $\mathbf{O}(k \rightarrow d \mid \top)$ and $\mathbf{O}(d \mid k)$ best captures the effects of a strong permission introducing exceptions for $\mathbf{O}(\varphi \mid \top)$?[23]

Although within different formal setups, the first option has been investigated among others in [14] and [6], where a dynamic logic is used. Here, we favor the second interpretation, as dropping $\mathbf{O}(\neg k \mid \top)$ altogether might arguably be seen too radical an effect for the introduction of a permission to kill under given circumstances. By doing this we highlight a relation between strong permission and CTDs which, to the best of our knowledge, has thus far never been explicitly stated, despite being to a sure extent adumbrated already in the input/output logic [36] literature (cf. [37] and, more recently, [49]).

[21] We substitute murder with killing as in the common usage of the term murder is usually taken to mean "unlawful killing". The example will make use of the weaker notion of "killing".

[22] The link between strong permission and derogation has often been stressed in the theory of law. We report here an excerpt quoted in [10, 11]:

> "The difference between weak and strong permission becomes clear when thinking about the function of permissive norms. [...] A permissive norm is necessary when we have to repeal a preceding imperative norm or to derogate to it. That is to abolish a part of it [...]." [8, p. 891-892]

[23] It is worth noticing that this dichotomy has a counterpart in doxastic logic where beliefs about conditional statements are distinguished from conditional beliefs.

As argued in Example 8, our claim here is that the process of incremental specification of a CTD sequence via betterness update, be it syntactic or semantic, can legitimately be viewed as the enactment of strong permissions. Such permissions are refinements of existing obligations or, to say it otherwise, exceptions to existing obligations that do not reject such obligations altogether, but rather specify conditions under which a violation of such obligations is tolerable. Notice also that, under this interpretation and against the intuitions that might be dictated by the natural language formulation of the notion, strong permission has actually more to do with **O**-statements of (although of a CTD type) rather than with $\neg\mathbf{O}\neg$-statements of (weak) permission.

Example 9 (Two-level dynamics of strong permission). Following Example 8 consider a model P-sequence $\langle\neg k\rangle$, whose derived model $\mathcal{M}$ validates $\mathbf{O}\neg k$. The enactment of a strong permission to kill if acting in self-defense can be modeled by the introduction of the CTD establishing the conditions that, when satisfied, make killing better than killing when those conditions are violated: either not killing or, if killing, may that be in self-defence: $\neg k \vee (k \wedge d)$, i.e., $k \rightarrow d$. So a strong permission can be introduced either at the semantic level by a betterness upgrade that changes the betterness ordering yielded by the existing obligations, or at the syntactic level by an operation on the P-sequence which extends it to cover the conditions under which the exception holds. What we get is again an instantiation of Theorem 3:

$$\begin{array}{ccc} \langle\neg k\rangle & \xrightarrow{;k\rightarrow d} & \langle\neg k\rangle;\langle k\rightarrow d\rangle \\ {\scriptstyle IM}\downarrow & & \downarrow{\scriptstyle IM} \\ \langle S, \preceq^{IM}_{\langle\neg k\rangle}\rangle & \xrightarrow{\Uparrow k\rightarrow d} & \langle S, \Uparrow k\rightarrow g(\preceq^{IM}_{\langle\neg k\rangle})\rangle \end{array}$$

Before moving to the next section, one more view of strong permission is worth mentioning which has been advocated by [15] and taken up by [10, 11] but also, earlier, by [37] under the name *positive dynamic permission*. Although not proposing a view in direct conflict with the one upon which we ground our analysis, [15] questions the thesis that strong permission always presupposes some pre-existing obligation. To put it in its own words, "the role played by permissive norms is not exhausted by derogation of former prohibition" [15, p. 253]. This claim is grounded on the observation that, in actual normative systems such as the law, strong permissions can occur without obligations from which they derogate. This is the case when strong permissions are enacted by some some normative authority as a means for setting boundaries to the prohibitions that subjected normative authorities could possibly enact.

5.4 Betterness dynamics and norm change: a brief discussion

The dynamic aspects of norms—the so-called *norm change* problem—have recently gained much attention from researchers in deontic logic, legal theory and

multi-agent systems. Before concluding this section on betterness dynamics we briefly want to put our work in perspective with some of the more recent contributions available in the literature on this topic, highlight similarities and future research lines.

