Beyond reasons and obligations: A dual-role approach to reasons and supererogation

Aleks Knoks [1]

University of Luxembourg
2, avenue de l'Université
L-4365 Esch-sur-Alzette

David Streit [1]

University of Luxembourg
2, avenue de l'Université
L-4365 Esch-sur-Alzette

Abstract

Dual-role approaches to reasons say, roughly, that reasons can relate to actions in two fundamentally different ways: they can either require conformity, or justify an action without requiring that it be taken. This paper develops a formal dual-role approach, combining ideas from defeasible logic and practical philosophy. It then uses the approach to shed light on the phenomenon of supererogation and resolve a well-known puzzle about supererogation, namely, Horton's All or Nothing Problem.

Keywords: reasons, dual-role approaches, defeasible logic, supererogation, all or nothing problem

1 Introduction

This paper has two goals. The first is to capture the core idea behind what we call *dual-role approaches* to reasons in a simple defeasible logic—in the style of the logic presented in [18]. The idea in question is, roughly, that the normative forces associated with reasons are of two fundamentally different kinds: some reasons for action *require* conformity, while others *justify* actions without requiring that they be taken. The second goal is to apply the resulting formal dual-role approach to the phenomenon of supererogation and to develop a unified response to the puzzles surrounding it.

[1] The first author acknowledges the support of the Fonds National de la Recherche Luxembourg (FNR) through the projects DELIGHT (OPEN O20/14776480). Both authors acknowledge the travel support from the same funding institution through the project INTEGRAUTO (INTER/AUDACE/21/16695098). We also thank our three anonymous reviewers for their insightful comments. The second reviewer's generous set of comments was particularly helpful.

The remainder of this paper is structured as follows. Section 2 provides a quick survey of the relevant literature on reasons, dual-role approaches, and formal work. Section 3 sets up the formal model. Section 4 shifts the focus to supererogation and explains how the dual-role approach that comes with the model responds to the basic challenge supererogation poses. Section 5 discusses a further puzzle of supererogation, namely, the *All or Nothing Problem*, [14]. Section 6 discusses related (formal) work on supererogation. The concluding Section 7 is followed by an appendix that contains the proofs of the most important results.

2 Reasons and dual-role approaches

Practical normative reasons are standardly characterized as considerations that count in favor or against actions.[2] Schroeder [41] helpfully points out the three "marks" that are characteristic of such reasons: they compete against each other, they are act-oriented, and they are the sorts of considerations that one can act for. The notion has become a mainstay of practical philosophy, where it is routinely made use of in answering various normative and metanormative questions. This is taken to the extreme in the *reasons-first program* which holds, roughly, that the notion of reasons is basic and that all other normative notions are to be analyzed in terms of it.[3]

For our purposes, two more recent developments in the literature on reasons will be particularly important. The first is the formal work on reasons, and, in particular, Horty's default logic-based model of the way reasons interact to support oughts.[4] Since the publication of *Reasons as Defaults* [18], this model has been extended in several ways and applied to many new problems, even finding a path into a more orthodox (that is, nonformal) monograph on reasons.[5] Other frameworks have been used to model reasons too—see, e.g., [6] and [7]—but default logic and defeasible logics more generally have been more influential. The second body of literature crucial for our interests develops what we call *dual-role approaches* to reasons.[6] Lately, several authors—most notably, Gert [8], [9], [10] and Greenspan [11], [12]—have argued that we need to distinguish two fundamentally different dimensions in the normative forces associated with reasons.[7] Thus, Gert discusses "requiring and justifying strengths" of reasons: the requiring strength is said to ground potential criticism and, through that,

[2] See, e.g., [34], [37], [38], [40], [44]. Whenever we say *reasons*, we always mean *normative* reasons, as opposed to explanatory or motivational reasons—see [1], for a discussion.

[3] The locus classicus here is Scanlon [38]. But see also, e.g., [34], [37], [41].

[4] We use the terms *ought* and *oughts* to refer to conclusions about what we ought to do.

[5] See, e.g., [15], [28], [32], and [41, Chs. 4.4–5].

[6] We borrow the term from [29].

[7] Both Gert and Greenspan urge to draw the distinction since it allows one to resolve various foundational issues in practical philosophy. For instance, Gert [8] shows how it can be used to the benefit of certain moral theories which, without the distinction, allow for cases in which the agent is forced to choose between an irrational moral action and an immoral rational one.

require conformity, while the justifying strength is said to ground answers to potential criticism and, thereby, *justify* nonconformity—see, e.g., [9, p. 541]. Importantly, some reasons are meant to be "purely justifying", meaning that they possess only the latter type of strength. In a similar vein, Greenspan [11,12] discusses "negative reasons" which count against an action and, without sufficient counterbalancing reasons, ".. [ground] a requirement to take some alternative option.." [11, p. 387] and "positive reasons" which count in favor of acts. Greenspan takes purely positive reasons to "ground at most only a recommendation". They "do not compel, but instead are optional, rendering an option eligible for choice, or justifying it, without requiring it" [11, p. 389]. [8] Gert's and Greenspan's views differ in details, but these won't matter for our purposes. Our main takeaway is their (common) core insight: reasons can relate to actions in two fundamentally different ways: They can have requiring force or (merely) justify an action. [9] The goal of the next section is to make this precise by expressing the core insight of dual-role approaches in a defeasible logic and combining the two strands found in the literature. The only other published attempt at formalizing dual-role approaches that we are aware of is due to Mullins [29]. While Mullins builds on Horty's model, like we do, his formalization differs from ours in several important respects. We compare our approach to his in Section 6.

3 The formal model

Let's start with a simple scenario:

> **Save One or Two**. Alice and Bob are trapped in a collapsing building. You can easily and without costs to yourself save one of them. You can also save both, but that would involve serious harm to you: you would lose your legs. [10]

Notice the three reasons that are particularly salient in this scenario: the fact that Alice will die, unless you save her; the fact that Bob will die, unless you save him; and the fact that you will lose your legs if you save both. Notice too, that all of the following judgments seem very intuitive: you *have to* save either Alice or Bob (we'd blame you if you walked away); it's not the case that you *have to* save both (we wouldn't blame you for deciding to keep your legs); but

[8] It pays noting that, in the literature on reasons, the terms *positive* and *negative reasons* are often applied to, respectively, reasons that count in favor of an action and those that count against—see, e.g., [37]. Clearly, this is very different from the way Greenspan uses these terms. To avoid confusion, we adopt Gert's terminology.

[9] Many other authors have drawn similar distinctions. This includes Dancy's [5] distinction between "enticing" and "peremptory" reasons, Parfit's [34] distinction between "partial" and "impartial" reasons, Portmore's [35] distinction between "moral" and "nonmoral" reasons, and Muñoz's [30] distinction between reasons and "prerogatives". The idea is always that we can distinguish two different dimensions in the way reasons—or reasons and considerations that aren't reasons—relate to actions.

[10] The scenario comes from [31].

were you to save both, your action would be highly admirable. [11]

We will now devise a formal notation that is just rich enough for a dual-role analysis of this scenario. As background, we assume the language of propositional logic with the standard connectives (including $\bot$), and we let the customary symbol $\vdash$ stand for classical logical consequence. Thus, we can use the propositional letters A, B, and L to express the propositions, respectively, that you save Alice, that you save Bob, and that you lose your legs. The constraint that you can't save Alice and Bob, as well as keep your legs, can be expressed as the material conditional $(A\&B) \supset L$. Extending the language slightly, we allow for formulas of the form $!X$ and read them as saying that there is a reason supporting the proposition expressed by X. What $!A$, $!B$, and $!\neg L$, then, say is, respectively, that there's a reason supporting your saving Alice, that there's a reason supporting your saving Bob, and that there's a reason supporting your not losing your legs. [12] For our purposes, it is not important to explicitly represent the reasons that ground such formulas as $!A$, $!B$, and $!\neg L$. In all the cases we will discuss, it won't matter what these reasons are exactly. What's more, we won't encounter any cases where the fact that a proposition is supported by multiple different reasons can make a difference for its analysis. In effect, this means that a formula of the form $!X$ can be read as "there is a reason supporting X" and also used to refer to the reason that grounds it. This is why we will often call such formulas *reasons*. We use $\mathcal{R}$ and $\mathcal{J}$ to denote (finite) collections of !-formulas: these will represent, respectively, *requiring* and *justifying reasons*—we adopt Gert's terminology. We also introduce the function $Conclusion(\cdot)$ that transforms !-formulas (and sets of such formulas) into ordinary propositional ones: thus, $Conclusion(!A) = A$ and $Conclusion(\{!L\}) = \{L\}$. The intuitive idea that some reasons have more weight than others will be captured by supplementing sets of !-formulas with a strict partial order. An expression of the form $!Y < !X$ should, then, be read as saying that the reason that grounds $!X$ has more weight than the reason that grounds $!Y$. [13]