In our view, existing approaches to norm change fall into two main groups. In syntactic approaches—inspired by legal practice—norm change is an operation performed directly on the explicit provisions in the "code" of the normative system [23, 24, 9]. In semantic approaches, norm change follows the dynamic logic update paradigm (e.g., [6]). Our betterness dynamics belongs to this latter group. Thus, it can be naturally related to the sort of context dynamics—and the related dynamics of counts-as rules—studied in [6] via the bridge offered by Corollary 1: obligations defined via ideality and maximality are special kinds of classifications of an Andersonian-Kangerian type.

Also, the dynamic logic connection enables a unified treatment of *t*wo kinds of change that mix harmoniously in deontic reasoning: 'information change' given a fixed normative order, and 'evaluation change' in such an order. Their interplay reflects the *e*ntanglement of obligation, knowledge and belief studied in [35, 41].

Finally, despite its semantic flavour, the study of dynamics we propose here bridges very well with more syntactic analyses of norm change. In fact, Theorem 3 can be viewed as establishing a precise match between changes at the level of models with changes at the level of syntax of a normative code, i.e., the P-sequences.

6 Dilemmas and norm merging

In this section we briefly point to a natural generalization of our framework and set the stage for future investigations. We are going to touch upon issue of the representation and genesis of deontic dilemmas.

6.1 Deontic dilemmas and priority graphs

The rendering of deontic scenarios, such as the gentle murder one, in terms of P-sequences does not leave room for the representation of normative conflicts or deontic dilemmas. These arise, typically, when two obligations command contradictory formulae under the same conditions, e.g. $\mathbf{O}(\varphi \mid \psi)$ and $\mathbf{O}(\neg\varphi \mid \psi)$ (cf. [30]). It is easy to verify that when the betterness relation is yielded by a P-sequence the two formulae cannot are not satisfiable together since the set of maximal ψ-consistent elements of the given P-sequence is always a singleton.

It follows that, in order to allow for the representation of deontic dilemmas, Definition 1 needs to be generalized to cover a broader class of structures beyond finite sequences.

Definition 8 (P-graphs). *Let $\mathcal{L}(\mathbf{P})$ be a propositional language built on the set of atoms $\mathbf{P}$, S a non-empty set of states and $\mathcal{I} : \mathbf{P} \longrightarrow 2^S$ a valuation function. A P-graph for $\mathcal{I}$ is a tuple $\mathcal{G}^{\mathcal{I}} = \langle G, \prec \rangle$ where:*

– $G \subset \mathcal{L}(\mathbf{P})$ *with* $|G| < \omega$*;*
– $\prec$ *is a strict pre-order on* G*, i.e., an irreflexive and transitive binary relation;*
– *for all* $\varphi, \psi \in G$*, if* $\varphi \prec \psi$ *then* $[\![\psi]\!]_{\mathcal{I}} \subset [\![\varphi]\!]_{\mathcal{I}}$

where $[\![\varphi]\!]_{\mathcal{I}}$ *denotes the truth-set of* φ *according to* $\mathcal{I}$*.* [24]

So, a P-graph is nothing but a P-sequence where the totality constraint on the priority relation is dropped. Also in this case, the third clause in Definition 8 forces the graph to correspond to the graph yielded by the strict part of set-theoretic inclusion on the elements of G.

Given a P-graph, we can then apply Definition 2 to derive the corresponding betterness relation among the states of the model. It is easy to see that such relation will not be a total pre-order any more, but simply a pre-order, i.e., a reflexive and transitive binary relation. It becomes then easy to see that the results presented in the paper can obtain a rather straightforward generalization to pre-orders and their well-known modal logic, i.e. **S4**.

Priority graphs can then be composed to yield new graphs. Natural operations on priority graphs are, for instance [4]:

– sequential composition (in symbols, $\mathcal{G}_1; \mathcal{G}_2$), which puts $\mathcal{G}_1$ above $\mathcal{G}_2$ in such a way that all its nodes are above the nodes of $\mathcal{G}_2$;
– parallel composition (in symbols, $\mathcal{G}_1 || \mathcal{G}_2$), which simply takes the disjoint union of the two graphs.

Further operations are studied, for instance, in [17]. In general, what we want to stress here is that the possibility of combining different priority structures suggests a generalization of the sort of dynamics in focus in Section 5. And from the point of view of deontics it suggests a 'constructive' look at systems of norms as the result of operations on the priority structures that represent them. This insight takes us to the next section.

6.2 Priority graphs as the result of *norm merging*

Deontic dilemmas have given rise to abundant literature in deontic logic (e.g. [30]), in the theory of law (e.g., [1]), and in the field of artificial intelligence and law (e.g. [42]). In this section we want to sketch a line of research which looks at such problems from the point of view of priority graphs. It is not our aim to develop a fully-fledged theory here, but rather to point to an interesting further development of the priority/betterness view of deontics we have been advocating.

Our main claim is that normative systems, or codes, arise from the combination of different P-graphs and, in the first instance, of different P-sequences. In deontic logic, the issue of the composition of different norms has recently been called *norm merging*, and has been recognized as one of the 'fundamental' open problems in the field [26]. The problem is best illustrated by means of an example.

[24] Like Definition 1, Definition 8 could be restated by stipulating that $\prec \subseteq \supset|_G$ (cf. Footnote 10).

Example 10 (Quick murder). Let us assume now there are two normative sources in the 'Murder code'. According to the first one:

1. It is obligatory under the law that Smith not murder Jones.
2. It is obligatory that, if Smith murders Jones, Smith murders Jones gently."

According to the second one:

1. It is obligatory under the law that Smith not murder Jones.
2. It is obligatory that, if Smith murders Jones, Smith murders Jones *quickly*."

Assuming the alphabet $\{m, g, q\}$ with the obvious intuitive interpretation, and a valuation $\mathcal{I}$ we can model this scenario by means of two P-sequences $\mathcal{B}_g$ with $\neg m \succ \neg m \vee (m \wedge g)(= m \rightarrow g)$, and $\mathcal{B}_q$ with $\neg m \succ \neg m \vee (m \wedge q)(= m \rightarrow q)$. In models where $\mathcal{I}(g) \cap \mathcal{I}(q) = \emptyset$, that is, where killing gently is incompatible with killing quickly, we seem to obtain an instance of a deontic conflict, although at a sub-ideal level: $\mathbf{O}(g \mid m)$ and $\mathbf{O}(\neg g \mid m)$ or, equivalently, $\mathbf{O}(g \mid m)$ and $\mathbf{O}(\neg g \mid m)$. What kind of P-graph would we obtain by merging the two P-sequences? And what kind of betterness pre-order would arise?

The problem illustrated by the example can be formulated in its generality as follows. Take two P-graphs $\mathcal{G}$ and $\mathcal{G}'$ for two valuations $\mathcal{I}$ and $\mathcal{I}'$ on domain S, and let $\odot$ denote the to-be-defined merging operation. Two questions present themselves, from a semantic and syntactic point of view:

1. How should the betterness pre-order on states $\preceq^{IM}_{\mathcal{G} \odot \mathcal{G}'}$ look like, which is yielded by $\mathcal{G} \odot \mathcal{G}'$?
2. How should $\odot$ be defined—as an operation on $\mathcal{G}$ and $\mathcal{G}'$—so that the application of Definition 2 to $\mathcal{G} \odot \mathcal{G}'$ delivers the desired pre-order?

The first question concerns the 'meaning' of the new P-graph resulting from the merging, that is, the betterness pre-order that would be induced by Definition 2. The second one concerns the method by means of which the P-graph can be built from the underlying sequences. Such method should be sound with respect to the desired 'meaning' of the graph.

6.3 Norm merging as parallel composition of P-graphs

Let us try to answer the two above questions with respect to Example 10. The first one does not seem difficult to answer. The resulting pre-order should be such that the best states are $\neg m$-states and the sub-ideal states are split into two incomparable classes, the class of $m \wedge g$-states and the class of $m \wedge q$-states (see Figure 3). This is nothing but the intersection of the betterness (total) pre-orders of the two P-sequences. As to the second question, the abovementioned operation of parallel composition yields the desired result. In fact, the application of Definition 2 to the disjoint union of the two P-sequences in Example 10 yields precisely the pre-order we just described.

More generally, it is the following fact [4] relating parallel composition of P-graphs and intersection of their derived pre-orders that serves us the answer to the two above questions.

Fact 2 (Semantics of parallel composition of P-graphs) *Let $\mathcal{G} = \langle G, \prec \rangle$ and $\mathcal{G}' = \langle G', \prec' \rangle$ be two P-graphs, and $\mathcal{I}$, $\mathcal{I}'$ their valuations. It holds that:*

$$\preceq^{IM}_{\mathcal{G}||\mathcal{G}'} = \preceq^{IM}_{\mathcal{G}} \cap \preceq^{IM}_{\mathcal{G}'} .$$

Proof (Sketch). The proof is given by the following equivalences obtained by iterated application of Definition 2.