We represent particular cases using the notion of a context:

Definition 3.1 [Contexts] A *context* Δ is a structure of the form $\langle \mathcal{W}, \mathcal{R}, \mathcal{J}, < \rangle$, where $\mathcal{W}$ is a consistent set of propositional formulas, $\mathcal{R}$ and $\mathcal{J}$ are finite sets of !-formulas, with the requirement that $\mathcal{R} \subseteq \mathcal{J}$, and $<$ is a strict partial order on $\mathcal{J}$. [14]

For illustration, we express Save One or Two in the context $\Delta_1 = \langle \mathcal{W}, \mathcal{R}, \mathcal{J}, < \rangle$ where $\mathcal{W} = \{(A\&B) \supset L\}$, $\mathcal{R} = \{!A, !B\}$, $\mathcal{J} = \{!A, !B, !\neg L\}$, and $<$ is empty.

---

[11] If your intuitions differ on this, consider upping the cost to yourself. Instead of losing your legs, you might lose your life.

[12] Similar notation is used in [17], [32, Appendix 2], and [45].

[13] We thank an anonymous reviewer for pressing us to clarify our conceptualization of formulas preceded by the ! (bang) operator.

[14] The constraint that $\mathcal{J}$ is finite keeps proofs in the appendix more manageable. It also fits well with the informal literature.

Given our intended interpretation of $\mathcal{R}$ and $\mathcal{J}$, the requirement that $\mathcal{R} \subseteq \mathcal{J}$ amounts to the idea that every requiring reason can serve as a justifying one. And the fact that $!\neg L$ is in $\mathcal{J}$, but not $\mathcal{R}$ formalizes the idea that the reasons that speaks against you losing your legs are purely justifying.

As our next step, we extend the language with three deontic operators. Thus, henceforth, we allow for formulas of the form $Ought(X)$, $Must(X)$, and $Can(X)$; and read them as saying, respectively, that it ought to be the case that X, that it is required, or that it must be the case, that X, and that it is permitted that X.[15] In what follows, we will often refer to these formulas as, respectively, *oughts*, *requirements*, and *permissions*. Before we specify a procedure for deriving such formulas from contexts, it pays noting that the emerging consensus in linguistics is that there are two distinct deontic necessities: a weaker one—typically ascribed using the modals *ought to* and *should*—and a stronger one—typically ascribed using *must* and *have to*.[16] Our oughts are meant to capture the weaker modality, while the requirements are meant to capture the stronger one.

Turning to the procedure, we need to specify how conflicts between reasons of different strength get resolved. A standard albeit simplistic move is to classify a reason r as "undefeated" if there is no stronger (requiring) reason r' such that $\mathcal{W} \cup Conclusion(r') \vdash \neg Conclusion(r)$.[17] Unfortunately, this approach won't do for us.[18] So, instead, we make use of a slightly more complex approach, motivated by the work of Brewka [4] and its characterization in [18, Ch. 8.2]. We start by defining two notions.

Definition 3.2 [Active reasons] Given a context $\Delta = \langle \mathcal{W}, \mathcal{R}, \mathcal{J}, < \rangle$ and $\mathcal{D} \subseteq \mathcal{J}$, let
$Active_{\Delta}(\mathcal{D}) = \{r \in \mathcal{J} : \mathcal{W} \cup Conclusion(\mathcal{D}) \cup Conclusion(r) \nvdash \bot \text{ and } r \notin \mathcal{D}\}$.

Thus, a reason r is active relative to a set of reasons $\mathcal{D}$ in case it is consistent with $\mathcal{D}$, but not (yet) in $\mathcal{D}$. The second notion we need is that of $<$-maximal elements:

Definition 3.3 [Maximal element] Given a set of reasons $\mathcal{D}$ and a preorder $<$ on $\mathcal{D}$, let $Max_{<}(\mathcal{D}) = \{r \in \mathcal{D} : \text{there is no } r' \in \mathcal{D} \text{ with } r < r'\}$.

Here is the basic idea of the Brewka-motivated approach: given a context $\langle \mathcal{W}, \mathcal{R}, \mathcal{J}, < \rangle$, we look at all possible ways of extending $<$ to a total order $<'$ on $\mathcal{J}$, and then, for each of those ways, we build a set of reasons whose

[15] The distinction between impersonal and personal obligations—as well as requirements and permissions—is orthogonal to our goals. So, we follow what Horty [17] calls the *policy of intentional, but harmless equivocation* and move freely between impersonal and personal reading of $Ought(X)$, $Must(X)$, and $Can(X)$.

[16] For the discussion of linguistic data, see e.g., [46], [36, pp. 79–81]; for its importance for ethical theory and reasons-first views in particular, see [3], [42], and [43], and for its importance for deontic logic, see [24].

[17] See, e.g., [17], [20], [29].

[18] For a critical discussion of this approach and a number of others, see [18, Ch. 8]. We can't use it, because it gives rise to counterexamples to our Proposition 5.1.

conclusions are consistent, starting with the empty set and iteratively selecting the $<'$-maximal element from among the reasons that are active at a given step. Our next two definitions make this idea precise.

Definition 3.4 [Brewka scenarios, for totally ordered contexts] Let $\Delta = \langle \mathcal{W}, \mathcal{R}, \mathcal{J}, < \rangle$ be a context where $<$ totally orders $\mathcal{J}$. Then $\mathcal{B}$ is the *Brewka scenario* of Δ just in case $\mathcal{B} = \bigcup_{i \geq 0} \mathcal{B}_i$, where the sequence $\mathcal{B}_0, \mathcal{B}_1, \mathcal{B}_2, \ldots$ is defined as follows:

$$\mathcal{B}_0 = \emptyset,$$

$$\mathcal{B}_{i+1} = \begin{cases} \mathcal{D}_i & \text{if } Active_{\Delta}(\mathcal{B}_i) = \emptyset \\ \mathcal{D}_i \cup Max_{<}(Active_{\Delta}(\mathcal{B}_i)) & \text{otherwise} \end{cases}$$

To illustrate, consider the context $\Delta_2 = \langle \mathcal{W}, \mathcal{R}, \mathcal{J}, < \rangle$ where $\mathcal{W} = \{A \supset \neg B, B \supset \neg C\}$, $\mathcal{R} = \mathcal{J} = \{!A, !B, !C\}$ and $!A < !B < !C$. Notice that $!A$ and $!C$ are compatible, while $!B$ conflicts with both of them. Now let's determine the unique Brewka scenario $\mathcal{B}$ of this context by constructing the sequence $\mathcal{B}_0, \mathcal{B}_1, \mathcal{B}_2, \ldots$ such that $\mathcal{B} = \bigcup_{i \geq 0} \mathcal{B}_i$. Clearly, $\mathcal{B}_0$ is the empty set. Since $Max_{<}(Active_{\Delta}(\mathcal{B}_0))$ equals $\{!C\}$, we have $\mathcal{B}_1 = \{!C\}$. Further, it is not difficult to see that $Max_{<}(Active_{\Delta}(\mathcal{B}_1))$ equals $\{!A\}$. Since $\mathcal{W} \cup Conclusion(\mathcal{B}_1) = \{A \supset \neg B, B \supset \neg C, A\}$ entails $\neg B$, the reason $!B$ is not in $Active_{\Delta}(\mathcal{B}_1)$, while the reason $!C$ is. As a reasult, we have $\mathcal{B}_2 = \{!A, !C\}$. After this step, there are no further active reasons that could be added, and so we have $\mathcal{B}_i = \mathcal{B}_2$ for every $i \geq 2$. At this point it should be clear that the Brewka scenario $\mathcal{B} = \bigcup_{i \geq 0} \mathcal{B}_i$ that we were looking for is $\{!C, !A\}$.

Our next definition extends the notion of a Brewka scenario to contexts that are not totally ordered.

Definition 3.5 [Brewka scenarios] Let $\Delta = \langle \mathcal{W}, \mathcal{R}, \mathcal{J}, < \rangle$ be any context. Then $\mathcal{B}$ is a *Brewka scenario* based on Δ just in case $\mathcal{B}$ is the Brewka scenario of some context $\langle \mathcal{W}, \mathcal{R}, \mathcal{J}, <' \rangle$ where $<'$ is a total order extending $<$.

Returning to the earlier Save One or Two scenario, there are six ways to extend the empty relation of Δ_1 to a total order. What is important to determining the Brewka scenarios of the resulting totally ordered contexts are only the two highest ranked reasons—how they are related to each other doesn't matter. If $!A$ and $!B$ are ranked the highest, the Brewka scenario is $\{!A, !B\}$. If $!A$ and $!\neg L$ are ranked the highest, the Brewka scenario is $\{!A, !\neg L\}$. Lastly, if $!B$ and $!\neg L$ are ranked the highest, we get $\{!B, !\neg L\}$. Thus, in total, there are three Brewka scenarios based on Δ_1.

In addition to Brewka scenarios, our procedure for deriving oughts, requirements, and permissions, will make use of the following auxiliary notion:

Definition 3.6 [Stable scenarios, restricted contexts] Given a context $\Delta = \langle \mathcal{W}, \mathcal{R}, \mathcal{J}, < \rangle$, a *stable scenario* based on Δ is any set $\mathcal{D}$ such that $\mathcal{R} \subseteq \mathcal{D} \subseteq \mathcal{J}$. Letting $<^{\mathcal{D}}$ stand for $<$ restricted to $\mathcal{D}$, we call the context $\langle \mathcal{W}, \mathcal{R}, \mathcal{D}, <^{\mathcal{D}} \rangle$ the *restriction of* Δ *to* $\mathcal{D}$ and denote it by $\Delta^{\mathcal{D}}$.

So, a stable scenario includes all requiring reasons and any set of justifying ones—which implies that $\mathcal{R}$ always qualifies as a stable scenario. For illustration, there are two stable scenarios based on Δ_1: $\{!A, !B\}$ and $\{!A, !B, !\neg L\}$.

We are finally in a position to specify the conditions under which oughts, requirements, and permissions follow from contexts. We start with oughts. Intuitively, these are obtained by restricting attention to requiring reasons and completely ignoring the justifying ones, and then looking at what follows from all Brewka scenarios that can be constructed from them.

Definition 3.7 [Oughts] Given a context $\Delta = \langle \mathcal{W}, \mathcal{R}, \mathcal{J}, < \rangle$, the formula $Ought(X)$ follows from Δ, written as $\Delta \mathrel{|\!\sim} Ought(X)$, just in case, $\mathcal{W} \cup Conclusion(\mathcal{B}) \vdash X$ for every Brewka scenario $\mathcal{B}$ based on $\Delta^{\mathcal{R}}$.

It's not difficult to verify that $Ought(A\&B)$ follows from Δ_1: you ought to save both Alice and Bob. Whereas oughts are determined on the basis of requiring reasons alone, requirements and permissions are determined on the basis of both types of reasons. The idea underlying our definitions is simple: $Must(X)$ follows from a context when, for every stable scenario based on the context, X is a consequence of all of its Brewka scenarios; and $Can(X)$ follows when, for some stable scenario based on the context, X is a consequence of one of its Brewka scenarios.

Definition 3.8 [Requirements] Given a context $\Delta = \langle \mathcal{W}, \mathcal{R}, \mathcal{J}, < \rangle$, the formula $Must(X)$ follows from it, $\Delta \mathrel{|\!\sim} Must(X)$, just in case, for every stable scenario $\mathcal{D}$ based on Δ, we have $\mathcal{W} \cup Conclusion(\mathcal{B}) \vdash X$ for every Brewka scenario $\mathcal{B}$ based on $\Delta^{\mathcal{D}}$.

Definition 3.9 [Permissions] Given a context $\Delta = \langle \mathcal{W}, \mathcal{R}, \mathcal{J}, < \rangle$, the formula $Can(X)$ follows from it, $\Delta \mathrel{|\!\sim} Can(X)$, just in case, for some stable scenario $\mathcal{D}$ based on Δ, we have $\mathcal{W} \cup Conclusion(\mathcal{B}) \vdash X$ for some Brewka scenario $\mathcal{B}$ based on $\Delta^{\mathcal{D}}$.

For illustration, the two stable scenarios based on Δ_1, one of which we've discussed in detail above, give rise to three Brewka scenarios: $\{A, B\}$, $\{A, \neg L\}$, and $\{B, \neg L\}$. Since $A \vee B$ follows from all of them, we have $\Delta_1 \mathrel{|\!\sim} Must(A \vee B)$. And since $A \vee B$ and $\neg L$ follow from some, we have $\Delta_1 \mathrel{|\!\sim} Can(A\&B)$ and $\Delta_1 \mathrel{|\!\sim} Can(\neg L)$. You *have to* save either Alice, or Bob; you can (and ought to) save both of them; and you can keep your legs. Thus, the model gets all the intuitions about Save One or Two right.

The model also has some nice properties. We register them here as a set of propositions—the proof of Proposition 3.2 is given in the appendix, the other two follow straightforwardly from the definitions:

Proposition 3.1 *For any context* $\Delta = \langle \mathcal{W}, \mathcal{R}, \mathcal{J}, < \rangle$*, (i) if* $\Delta \mathrel{|\!\sim} Must(X)$*, then* $\Delta \mathrel{|\!\sim} Ought(X)$*; and (ii) if* $\Delta \mathrel{|\!\sim} Ought(X)$*, then* $\Delta \mathrel{|\!\sim} Can(X)$*.*

Proposition 3.2 *For any context* $\Delta = \langle \mathcal{W}, \mathcal{R}, \mathcal{J}, < \rangle$*, neither* $\Delta \mathrel{|\!\sim} Ought(\bot)$*, nor* $\Delta \mathrel{|\!\sim} Must(\bot)$*, nor* $\Delta \mathrel{|\!\sim} Can(\bot)$*.*

Proposition 3.3 *Let* $\Delta = \langle \mathcal{W}, \mathcal{R}, \mathcal{J}, < \rangle$ *be an arbitrary context. Then (i)* $\Delta \mathrel{|\!\sim} Ought(X\&Y)$ *just in case both* $\Delta \mathrel{|\!\sim} Ought(X)$ *and* $\Delta \mathrel{|\!\sim} Ougth(Y)$ *and (ii)* $\Delta \mathrel{|\!\sim} Must(X\&Y)$ *just in case both* $\Delta \mathrel{|\!\sim} Must(X)$ *and* $\Delta \mathrel{|\!\sim} Must(Y)$.

Before we leave this section, let us answer two natural questions. The first concerns conditional oughts, requirements, and permissions. It's natural to wonder how these might be captured in our framework. It turns out that we can capture them by generalizing a familiar idea, going back at least to [16]. As a first step, we define the notion of updated contexts:

Definition 3.10 [Updated contexts] Given a context $\Delta = \langle \mathcal{W}, \mathcal{R}, \mathcal{J}, < \rangle$ and a formula X consistent with $\mathcal{W}$, the result of updating, or supplementing, Δ with X, written as $\Delta[X]$, is the context $\langle \mathcal{W} \cup \{X\}, \mathcal{R}, \mathcal{J}, < \rangle$.

Thus, the context $\Delta[X]$ is just like Δ, except that X is now taken to be an established fact. With the notion of updated contexts, we can specify when conditional deontic statements follow from a context as follows:

Definition 3.11 [Conditional oughts, requirements, and permissions] Let Δ be an arbitrary context. Then:
$\Delta \mathrel{|\!\sim} Ought(Y|X)$ just in case $\Delta[X] \mathrel{|\!\sim} Ought(Y)$;
$\Delta \mathrel{|\!\sim} Must(Y|X)$ just in case $\Delta[X] \mathrel{|\!\sim} Must(Y)$;
$\Delta \mathrel{|\!\sim} Can(Y|X)$ just in case $\Delta[X] \mathrel{|\!\sim} Can(Y)$.