$$\begin{aligned} s \preceq^{IM}_{\mathcal{G}||\mathcal{G}'} s' &\iff \forall \varphi \in G \cup G' : s \in [\![\varphi]\!] \Rightarrow s' \in [\![\varphi]\!] \\ &\iff \forall \varphi \in G : s \in [\![\varphi]\!] \Rightarrow s' \in [\![\varphi]\!] \text{ and } \forall \varphi \in G' : s \in [\![\varphi]\!] \Rightarrow s' \in [\![\varphi]\!] \\ &\iff s \preceq^{IM}_{\mathcal{G}} s' \text{ and } s \preceq^{IM}_{\mathcal{G}'} s'. \end{aligned}$$

This completes the proof. □

In the light of this result, we can recapitulate our view of norm merging as follows. Sets of norms, allowing for conflicts, can be represented as P-graphs. The process of norm merging can then be viewed as a generalization of the betterness dynamics process we have studied in Section 5 where, instead of updating a P-sequence or betterness total pre-order via formulae, we update a P-graph by means of another P-graph and, correspondingly, a betterness pre-order by means of another betterness pre-order. Once again, two faces of a same coin appear.

Corollary 3 (The two faces of merging). *Let $\mathcal{G}$ and $\mathcal{G}'$ be two P-graphs and S a non-empty set of states. The following diagram commutes.*

$$\begin{array}{ccc} \mathcal{G} & \xrightarrow{\;||\mathcal{G}'\;} & \mathcal{G}||\mathcal{G}' \\ {\scriptstyle IM}\downarrow & & \downarrow{\scriptstyle IM} \\ \langle S, \preceq^{IM}_{\mathcal{G}} \rangle & \xrightarrow{\;\cap \preceq^{IM}_{\mathcal{G}'}\;} & \langle S, \preceq^{IM}_{\mathcal{G}||\mathcal{G}'} \rangle \end{array}$$

Proof. Follows directly from Fact 2.

In other words, the 'merging' of $\mathcal{G}$ with $\mathcal{G}'$ yields the same pre-order that we would obtain by intersecting the pre-orders derived from $\mathcal{G}$ and, respectively, $\mathcal{G}'$.

Although parallel composition seems to provide, on the ground of the above results, a readily available and neat characterization of norm merging, there is one more technical subtlety in the merging problem which deserves a few words. Going back to Example 10 we can notice that the two P-sequences share part of their logical alphabet (atom m) and differ in the rest (atoms g and q). Intuitively, this suggests that node $\neg m$, which is present in both P-sequences, should be represented as only one node in the resulting graph, and not as two nodes like in what would result from parallel composition. This issue could be tackled by resorting to a transformation of the graph to the kind of normal form for P-graphs studied in [4], where the normal form graph exhibits minimal complexity in the sense of having a minimal number of nodes.

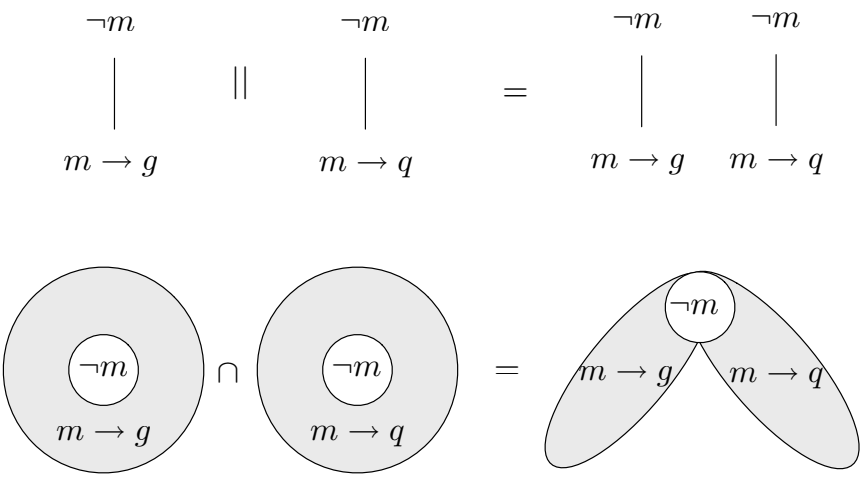

Fig. 3. Merging of P-sequences in Example 10. The merging is depicted both in the form of parallel composition of P-sequences (top) and as intersection of total pre-orders in the form of spheres (bottom).