The second natural question concerns Definitions 3.7, 3.8, and 3.9: one may wonder what prompts the choice of what's known as the *disjunctive account* (over the *conflict account*).[19] The short answer is that not much seems to hinge on it, given our purposes, and that, in deontic settings, the disjunctive account is the less committal of the two and so also safer to work with.

4 Supererogation and the standard account

Having set up the formal model, let's take a step back from it and reconsider the Save One or Two scenario. As we have already noted, there seems to be an intuitive sense of *ought* in which you ought to save Alice *and* Bob, but it's not the case that you *have to* do it. Still, saving Alice and Bob is not only permissible, but would also be highly admirable. In fact, there seems to be a clear intuitive sense in which it is the best thing you could do. From a third-person perspective, it certainly looks like this action leads to the best possible outcome, with all three people involved staying alive—although one of them severely inured.

And this means that saving Alice and Bob is a *supererogatory action* as it is an action that is ostensibly best, and yet it isn't obligatory. What Muñoz [30] calls the *Classic Paradox of Supererogation* is the challenge to explain the very possibility of such actions. Our formal approach has the resources to meet this challenge—which it inherits from the core idea of dual-role approaches. Thus, in response to the question of why saving Alice and Bob is the best action,

[19] See, e.g., [17] for a discussion.

we can say that it maximizes compliance with the requiring reasons at play in the scenario. In fact, carrying out $A\&B$, that is, saving Alice and Bob, means complying with all the requiring reasons at play in Δ_1, that is, $!A$ and $!B$. And in response to the question of why saving Alice and Bob isn't obligatory, our approach lets us point to the (purely) justifying reason $!\neg L$ and say that it can serve as an excuse to not comply with one of the requiring reasons. Notice that these answers straightforwardly generalize to other cases involving supererogatory actions, giving us a general response to the classic paradox.

There's another formal approach to supererogation—the titular standard approach—that resolves the classic paradox, namely, McNamara's *Doing Well Enough* framework [23], [24], [25].[20] McNamara works with ranked possible worlds: the higher a world's ranking, the (morally) better it is. Requirements are determined by a threshold: if X is true in all worlds above it, it's required that X. Permissions are duals of requirements: if it's not the case that $\neg X$ is required, it's permissible that X. Oughts in our sense are determined by the best worlds, they are "the most one can do": if X is true in all the best worlds, it ought to be that X. Also, since the best worlds are above the threshold, this gives the intuitive principle that requirements imply oughts.

This setup lets McNamara account for the intuitions in Save One or Two and respond to the challenge: saving Alice and Bob is best because the worlds where both get saved are ranked the highest; it is not obligatory because there are other worlds above the threshold where only one person is saved.[21]

So now we have seen two formal accounts of supererogation. The standard one might look more elegant and simple, but there's a serious problem with an account like this: transitively ranking all worlds and determining acceptability by means of a threshold imposes serious restrictions. It rules out scenarios where an impermissible act is superior to a permissible one—cf. [47]. The problem is that such scenarios seem possible:[22]

All or Nothing. Alice and Bob are, again, trapped in a collapsing building, but this time you will lose your legs whether you save one or both of them.[23]

Intuitively, worlds where only one person is saved are superior to those where none are. Nevertheless, walking away seems permissible, while saving only one person does not—it involves gratuitous loss of life. The threshold framework

[20] We see McNamara's work as a representative of the dominant approach to deontic modality in philosophy and linguistics, associated, among others, with [21] and [22]. The difficulties that McNamara faces are symptomatic of problems for this dominant approach. This is evidenced by the fact that Åqvist in [2], who defends an even more fine-grained threshold model with an arbitrary number of levels of goodness, still cannot accommodate the scenario we discuss in the next section in a natural and intuitive manner—at least not without giving up the intuitive notion of a threshold.

[21] Perhaps, a fully satisfactory explanation would need to say more about the ranking and threshold, but there are several plausible things to say here.

[22] In Section 6, we consider the question of how the standard account might be changed to address this problem.

[23] The case comes from [14].

says otherwise: since worlds where one person gets saved are better than those where none are, and it's permissible to walk away, it must be permissible for you to save only one.

Our model, by contrasts, easily handles the case. We express it as the context $\Delta_3 = \langle \mathcal{W}, \mathcal{R}, \mathcal{J}, < \rangle$, where $\mathcal{W} = \{A \supset L, B \supset L\}$, $\mathcal{R} = \{!A, !B\}$, $\mathcal{J} = \{!A, !B, !\neg L\}$, and $<$ is empty. It's not difficult to verify that $Can(\neg A \& \neg B)$ follows from Δ_3, and that neither $Can(A \& \neg B)$, nor $Can(\neg A \& B)$ do: while it's permissible for you to save neither Alice, nor Bob, it's not the case that you can save only one.

5 Horton's All or Nothing Problem

In addition to the classic paradox, supererogation gives rise to at least two other puzzles. Following [30], we call them the *All or Nothing Problem* [14] and the *Intransitivity Paradox* [19]. While our dual-role approach can resolve both puzzles, here we discuss only the former one, for reasons of space. It emerges as combinations of intuitions about the All or Nothing scenario and a plausible principle. We present the problem as a set of jointly inconsistent claims in English, staying close to Horton's [14] original formulation:

1. It's morally permissible to save neither Alice nor Bob. (intuition)
2. It's morally wrong for you to save only one of them. (intuition)
3. If an act X is morally permissible and an act Y is morally wrong—and X and Y are the only two available acts—one ought to do X, rather than Y. (intuitive principle)
4. You ought to save neither Alice nor Bob rather than save only one of them. (from 1–3)
5. But, clearly, (4) is false. (intuition)

Two notes are in order. First, the *oughts* in claims (3) and (4) aren't meant to immediately map onto our technical notion of ought. Rather, at this point, claims (1)–(5) are meant to express pretheoretical intuitions—as they do in Horton's statement of the puzzle. Second, the paradox appeals to the notion of comparative obligations. While this notion makes intuitive sense and is used by Horton, it certainly hasn't been the focus of much research in deontic logic. Luckily, it seems possible to capture this notion in terms of conditional obligations: to say that one ought to do X, rather than Y is just to say that one ought to do X in case $X \vee Y$. [24] Bearing this in mind and letting A and B express the same propositions they did before, we propose to express the problem in our formal notation as follows—which, we contend, sharpens it:

1. $Can(\neg A \& \neg B)$ (intuition)
2. $Must(\neg([A \& \neg B] \vee [\neg A \& B]))$ (intuition)
3. If $Can(X)$ and $Must(\neg Y)$, then $Must(X | X \vee Y)$ (intuitive principle)

[24] In his original statement of the problem, Horton suggests this much—see [14, fn. 2].

4. If $Can(\neg A \& \neg B)$ and $Must(\neg([A \& \neg B] \vee [\neg A \& B]))$, then $Must(\neg A \& \neg B | [\neg A \& \neg B] \vee ([A \& \neg B] \vee [\neg A \& B]))$ (instance of the principle)
5. $Must(\neg A \& \neg B | [\neg A \& \neg B] \vee ([A \& \neg B] \vee [\neg A \& B]))$ (from 1, 2, and 4)
6. $Must(\neg A \& \neg B | \neg [A \& B])$ (substitution of equivalent formulas)
7. But, clearly, not $Must(\neg A \& \neg B | \neg [A \& B])$ (intuition)

It's worth being explicit about two assumptions in the background of our formalization. First, we take the oughts in the original claims to express the stronger deontic modals, what we called *requirements*. Second, we are assuming that if an action is morally wrong, there's a requirement forbidding taking this action. Both assumptions strike us as very plausible. What our formalization, then, does is show that All or Nothing is, indeed, a genuine puzzle, and that, their intuitive character notwithstanding, we cannot hold onto claims (1)–(3) and (7) on pain of inconsistency.

Our model happens to solve this puzzle, suggesting that the fault lies with the principle expressed in (3). First off, the principle's counterpart

If $\Delta \mathrel{|\!\sim} Can(X)$ *and* $\Delta \mathrel{|\!\sim} Must(\neg Y)$*, then* $\Delta \mathrel{|\!\sim} Must(X | X \vee Y)$

is demonstrably false. This is witnessed by the context $\Delta_4 = \langle \mathcal{W}, \mathcal{R}, \mathcal{J}, < \rangle$ where $\mathcal{W} = \{C \supset \neg D, D \supset \neg E, E \supset \neg C\}$, $\mathcal{R} = \{!C, !D\}$, $\mathcal{J} = \{!C, !D, !E\}$, and $!D < !C$. It's quite easy to verify that we have both $\Delta_4 \mathrel{|\!\sim} Can(E)$ and $\Delta_4 \mathrel{|\!\sim} Must(\neg D)$, while we don't have $\Delta_4 \mathrel{|\!\sim} Must(E | D \vee E)$. What's more, it can be shown that two principles in the vicinity hold true in the model—the proofs are provided in the appendix:

Proposition 5.1 *For any context* Δ*,*
(i) if $\Delta \mathrel{|\!\sim} Can(X)$ *and* $\Delta \mathrel{|\!\sim} Must(\neg Y)$*, then* $\Delta \mathrel{|\!\sim} Can(X | X \vee Y)$*;*
(ii) if $\Delta \mathrel{|\!\sim} Ought(X)$ *and* $\Delta \mathrel{|\!\sim} Must(\neg Y)$*, then* $\Delta \mathrel{|\!\sim} Ought(X | X \vee Y)$*.* [25]

The fact that these principles hold can explain the intuitive pull of the original principle. Our approach also makes clear where the original principle goes wrong: it attempts to bridge unconditional and conditional deontic statements without keeping track of the types of reasons that these statements depend on.

6 Related work

This section compares our model to Mullin's [29] dual-role approach to reasons and briefly discusses related work on supererogation. [26] After discussing

[25] To be fair, both principles are immediate consequences of more general principles that hold in the model, as the proofs in the appendix make manifest. An anonymous reviewer suggests that this weakens our claim that the principles we propose account for the intuitive pull of the original principle. While we share the intuition that, it would be a nice feature of the model, if our principles wouldn't be mere corollaries of more general ones, it is not immediately clear to us why the claim is weakened. In any event, Proposition 5.1 is the best we have for now, and it might well be that our model validates other principles that could serve its function, or serve it better.

[26] We focus on recent work on supererogation. For a historical perspective and its relevance to current topics see [26].

Mullins, we revisit McNamara's threshold account and consider his recent extension to conditional operators. Then we discuss Wessels' [47] quasi decision-theoretic approach—which discusses cases like All or Nothing—and Hansson's [13] approach.

6.1 Mullins dual-role approach

Mullins starts with Horty's model [18] and discusses two ways to capture the distinction between requirements and oughts. The first appeals to a threshold of strength, the second one—which we focus on—distinguishes between two distinct types of reasons in the spirit of dual-role approaches. [27]

Unlike us, Mullins relies on the simple approach to defeat, and his strategy is to, first, specify when a context entails a requirement—that is, a *Must*-formula—and then, in the second step, use this as a basis for determining which permissions and oughts this context entails. More precisely, $Can(X)$ is set to follow from a context just in case $Must(\neg X)$ does not follow, and $Ought(X)$ is set to follow just in case, roughly, the reasons that entail X are compatible with the reasons that allow for the derivation of *Must*-formulas. Explaining his strategy, Mullins writes:

> We first identify our undefeated requiring reasons, in order to determine what is required or impermissible. Oughts are then supported by our best justifying reasons, provided the consequences of their conclusions are consistent with some maximal subset of requiring reasons [29, p. 586].

To see Mullins' model at work, we revisit the familiar Save One or Two scenario. We captured it in the context $\Delta_1 = \langle \mathcal{W}, \mathcal{R}, \mathcal{J}, < \rangle$ where $\mathcal{W} = \{(A\&B) \supset L\}$, $\mathcal{R} = \{!A, !B\}$, $\mathcal{J} = \{!A, !B, !\neg L\}$, and $<$ empty. To determine which *Must*-formulas follow from it, we are to look at what follows from the subsets of undefeated requiring reasons—that is, the subsets of $Conclusion(\mathcal{R}) = \{A, B\}$—that are maxiconsistent with $\mathcal{W}$. This, however, gives us the counterintuitive result that both $Must(A\&B)$ and $Must(L)$ follow from Δ_1: you have to save Alice and Bob, and lose your legs. One might take this to mean that the scenario has to be captured in a different context, and the most natural alternative that suggests itself is $\Delta_5 = \langle \mathcal{W}, \mathcal{R}, \mathcal{J}, < \rangle$ where $\mathcal{W} = \{(A\&B) \supset L\}$, $\mathcal{R} = \emptyset$, $\mathcal{J} = \{!A, !B, !\neg L\}$, and $<$ is empty. [28] Even barring the counterintuitive im-

[27] See [29, Secs. 4 and 5]. The basic idea behind the first way to capture the distinction is that only reasons above a certain threshold can support requirements. Mullins attributes the idea to Scanlon [39].

[28] Another possible candidate is the context $\Delta_6 = \langle \mathcal{W}, \mathcal{R}, \mathcal{J}, < \rangle$ where $\mathcal{R} = \{!(A \vee B)\}$ and the rest is like in Δ_5. But while this secures the intuitive results that both $Must(A \vee B)$ and $Ought(A\&B)$ are derivable, there's good reason to be dissatisfied with this context. Most importantly, the inclusion of a *disjunctive* requiring reason looks terribly ad hoc, since it amounts to hard-coding the desired intuition. Also, $Ought(\neg L)$ follows from Δ_6, just like it does in the case of Δ_5. An anonymous reviewer worries that leaving the ordering $<$ empty in the representation of the case stacks the cards, since, in Mullins' model, requiring reasons can get defeated by justifying ones. While the reviewer's reaction is certainly reasonable, we couldn't think of any way the ordering might be used to get Mullins' account to deliver the right result: setting $!A$, $!B < !\neg L$ fails to deliver the intuitive $Must(A \vee B)$, while adding

plication that the fact that Alice and Bob are in danger doesn't exert requiring normative force, we don't get a good match with intuitions. First, $Must(A \vee B)$ doesn't follow from Δ_5. Second, while $Ought(A\&B)$ follows from Δ_5, so does $Ought(\neg L)$, suggesting that not losing your legs is optimal.[29]

This invites the conclusion that Mullins' model has serious trouble accommodating the Save One or Two scenario. There are other issues with it too, but we won't dwell on them and simply state (what we take to be) the underlying problem: it determines requirements almost exclusively on the basis of requiring reasons, not giving justifying reasons their due.[30] Admittedly, this appears to be the default approach in the (informal) philosophical literature—see, e.g., [43]—but it doesn't seem to work once expressed in a defeasible logic-based framework.

6.2 Doing Well Enough

Since McNamara's framework was already introduced in Section 4, here we confine ourselves to some brief remarks focusing on its conditionalized version, as developed in [27], and briefly sketch some worries whether, if at all, it might accommodate cases like Horton's *All or Nothing* scenario. The main advance of [27] is the provision of formal tools to capture *conditionally* acceptable worlds, in the style of Dyadic Deontic Logic [33, Ch. 2]. Acceptable worlds are those worlds that are above the threshold or "good enough", and any proposition true in one of these world is permitted. This allows one to formalize Horton's All or Nothing Problem using conditional obligations, like we suggest in Section 5. McNamara's analogues of our *Must*-, *Can*-, and *Ought*-operators are, respectively, $OB(\cdot)$, $PE(\cdot)$, and $MA(\cdot)$, "the most one can do". Both $OB(\cdot)$ and $MA(\cdot)$ function like standard dyadic operators. This has the consequence that the principle $PE(X)\&OB(\neg Y) \supset OB(X|X \vee Y)$ is a theorem in McNamara's logic as he presents it in [27]. The fact that this principle holds, depends crucially on the semantic principle that there is a threshold: if a world is acceptable, then any world better than it is acceptable as well. An anonymous reviewer notes that it is possible to give up this principle. This is true, but a challenge remains: one has to account for the (remaining) claims that comprise the All or Nothing puzzle without simply hard-coding which actions are

$!(A\&B)$ to $\mathcal{R}$ and setting $!(A\&B) < !\neg L$ (as the reviewer appears to suggest) doesn't really change anything.