To conclude, P-graphs offer the sort of generalization of P-sequences that allow us to enrich CTD structures with the representation of dilemmas. P-graphs themselves, and therefore deontic dilemmas, can then be viewed as the result of a process of norm merging also exhibiting a syntactic (parallel composition) and a semantic (conjunction of pre-orders) face. The natural question then arises of how to solve those dilemmas or, in other words, of how to *linearize* P-graphs and pre-orders back to P-sequences and total pre-orders. This research question, which we leave for future investigations, seems to naturally bridge with the aggregation problem as studied, for instance, in social choice theory [20].

7 Conclusions

This paper has revisited the preference-based semantics of deontic logic first presented in [27], and then developed it further into a richer medium for analyzing deontic reasoning. This was done in two related ways: by introducing a two-level perspective on deontic ideality which enables a two-faceted modal analysis of deontic concepts (Theorem 2 and Corollary 1), and also a rich view of deontic dynamics as betterness change (Theorem 3 and Corollary 2) as well as of norm merging as composition of priority graphs (Corollary 3).

Along with exposing these results, the paper has also punctually put the work in perspective with relevant literature. This has been done in several places: in Section 2 we have related the application of P-sequences to CTDs to existing work in deontic logic stressing the sub-ideality component of CTDs; in Section 3 we have taken care of relating our framework to the tradition of preference

logic based on standard modal logic; in Section 5.4 we have discussed the place of betterness dynamics within deontic logic literature on norm change and, from that point of view, we have also provided an overview of related work on strong permission (Section 5.3).

Acknowledgments The authors would like to thank the anonymous reviewers of DEON'10, where a previous version of the paper has been published, as well as Jörg Hansen and Leon van der Torre for useful pointers to the literature. Davide Grossi is supported by the *Nederlandse Organisatie voor Wetenschappelijk Onderzoek* (VENI grant 639.021.816).

References

1. C. E. Alchourrón and E. Bulygin. *Normative Systems*. Springer Verlag, 1971.
2. C. E. Alchourrón and E. Bulygin. The expressive conception of norms. In R. Hilpinen, editor, *New Studies in Deontic Logic*, pages 95–124. Reidel Publishing Company, 1981.
3. A.R. Anderson. The formal analysis of normative concepts. *American Sociological Review*, 22:9–17, 1957.
4. H. Andreka, M. Ryan, and P. Schobbens. Operators and laws for combining preference relations. *Journal of Logic and Computation*, 12:13–53, 2002.
5. L. Åqvist. Good samaritans, contrary-to-duty imperatives and epistemic obligations. *Nous*, 1:361–379, 1967.
6. G. Aucher, D. Grossi, A. Herzig, and E. Lorini. Dynamic context logic. In X. He, J. Horty, and E. Pacuit, editors, *Proceedings of LORI 2009*, volume 5834 of *LNAI*. Springer, 2009.
7. P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, Cambridge, 2001.
8. N. Bobbio. Norma. In *Enciclopedia Einaudi*, volume 9, pages 876–907. Einaudi, Torino, Italy, 1980.
9. G. Boella, G. Pigozzi, and L. van der Torre. Normative framework for normative system change. In Decker P., Sichman J., Sierra C., and Castelfranchi C., editors, *Proceedings of the Eighth International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, 2009.
10. G. Boella and L. van der Torre. Permissions and obligations in hierarchical normative systems. In *Proceedings of ICAIL'03*, pages 109–118, New York, 2003. ACM Press.
11. G. Boella and L. van der Torre. Institutions with a hierarchy of authorities in distributed dynamic environments. *Journal of Artificial Intelligence and Law*, 16:53–71, 2008.
12. C. Boutilier. Conditional logics of normality as modal systems. In *Proceedings of AAAI'90*, pages 594–599, 1990.
13. C. Boutilier. Conditional logics of normality: A modal approach. *Artificial Intelligence*, 68:87–154, 1994.
14. M. Brown. Conditional obligation and positive permission for agents in time. *Nordic Journal of Philosophical Logic*, 5(2):83–112, 2000.
15. E. Bulygin. Permissive norms and normative systems. In A. Martino and F. Socci Natali, editors, *Automated analysis of legal texts*, pages 211–218. Publishing company, Amsterdam, 1986.

16. R. M. Chisholm. Contrary-to-duty imperatives and deontic logic. *Analysis*, 23:33–36, 1963.
17. D. de Jongh and F. Liu. Preference, priorities and belief. In T. Grüne-Yanoff and S. O. Hansson, editors, *Preference Change: Approaches from Philosophy, Economics and Psychology*, Theory and Decision Library. Springer, 2009.
18. D. Etherington and R. Reiter. On inheritance hierarchies with exceptions. In *Proceedings of AAAI'83*, pages 104–108, 1983.
19. J. Forrester. Gentle murder, or the adverbial samaritan. *Journal of Philosophy*, 81:193–197, 1984.
20. W. Gaertner. *A Primer in Social Choice Theory.* Oxford University Press, 2006.
21. P. Girard. *Modal Logic for Belief and Preference Change.* PhD thesis, University of Amsterdam, 2008.
22. G. Governatori and A. Rotolo. Logic of violations: A gentzen system for reasoning with contrary-to-duty obligations. *Australasian Journal of Logic*, 3, 2005.
23. G. Governatori and A. Rotolo. Changing legal systems: Abrogation and annulment. part 1: Revision and defeasible theories. In Ron van der Meyden and Leendert van der Torre, editors, *Proceedings of the 9th International Conference on Deontic Logic in Computer Science (DEON 2008), Luxembourg, Luxembourg, July 15-18, 2008*, volume 5076 of *LNAI*, pages 3–18. Springer, 2008.
24. G. Governatori and A. Rotolo. Changing legal systems: Abrogation and annulment. part 2: Temporalised defeasible logic. In G. Boella, G. Pigozzi, M. P. Singh, and H. Verhagen, editors, *Proceedings of the 3rd International Workshop on Normative Multiagent System (NorMAS 2008), Luxembourg, Luxembourg, July 14-15, 2008*, pages 112–127, 2008.
25. A. Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17:157–170, 1988.
26. J. Hansen, G. Pigozzi, and L. van der Torre. Ten philosophical problems in deontic logic. In Guido Boella, Leon van der Torre, and Harko Verhagen, editors, *Normative Multi-agent Systems*, number 07122 in DROPS Proceedings. Internationales Begegnungs- und Forschungszentrum fuer Informatik (IBFI), Schloss Dagstuhl, Germany, 2007.
27. B. Hansson. An analysis of some deontic logics. *Nous*, 3:373–398, 1969.
28. R. Hilpinen. On normative change. In E. Morscher and R. Stranzinger, editors, *Ethics. Foundations, Problems, and Applications. Proceedings of The 5th International Wittgenstein Symposium*, pages 155–164. Hölder-Pichler-Tempsky, 1980.
29. J. Horty. Deontic logic as founded on nonmonotonic logic. *Annals of Mathematics and Artificial Intelligence*, 9, 1993.
30. J. Horty. Moral dilemmas and nonmonotonic logic. *Journal of Philosophical Logic*, 23:35–65, 1994.
31. A. J. I. Jones and I. Pörn. Ideality, sub-ideality and deontic logic. *Synthese*, pages 275–290, 1985.
32. S. Kanger. New fondations for ethical theory. In R. Hilpinen, editor, *Deontic Logic: Introductory and Systematic Readings*, pages 36–58. Reidel Publishing Company, 1971.
33. D. Lewis. *Counterfactuals.* Blackwell, Oxford, 1973.
34. D. Lewis. Semantic analysis for dyadic deontic logic. In S. Stenlund, editor, *Logical Theory and Semantical Analysis*, pages 1–14. Reidel Publishing Company, 1974.
35. F. Liu. *Changing for the Better. Preferemce dynamics and Agent Diversity.* PhD thesis, University of Amsterdam, 2008.
36. D. Makinson and L. van der Torre. Input-output logics. *Journal of Philosophical Logic*, 29(4):383–408, 2000.