[29] Mullins' strategy uses the *conflict account* in determining which oughts follow from a context. A natural idea is to substitute it with the *disjunctive account*—Shyam Nair suggested this much in his keynote talk at the DEON2020/21 conference. Unfortunately, this move doesn't solve the problem: $Ought(\neg L)$ no longer follows from the context, but neither does $Ought(A\&B)$.

[30] There is only one way in which justifying reasons can have an impact on the requirements in Mullins' model: they can outright defeat requiring reasons. This, however, appears to be not enough to get the cases right—see Footnote 28. We thank an anonymous reviewer for pushing us to clarify our take on Mullins' approach.

permitted and which are not into the logical description of the case. [31]

All in all, while McNamara's (conditional) system allows one to express normative notions that we cannot easily capture—like "the least one can do"—we take it to be a serious challenge to modify it so that it can account for puzzles surrounding supererogation (of which All or Nothing is one) in a natural way. [32] In any event, we should be wary of any framework committed to the existence of a threshold, since it implies some of our intuitions about the All or Nothing scenario must be mistaken.

6.3 Other recent proposals

Wessels [47] proposes a very different account to accommodate supererogation. It is not a full-fledged logic, but still instructive. Wessels' explicit goal is to account for cases like the All or Nothing scenario, or cases involving what she calls "supererogation holes".

In Wessels' account, actions (instead of worlds) are totally ordered by their respective goodness, and an actions' "being supererogatory with respect to another action" is used to define supererogation simpliciter. The core idea is that an action is supererogatory with respect to another action just in case the relation between gained moral value and burden to the agent is above a threshold. [33] Using this construction, Wessels then defines supererogatory actions as follows.

> An action f_j is supererogatory just in case the answers to all three subquestions is yes:
> (1) Is there an action f_i such that f_j is supererogatory with respect to f_i?
> (2) Are all the actions that are morally better than f_j supererogatory with respect to f_j?
> (3) Are all the actions that are morally better than f_i supererogatory with respect to f_i?

Notice how this way of capturing supererogation lets her say that, in the All or Nothing scenario, saving either only Alice or only Bob is not supererogatory: since the act of saving both Alice and Bob doesn't put an additional burden on the agent when compared to saving either only Alice or only Bob, it's not supererogatory with respect to these other acts, and so the answer to the second question is negative.

[31] Note that our account makes no such assumptions. Which (conditional) actions are permitted, obligatory, or required follows from the interplay of reasons and their strength alone.

[32] In addition to the issue discussed in the previous paragraph, McNamara's framework faces a second problem, which we can only hint at here. As is, it validates the principle $PE(X|Y \vee X)\&PE(Z|X \vee Y) \supset PE(Z|Y \vee Z)$ which certain well-known cases involving supererogation bring into doubt. Here, too, the framework would have to be modified to account for this fact. See [19], as well as [31] for a discussion.

[33] Wessels uses real numbers to represent this in the style of rational choice theory with some restrictions. For instance, one action is allowed to be supererogatory with respect to another one just in case the moral value of the first action is at least as high as the moral value of the second one.

While we set out to develop a formal model that can solve some puzzles surrounding supererogation, Wessels' aims are more moderate. What her framework shows, in effect, is that if a logic can solve the problem of defining "X supererogatory with respect to Y", then a preference-based logic might take care of the rest. What Wessels doesn't do is develop such a logic. [34]

Hansson [13], unlike Wessels, proposes a logic-based account of supererogation which, like Wessels' account, builds on the relation "p is supererogatory relative to q". Hansson's idea is to set it that p is supererogatory relative to q if q is obligatory and p is "a better variant of q". Betterness is spelled out in terms of a preference relation, whereas "is a version of" is a primitive spelled out in terms of logical strength: thus, $p \vdash q$ means that p is a variant of q. This approach seems to be overly simplistic as it faces two challenges. First, not every supererogatory action is a variant of some obligatory action. For example, in the All or Nothing case, the supererogatory action is saving Alice and Bob, while no action at all appears to be obligatory. The second challenge is that, in modeling such scenarios, the choice of which actions are variants of one another is threatened to become entirely ad hoc. [35]

7 Conclusion

We set ourselves two goals in this paper. The first was to express the core of dual-role approaches to reasons in a defeasible logic. To reach this goal, we extended Horty's influential default logic-based model [18] in a number of ways. Our second goal had to do with supererogation, and we saw how our dual-role approach provides a unified response to the Classic Paradox of Supererogation and the All or Nothing Problem. What's more, we noted some advantages that our model has over alternative (formal) approaches to supererogation.

We see several promising directions for future research. First, our approach seems to let us solve another notorious puzzle about supererogation, namely, Kamm's *Intransitivity Paradox* [19], and we plan to discuss the issue in detail in a follow-up paper. Second, it would be interesting to explore how dual-role approaches to reasons might be captured in other frameworks that have been used to model reasons, such as structured argumentation or justification logic. Relatedly, it seems worthwhile to relate our model to input/output logic with permission—the latter looks like a more general system. Third, it might pay exploring further applications of formalized dual-role approaches. For instance, our model might have something interesting to say about the puzzles associated with permission. Lastly, it's worth thinking about the commitments of the particular dual-role approach that comes with the model and its potential advantages over the dual-role views defended in the philosophical literature.

[34] See [26] for a critical discussion of Wessels in the context of deontic logic.

[35] It pays noting that, in a critical discussion of Hansson [13], McNamara [26] suggests a way to ameliorate the second challenge by means of introducing a dyadic action operator standing for "an agent brings it about that q by bringing about that p". We suspect, however, that this move makes the first challenge more pressing, since it imposes further restrictions on what can count as a variant of an action.

Appendix

Proposition 3.2 *For any context* $\Delta = \langle \mathcal{W}, \mathcal{R}, \mathcal{J}, < \rangle$*, neither* $\Delta \mathrel{|\!\sim} Ought(\bot)$*, nor* $\Delta \mathrel{|\!\sim} Must(\bot)$*, nor* $\Delta \mathrel{|\!\sim} Can(\bot)$*.*

Proof. It is enough to show that $\Delta \mathrel{|\!\sim} Can(\bot)$ does not hold, since by Proposition 3.1, if one of the other statements were to hold, so would $\Delta \mathrel{|\!\sim} Can(\bot)$.

Suppose, toward a contradiction, that $\Delta \mathrel{|\!\sim} Can(\bot)$. This implies that there is a stable scenario $\mathcal{D}$ and a Brewka scenario $\mathcal{B}$ based on $\Delta^{\mathcal{D}} = \langle \mathcal{W}, \mathcal{R}, \mathcal{D}, <^{\mathcal{D}} \rangle$ such that $\mathcal{W} \cup Conclusion(\mathcal{B}) \vdash \bot$. Let $<^*$ be the ordering that extends $<^{\mathcal{D}}$ to a total order over $\mathcal{D}$, and that is used in the construction of $\mathcal{B}$. Given that $\mathcal{W}$ is consistent and that $\mathcal{B}$ is the limit of the sequence $\mathcal{B}_0, \mathcal{B}_1, \ldots$, we can be sure that there is some i such that $\mathcal{W} \cup Conclusion(\mathcal{B}_i) \nvdash \bot$, while $\mathcal{W} \cup Conclusion(\mathcal{B}_{i+1}) \vdash \bot$. But $\mathcal{B}_{i+1} = \mathcal{B}_i \cup Max_{<^*}(Active_{\langle \mathcal{W}, \mathcal{R}, \mathcal{D}, <^* \rangle}(\mathcal{B}_i))$, and $Max_{<^*}(Active_{\langle \mathcal{W}, \mathcal{R}, \mathcal{D}, <^* \rangle}(\mathcal{B}_i))$ is the singleton set $\{r \in \mathcal{D} : \mathcal{W} \cup Conclusion(\mathcal{B}_i) \cup Conclusion(r) \nvdash \bot$ and $r \notin \mathcal{B}_i\}$. Given that $\mathcal{W} \cup Conclusion(\mathcal{B}_i) \cup Conclusion(r) = \mathcal{W} \cup Conclusion(\mathcal{B}_{i+1})$, we have arrived at a contradiction. □

Before we turn to the proof of Proposition 5.1, we establish two lemmas.

Lemma 1 Given a context $\Delta = \langle \mathcal{W}, \mathcal{R}, \mathcal{J}, <' \rangle$, where $<'$ is a total order over $\mathcal{J}$, and a Brewka scenario $\mathcal{B}$ based on Δ, we have $\mathcal{W} \cup Cocnlusion(\mathcal{B}) \nvdash \bot$.