37. D. Makinson and L. van der Torre. Permission from an input/output perspective. *Journal of Philosophical Logic*, 32:391–416, 2003.
38. L. McCarty. Defeasible deontic reasoning. *Fundamenta Informaticae*, 21:125–148, 1994.
39. G. E. Moore. *Principia Ethica*. Cambridge University Press, 1903.
40. A. Nayak, P. Nelson, and H. Polansky. Belief change as change in epistemic entrenchment. *Synthese*, 109:143–174, 1996.
41. E. Pacuit, R. Parikh, and E. Cogan. The logic of knowledge based on obligation. *Knowledge, Rationality and Action*, 149:311–341, 2006.
42. H. Prakken. Two approaches to the formalization of defeasible deontic reasoning. *Studia Logica*, 57:73–90, 1996.
43. H. Prakken and M. Sergot. Contrary-to-duty obligations. *Studia Logica*, 57:91–115, 1996.
44. A. Ross. *Directives and Norms*. Routledge & Kegan Paul, London, 1968.
45. Y. Shoham. A semantical approach to nonmonotonic logics. In M. Ginsberg, editor, *Readings in Nonmonotonic Reasoning*, pages 227–250. Morgan Kaufmann Publishers, 1987.
46. St. Paul. First letter to the corinthians. 7th Book New Testament.
47. St Thomas Aquinas. Summa theologiae. Translation by Thomas Gilby.
48. R. Stalnaker. A theory of conditionals. In N. Rescher, editor, *Studies in Logical Theory*, pages 98–112. Oxford, 1968.
49. A. Stolpe. A theory of permission based on the notion of derogation. *Journal of Applied Logic*, 8:97–113, 2010.
50. Y.-H. Tan and L. van der Torre. How to combine ordering and minimizing in a deontic logic based on preferences. In M. Brown and J. Carmo, editors, *Deontic Logic, Agency and Normative Systems, DEON '96: Third International Workshop on Deontic Logic in Computer Science*, pages 216–232, 1996.
51. J. van Benthem. Logical dynamics of information and interaction. To appear.
52. J. van Benthem, D. Grossi, and F. Liu. Deontics = betterness + priority. In G. Governatori and G. Sartor, editors, *Proceedings of the 10th International Conference on Deontic Logic in Computer Science (DEON 2010)*, LNCS, pages 50–65. Springer, 2010.
53. L. van der Torre. *Reasoning about obligations: defeasibility in preference-based deontic logic*. PhD thesis, Erasmus University Rotterdam, 1997.
54. L. van der Torre and Y.-H. Tan. The many faces of defeasibility in defeasible deontic logic: Essays in nonmonotonic normative reasoning. In D. Nute, editor, *Defeasible Deontic Logic*, volume 263 of *Synthese Library*, pages 79–121. Kluwer Academic Publishers, 1997.
55. B. van Fraassen. Values and the heart's command. *The Journal of Philosophy*, 70(1):5–19, 1973.
56. G. H. Von Wright. *Norm and Action. A Logical Inquiry*. Routledge, London, 1963.
57. W. Woods. What is a link: Foundations for semantic networks. In D. Bobrow and Collins A., editors, *Representation and Understanding: Studies in Cognitive Science*, pages 35–82, New York, 1975. Academic Press.

Appendix: an alternative view of P-sequences

A P-sequences, as introduced in Definition 1, is nothing but a system of nested Lewis' spheres [33] where each sphere consists in the truth-set of a given propositional formula. The innermost sphere is the smallest sphere and, on the ground

of Definition 2, it contains all the best states. The second sphere contains all the best states, plus the second best ones, and so on down to the last sphere, the biggest one. This contains all the second worst states, while the complement of the last sphere contains all the worst ones.

This, however, is only one specific option we chose for the definition of a P-sequence. Before moving to the next section and to the application of P-sequences to CTDs, we want to mention an alternative, but equivalent view, of P-sequences, which is to some extent closer in spirit to the general theory of priority developed in [35]. Here is the alternative definition. To distinguish this concept from the one of P-sequence, we refer to it as *disjoint* P-sequence.

Definition 9 (Disjoint P-sequence). *Let $\mathcal{L}(\mathbf{P})$ be a propositional language built on the set of atoms $\mathbf{P}$, S a non-empty set of states and $\mathcal{I} : \mathbf{P} \longrightarrow 2^S$ a valuation function. A P-sequence for $\mathcal{I}$ is a tuple $\mathcal{B}^{\mathcal{I}} = \langle B, \prec \rangle$ where:*

- *$B \subset \mathcal{L}(\mathbf{P})$ with $|B| < \omega$;*
- *$\prec$ is a strict linear order on B;*
- *for all $\varphi, \psi \in B$, $[\![\psi]\!]_{\mathcal{I}} \cap [\![\varphi]\!]_{\mathcal{I}} = \emptyset$.*

where $[\![\varphi]\!]_{\mathcal{I}}$ denotes the truth-set of φ according to $\mathcal{I}$.

The difference between Definition 1 and Definition 9 is that the latter requires all elements of the sequence to be disjoint, instead of being ordered according to set-theoretic inclusion. The intuition here is that all the states belonging to the truth-set of some proposition in the sequence are to be considered equivalent, while states belonging to the truth-set of a proposition with a higher position in the sequence are to be considered better than states belonging to the truth-set of a proposition with a lower position in the sequence. This intuition is made formal by defining a new procedure for extracting orders on states from disjoint P-sequences.

Definition 10 (Deriving betterness from disjoint P-sequences). *Let $\mathcal{B} = \langle B, \prec \rangle$ be a (disjoint) P-sequence, S a non-empty set of states and $\mathcal{I} : \mathbf{P} \longrightarrow 2^S$ a valuation function. The preference relation $\preceq_{\mathcal{B}}^{LEX} \subseteq S^2$ is defined as follows:*

$$s \preceq_{\mathcal{B}}^{LEX} s' := \forall \varphi \in B : (s \in [\![\varphi]\!] \Rightarrow s' \in [\![\varphi]\!] \text{ or } \exists \varphi' : (\varphi \prec \varphi' \text{ and } s \notin [\![\varphi']\!] \text{ and } s' \in [\![\varphi']\!])). \quad (23)$$

where LEX is just a mnemonics for 'lexicographic'.

The procedure works as a form of lexicographic ordering. The relation $s \preceq_{\mathcal{B}}^{LEX} s'$ holds if s satisfies all formulae in the sequence that also s' satisfies or, if that is not the case, if s' satisfies a formula which s does not satisfy and which is ranked higher in the P-sequence. In a disjoint P-sequence the effect of this latter clause is clear. Two states belonging to the truth-sets of two disjoint propositions will be ranked according to the ranking of the proposition whose truth-set they belong to.

It has by now probably not gone unnoticed that Definition 2 is a simplified version of Definition 10 consisting of only the first clause of the latter. In general, we can prove the following result connecting Definitions 1 and 2 to Definitions 9 and 10.

Fact 3 (Equivalence of P-sequence definitions) *The following holds:*

1. *Let $\mathcal{B} = (\varphi_1, \ldots, \varphi_n)$ be a P-sequence. There exists a disjoint P-sequence $\mathcal{B}'$ such that $\preceq^{IM}_{\mathcal{B}} = \preceq^{LEX}_{\mathcal{B}'}$.*
2. *Let $\mathcal{B} = (\varphi_1, \ldots, \varphi_n)$ be a disjoint P-sequence. There exists a P-sequence $\mathcal{B}'$ such that $\preceq^{LEX}_{\mathcal{B}} = \preceq^{IM}_{\mathcal{B}'}$.*

Proof (Sketch). The proof of both claims is by construction. [Claim 1:] Define $\mathcal{B}'$ as follows: $\mathcal{B}' = (\varphi_1, \neg\varphi_1 \wedge \varphi_2, \ldots, \neg\varphi_{n-1} \wedge \varphi_n)$. It is clear that all elements of $\mathcal{B}'$ are therefore disjoint (notice also that $\mathcal{B}$ and $\mathcal{B}'$ have the same length). From this construction it also follows directly that, if two states belong to a same equivalence class with respect to $\preceq^{IM}_{\mathcal{B}}$, so they do with respect to $\preceq^{LEX}_{\mathcal{B}'}$ and vice versa. Now for the strict case. Suppose $s \prec^{IM}_{\mathcal{B}} s'$. By Definition 2 this implies that $s' \in [\![\varphi_i]\!]$ and $s \notin [\![\varphi_i]\!]$ for $1 \leq i \leq n$. But this means that $s \in [\![\neg\varphi_i]\!]$ and hence that s belongs to the truth-set of some φ_j in $\mathcal{B}'$ with $i < j$. From which we can conclude that $s \prec^{LEX}_{\mathcal{B}'} s'$. Vice versa, if $s \prec^{LEX}_{\mathcal{B}'} s'$ then $s' \in [\![\psi_i]\!]$ and $s' \in [\![\psi_j]\!]$ for some $1 \leq i < j \leq n$. By the construction of $\mathcal{B}'$ we have that $\psi_i = \neg\varphi_{i-1} \wedge \varphi_i$ and $\psi_j = \neg\varphi_{j-1} \wedge \varphi_j$ (stipulate $\varphi_0 := \bot$) and therefore, since $i < j$, that $[\![\varphi_i]\!] \subset [\![\varphi_j]\!]$, from which we conclude that $s \prec^{IM}_{\mathcal{B}} s'$. [Claim 2:] Define $\mathcal{B}'$ as follows: $\mathcal{B}' = (\varphi_1, \varphi_1 \vee \varphi_2, \ldots, \varphi_{n-1} \vee \varphi_n)$. Clearly this is a P-sequence. The proof proceeds analogously to Claim 1. □

To conclude, the sort of P-sequences we have worked with in this paper are not the only one that are suitable for the deontic applications we have dealt with. All results we have presented can, in the light of Fact 3, be reformulated in terms of disjoint P-sequences and the lexicographic-like state ordering.