Proof. The lemma follows by an easy induction on the construction of $\mathcal{B}$. □

Lemma 2 Let $\Delta = \langle \mathcal{W}, \mathcal{R}, \mathcal{J}, <' \rangle$ be a context with $<'$ a total order on $\mathcal{J}$, $\mathcal{B}$ the Brewka scenario of Δ with $\mathcal{W} \cup Conclusion(\mathcal{B}) \vdash X$, and $\mathcal{B}^*$ the Brewka scenario of the context $\Delta[X \vee Y] = \langle \mathcal{W} \cup \{X \vee Y\}, \mathcal{R}, \mathcal{J}, <' \rangle$. Then $\mathcal{B} = \mathcal{B}^*$.

Proof. Before getting into the proof, note that both $\mathcal{B}$ and the sequence it is the limit of are unique. This is due ot the fact that at each step i of the construction of $\mathcal{B}$ there's at most one $<'$-maximal reason in $Active_{\Delta}(\mathcal{B}_i)$. We will show that, for each step i, $\mathcal{B}_i = \mathcal{B}_i^*$. We do this by induction on i.

The base case is trivial: $\mathcal{B}_0 = \emptyset = \mathcal{B}_0^*$.

For the induction step, assume that $\mathcal{B}_i = \mathcal{B}_i^*$. Given our definition of Brewka scenarios, it's enough to establish that $Max_{<'}(Active_{\Delta}(\mathcal{B}_i)) = Max_{<'}(Active_{\Delta[X \vee Y]}(\mathcal{B}_i^*))$. So that's what we turn to.

$\subseteq$: Consider some $r \in Max_{<'}(Active_{\Delta}(\mathcal{B}_i))$. Then $r \in \mathcal{B}_{i+1} \subseteq \mathcal{B}$. We know that $\mathcal{W} \cup Conclusion(\mathcal{B}) \vdash X$, and hence that $\mathcal{W} \cup Conclusion(\mathcal{B}) \vdash X \vee Y$. By Lemma 1, $\mathcal{W} \cup Conclusion(\mathcal{B})$ is consistent. Hence, as it entails $X \vee Y$, it is also consistent with $X \vee Y$. Since $\mathcal{W} \cup Conclusion(r) \cup Conclusion(\mathcal{B}_i)$ is a subset of $\mathcal{W} \cup Conclusin(\mathcal{B})$, we can be sure that $\mathcal{W} \cup \{X \vee Y\} \cup Conclusion(B_i) \cup Conclusion(r) \nvdash \bot$. This suffices to conclude that $r \in Active_{\Delta[X \vee Y]}(\mathcal{B}_i)$. We still need to show that r is $<'$-maximal in this set. We know that r is $<'$-maximal in $Active_{\Delta}(\mathcal{B}_i)$. Hence, for every $r' > r$ such that $r' \notin \mathcal{B}_i$, $r' \notin Active_{\Delta}(\mathcal{B}_i)$. But this means that, for every such r', $\mathcal{W} \cup Conclusion(\mathcal{B}_i) \cup Conclusion(r') \vdash \bot$, and, hence, by monotonicity, that $\mathcal{W} \cup \{X \vee Y\} \cup Conclusion(\mathcal{B}_i) \cup Conclusion(r') \vdash \bot$. This means that each such r' is not in $Active_{\Delta[X\vee Y]}(\mathcal{B}_i)$, and that r is indeed maximal here. By the inductive hypothesis, $\mathcal{B}_i = \mathcal{B}_i^*$, and so we have shown that $r \in Max_{<'}(Active_{\Delta[X\vee Y]}(\mathcal{B}_i^*))$.

$\supseteq$: Suppose that there is an $r \in Max_{<'}(Active_{\Delta[X\vee Y]}(\mathcal{B}_i^*))$. Further, suppose, toward a contradiction, that $r \notin Max_{<'}(Active_{\Delta}(\mathcal{B}_i))$. Either $r \in Active_{\Delta}(\mathcal{B}_i)$ or not. If not, then either (i) $r \in \mathcal{B}_i$, or (ii) $\mathcal{W} \cup Conclusion(\mathcal{B}_i) \cup Conclusion(r) \vdash \bot$. If (i), then $\mathcal{B}_i \neq \mathcal{B}_i^*$. If (ii), then $\mathcal{W} \cup Conclusion(\mathcal{B}_i^*) \cup Conclusion(r) \vdash \bot$, and $r \notin Active_{\Delta[X\vee Y]}(\mathcal{B}_i^*)$ after all. Hence, $r \in Active_{\Delta}(\mathcal{B}_i)$, but not $<'$-maximal. Let r' be the $<'$-maximal reason in $Active_{\Delta}(\mathcal{B}_i)$. Now we can reuse the argument we made use of above to conclude that $r' \in Active_{\Delta[X\vee Y]}(\mathcal{B}_i^*)$, and that $r \notin Max_{<'}(Active_{\Delta[X\vee Y]}(\mathcal{B}_i^*))$ after all. This gives us a contradiction. □

Proposition 5.1 *For any context* Δ*,*
(i) if $\Delta \mid\!\sim Can(X)$ *and* $\Delta \mid\!\sim Must(\neg Y)$*, then* $\Delta \mid\!\sim Can(X|X \vee Y)$*;*
(ii) if $\Delta \mid\!\sim Ought(X)$ *and* $\Delta \mid\!\sim Must(\neg Y)$*, then* $\Delta \mid\!\sim Ought(X|X \vee Y)$*.*[36]

Proof. We establish claim (i) by proving a stronger claim, namely, that if $\Delta \mid\!\sim Can(X)$, then $\Delta \mid\!\sim Can(X|X \vee Y)$. Suppose that $\Delta \mid\!\sim Can(X)$. It follows that there is a stable scenario $\mathcal{D}$ based on Δ and a total order $<'$ on $\mathcal{D}$ that extends $<$ such that $\mathcal{W} \cup Conclusion(\mathcal{B}) \vdash X$ for the Brewka scenario based on $\langle \mathcal{W}, \mathcal{R}, \mathcal{D}, <' \rangle$. Note that $\mathcal{D}$ is a stable scenario of $\Delta[X \vee Y]$, and that $<'$ is a total order extending $<$ in this restricted updated context as well. Set Δ^* to be the context $\langle \mathcal{W} \cup \{X \vee Y\}, \mathcal{R}, \mathcal{D}, <' \rangle$. By Lemma 2, we know that $\mathcal{B}^* = \mathcal{B}$ where $\mathcal{B}^*$ is the Brewka scenario based on Δ^*. Since $\mathcal{W} \cup Conclusion(\mathcal{B}) \vdash X$, we immediately get $\mathcal{W} \cup Conclusion(\mathcal{B}^*) \vdash X$, and, by monotonicity of classical logic, $\mathcal{W} \cup \{X \vee Y\} \cup Conclusion(\mathcal{B}^*) \vdash X$. This means that there is a stable scenario of $\Delta[X \vee Y]$, namely, $\mathcal{D}$, and a Brewka scenario based on $\Delta[X \vee Y]^{\mathcal{D}}$, namely, $\mathcal{B}^*$, such that $\mathcal{W} \cup \{X \vee Y\} \cup Conclusion(\mathcal{B}^*) \vdash X$. Given our definition of permissions, this is enough to conclude that $\Delta[X \vee Y] \mid\!\sim Can(X)$, and hence that $\Delta \mid\!\sim Can(X|X \vee Y)$.

For Claim (ii), we prove something stronger, namely, that if $\Delta \mid\!\sim Ought(X)$, then also $\Delta \mid\!\sim Ought(X|X \vee Y)$. Suppose that $\Delta \mid\!\sim Ought(X)$. This means that, for any Brewka scenario $\mathcal{B}$ of the context $\Delta^{\mathcal{R}}$, we have $\mathcal{W} \cup Conclusion(\mathcal{B}) \vdash X$. What we need to show is that, for any Brewka scenario $\mathcal{B}^*$ based on $\Delta[X \vee Y]^{\mathcal{R}}$, we have $\mathcal{W} \cup \{X \vee Y\} \cup Conclusion(\mathcal{B}^*) \vdash X$. Suppose, toward a contradiction, that this wasn't the case. So there is a context $\Delta^* = \langle \mathcal{W} \cup \{X \vee Y\}, \mathcal{R}, \mathcal{R}, <' \rangle$, where $<'$ extends $<$ to a total order over $\mathcal{R}$, such that $\mathcal{W} \cup \{X \vee Y\} \cup Conclusion(\mathcal{B}^*) \nvdash X$ for the

[36] To be fair, both principles are immediate consequences of more general principles that hold in the model, as the proofs in the appendix make manifest. An anonymous reviewer suggests that this weakens our claim that the principles we propose account for the intuitive pull of the original principle. While we share the intuition that, it would be a nice feature of the model, if our principles wouldn't be mere corollaries of more general ones, it is not immediately clear to us why the claim is weakened. In any event, Proposition 5.1 is the best we have for now, and it might well be that our model validates other principles that could serve its function, or serve it better.

Brewka scenario $\mathcal{B}^*$ based on Δ^*. Consider the context $\langle \mathcal{W}, \mathcal{R}, \mathcal{R}, <' \rangle$. From above, we can be sure that, for the Brewka scenario $\mathcal{B}$ based on it, we have $\mathcal{W} \cup Conclusion(\mathcal{B}) \vdash X$. By Lemma 2, we have $\mathcal{B}^* = \mathcal{B}$. (Recall that $\mathcal{B}^*$ is unique.) Hence, $\mathcal{W} \cup Conclusion(\mathcal{B}^*) \vdash X$, and, by the monotonicity of classical logic, $\mathcal{W} \cup \{X \vee Y\} \cup Conclsuion(\mathcal{B}^*) \vdash X$. And this is a contradiction. □

References

[1] Alvarez, M., *Reasons for action: Justification, motivation, explanation*, in: E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*, 2016, winter 2016 edition .

[2] Åqvist, L., *Three characterizability problems in deontic logic*, Nordic Journal of Philosophical Logic **5** (2000), pp. 65–82.

[3] Bedke, M., *Passing the deontic buck*, Oxford Studies in Metaethcis **6** (2011), pp. 128–53.

[4] Brewka, G., *Reasoning about priorities in default logic*, in: *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-94)*, 1994, pp. 940–5.

[5] Dancy, J., *Enticing reasons*, in: R. J. Wallace, P. Pettit, S. Scheffler and M. Smith, editors, *Reasons and Values: Themes from the Moral Philosophy of Joseph Raz*, Oxford University Press, 2004 pp. 91–118.

[6] Dietrich, F. and C. List, *A reason-based theory of rational choice*, Noûs **47(1)** (2013), pp. 104–34.

[7] Faroldi, F., *Common law precedent in the logic of reasons*, in: S. Rahman, M. Armgardt and H. Kvernenes, editors, *New Systems and Historic Studies in Legal Reasoning and Logic*, Springer, 2022 pp. 301–19.

[8] Gert, J., "Brute Rationality: Normativity and Human Action," Oxford University Press, 2004.

[9] Gert, J., *Normative strength and the balance of reasons*, Philosophical Review **116** (2007), pp. 533–62.

[10] Gert, J., *Perform a justified option*, Utilitas **26** (2014), pp. 206–17.

[11] Greenspan, P., *Asymmetrical practical reasons*, in: J. C. Marek and M. E. Reicher, editors, *Experience and Analysis: Proceedings of the 27th International Wittgenstein Symposium*, Vienna: ÖBV and HPT, 2005 pp. 387–94.

[12] Greenspan, P., *Practical reasons and moral 'ought'*, Oxford Studies in Metaethcis **2** (2007), pp. 181–205.

[13] Hansson, S. O., *Representing supererogation*, Journal of Logic and Computation **25** (2015), pp. 443–451.

[14] Horton, J., *The all or nothing problem*, Journal of Philosophy **114** (2017), pp. 94–104.

[15] Horty, J., "The Logic of Precedent: Constraint and Freedom in Common Law Reasoning," Cambridge University Press, forthcoming.

[16] Horty, J. F., *Moral dilemmas and nonmonotonic logic*, Journal of Philosophical Logic **23** (1994), pp. 35–65.

[17] Horty, J. F., *Reasoning with moral conflicts*, Noûs **37** (2003), pp. 557–605.

[18] Horty, J. F., "Reasons as Defaults," New York: Oxford University Press, 2012.

[19] Kamm, F., *Supererogation and obligation*, Journal of Philosophy **82** (1985), pp. 118–38.

[20] Knoks, A., *Misleading higher-order evidence, conflicting ideals, and defeasible logic*, Ergo **8** (2021), pp. 141–74.

[21] Kratzer, A., "Modals and Conditionals: New and Revised Perspectives," Oxford: Oxford University Press, 2012.

[22] Lewis, D., "Counterfactuals," Blackwell, 1973.

[23] McNamara, P., *Doing well enough: Toward a logic for common-sense morality*, Studia Logica **57** (1996), pp. 167–92.

[24] McNamara, P., *Must I do what I ought (or will the least I can do do)?*, in: M. Brown and J. Carmo, editors, *Deontic Logic, Agency and Normative Systems*, Springer-Verlag, 1996 pp. 154–73.

[25] McNamara, P., *Praise, blame, obligation, and DWE: Toward a framework for classical supererogation and kin*, Journal of Applied Logic **9** (2011), pp. 153–70.

[26] McNamara, P., *Logics for supererogation and allied normative concepts*, in: D. Gabbay, J. Horty, X. Parent, R. van der Meyden and L. van der Torre, editors, *Handbook of Deontic Logic and Normative Systems*, College Publications, 2022 pp. 155–306.

[27] McNamara, P., *A natural conditionalization of the dwe framework*, in: *Agency, Norms, Inquiry, and Artifacts: Essays in Honor of Risto Hilpinen*, Springer, 2022 pp. 113–136.

[28] Mullins, R., *Moral conflict and the logic of rights*, Philosophical Studies **177** (2020), pp. 633–51.

[29] Mullins, R., *Formalizing reasons, oughts, and requirements*, Ergo **7** (2021), pp. 568–99.

[30] Muñoz, D., *Three paradoxes of supererogation*, Noûs **55** (2021), pp. 669–716.

[31] Muñoz, D. and T. Pummer, *Supererogation and conditional obligation*, Philosophical Sstudies **179** (2022), pp. 1429–43.

[32] Nair, S., *Conflicting reasons, unconflicting 'oughts'*, Philosophical Studies **173** (2016), pp. 629–63.

[33] Parent, X. and L. van der Torre, "Introduction to Deontic Logic and Normative Systems," College Publications, 2018.

[34] Parfit, D., "On What Matters," Oxford University Press, 2011.

[35] Portmore, D., "Commonsense Consequentialism: Wherein Morality Meets Rationality," Oxford University Press, 2011.

[36] Portner, P., "Modality," Oxford University Press, 2009.

[37] Raz, J., "Practical reason and norms," Oxford University Press, 1990.

[38] Scanlon, T. M., "What We Owe to Each Other," Cambridge, MA: Harvard University Press, 1998.

[39] Scanlon, T. M., "Being Realistic about Reasons," Oxford University Press, 2014.

[40] Schroeder, M., "Slaves of the Passions," New York: Oxford University Press, 2007.

[41] Schroeder, M., "Reasons First," Oxford University Press, 2021.

[42] Silk, A., *What normative terms mean and why it matters for ethical theory*, Oxford Studies in Normative Ethics **25** (2015), pp. 296–325.

[43] Snedegar, J., *Reasons, oughts, and requirements*, Oxford Studies in Metaethcis **11** (2016), pp. 155–81.

[44] Star, D., *Introduction*, in: D. Star, editor, *The Oxford Handbook of Reasons and Normativity*, Oxford University Press, 2018 pp. 1–21.

[45] van Fraassen, B., *Values and the heart's command*, The Journal of Philosophy **70** (1973), pp. 5–19.

[46] von Fintel, K. and S. Iatridou, *Time and modality*, in: J. Guéron and J. Lecarme, editors, *How to say ought in foreign: The composition of weak necessity modals*, Springer-Verlag, 2008 pp. 115–41.

[47] Wessels, U., *Beyond the call of duty: The structure of a moral region*, Royal Institute of Philosophy Supplement: Supererogation **77** (2015), pp. 87–104.