Supererogation, Dual-Role Views, and the Logic of Reasons

Maryland Logic Seminar

Aleks Knoks
aleks.knoks@uni.lu
(joint work with David Streit)

April 3, 2023
Supererogation is tricky. It comes with at least three puzzles or paradoxes. My main goal is to solve them.

My solutions will draw on two resources:

(1) the core insight of the dual-role views of reasons; and

(2) a simple defeasible logic-inspired model of the way reasons interact to support oughts. (This is the “logic of reasons”.)
- Preliminaries and a baseline model
- Supererogation and its puzzles
- A formal dual-role account
- Resolving the puzzles
- Conclusion
Normative reason = a consideration that counts in favor of (or against) an action

Schroeder’s (2021) “three marks”: competition, act-orientation, and acting for

Standard view: reasons are to play a crucial role in our understanding of the structure of normativity and answering various normative questions.

Background idea, going back to W. D. Ross (1930): there’s a close tie between the structure of human deliberation and the structure of what makes it the case that you ought to do one thing rather than another

So how do oughts relate to reasons? Standard response: the “scales model” (or, perhaps, scales with conditions and modifiers).
I’m going to give you a better alternative.

- **Background:** PL, materially inconsistent formulas, ⊬

- **!(X) = “there is a reason to do X”**
  
  \[ \text{Conclusion}(!(X)) = X \quad \text{cf. van Fraassen (1973), Nair (2016)} \]

- **A reason theory** \( \Delta \) is a pair of the form \( \langle R, < \rangle \), where
  
  - \( R \) is a set of !-formulas (roughly, a set of reasons)
  - \( < \) is a preorder on \( R \)

- **Example (Drowning Child):** \( \Delta_1 = \langle R, < \rangle \) where
  
  - \( R = \{ !K, !R \} \)
  - \( !K < !R \)
  - \( K = \text{Keep the promise} \)
  - \( R = \text{Rescue the child} \)
  - \( R \vdash \neg K \)

- **Undefeated_{\langle R, < \rangle} = \{ r \in R : \text{no } r' \in R \text{ with } r < r' \text{ and } \text{Conclusion}(r') \vdash \neg \text{Conclusion}(r) \} \)

  \( \text{Undefeated}_{\Delta_1} = \{ !R \} \)
Let \( \Delta = \langle \mathcal{R}, < \rangle \) be a reason theory. Then \( \text{Ought}(X) \) follows from it, \( \Delta \mid \sim \text{Ought}(X) \), just in case \( M \vdash X \) for every \( M \), where

1. \( M \subseteq \text{Conclusion}(\text{Undefeated}_{\langle \mathcal{R}, < \rangle}) \), and
2. \( M \) is maximally consistent.

Notice that \( \text{Ought}(R) \) follows from \( \Delta_1 \) (Drowning Child).

**Example (Save One):** \( \Delta_2 = \langle \mathcal{R}, < \rangle \) where

\[
\mathcal{R} = \{!A, !B\} \\
< \text{ is empty } \\
A = \text{Save Alice} \\
B = \text{Save Bob} \\
A \vdash \neg B
\]

We have \( \text{Undefeated}_{\Delta_2} = \{!A, !B\} \); \( \text{Ought}(A) \) and \( \text{Ought}(B) \) do not, while \( \text{Ought}(A \lor B) \) does follow from \( \Delta_2 \).

Consequence relations that appeal to maxiconsistency have a venerable history—see, e.g., (Makinson, 2005).

I’ll be interested in permissions, \( \text{Can}(X) \), as well as requirements, \( \text{Must}(X) \), but I will introduce them later..
My mantra: keep the formalism as simple as possible

The model I’ve set up is simple, but one can extend it to Hory’s (2012) default logic-based model which is much more expressive..
Supererogation and its “classic paradox”

Save One or Two. Alice and Bob are in mortal danger. You can save one without cost. You can also save both, but you would lose your legs.

- **you ought to / should save Alice and Bob**
- **you must / have to save either Alice or Bob**

Shall we set $\neg L < !A, !B$? $!A, !B < !\neg L$? Or do yet something else?

Informal reason-first accounts fare no better. They face a dilemma:
- if most reason to save one, then saving two isn’t really better..
- if most reason to save both, then losing legs is obligatory..

Upshot: these views lack the resources to (i) account for supererogation and (ii) capture the distinction between **oughts** and **requirements**


(More problems in the vicinity: (1) Raz’s “basic belief”; (2) tragic dilemmas vs. optionality—cf. Mullins (2021))
Can’t the classical semantics handle this?

A natural thought: can’t classical semantics for deontic modals easily deal with this?

McNamara’s (1996) Doing Well Enough framework

\[
\begin{array}{c|c|c}
A, B, L & A, \neg B, \neg L & \neg A, B, \neg L \\
\hline
\neg A, \neg B, \neg L & & \\
\end{array}
\]

All or Nothing. Alice and Bob are trapped in a burning building. You can save one or both of them, either will result in you losing your legs. (Horton, 2017)

\[
\begin{array}{c|c|c}
A, B, L & A, \neg B, L & \neg A, B, L \\
\hline
\neg A, \neg B, \neg L & & \\
\end{array}
\]

\(A = \text{Save Alice}\)
\(B = \text{Save Bob}\)
\(L = \text{Lose legs}\)
Horton’s All or Nothing Problem

Horton (2017): the case actually exhibits a paradox

The All or Nothing Problem:

1. It’s morally permissible to save neither Alice, nor Bob. (intuition)
2. It is morally wrong for you to save only one of them. (intuition)
3. If an act \( X \) is morally permissible and \( Y \) 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 one of them. (from 1–3)
4∗ If you are not going to save Alice and Bob, you ought to save neither. (= 4)
5. But, surely, (4)/(4∗) is false. (intuition)

Next question: How can we restate this in formal notation?
A = Save Alice; B = Save Bob; L = Lose legs; A ⊨ L; B ⊨ L

1. Can(¬L) (intuition)
2. Must(¬([A&¬B] ∨ [¬A&B]))) (intuition)
   If X is morally wrong, then Must(¬X). (assumption)
3. If Can(X) AND Must(¬Y), then Must(X|X ∨ Y) (principle)
4. If Can(¬L) AND Must(¬([A&¬B] ∨ [¬A&B])), then
   Must(¬L|¬L ∨ ([A&¬B] ∨ [¬A&B])) (instance)
5. Must(¬L|¬L ∨ ([A&¬B] ∨ [¬A&B])) (from 1, 2, and 4)
6. Must(¬L|¬[A&B]) (substitution of equivalent formulas)
7. But, clearly, ¬Must(¬L|¬[A&B]) (intuition)

Heads up: on my account, (3) is false.
Life or Promise. Macabre version of the Drowning Child. (Kamm, 1985)

1. If act $X$ is permissible in the presence of act $Y$ AND act $Y$ is permissible in the presence of act $Z$, then act $X$ is permissible in the presence of act $Z$. (intuitive principle)

[Kamm talks about the “may take precedence over” relation.]

2. *Walk Away* is permissible in the presence of *Save Alice* (and lose legs). (intuition)

3. *Save Alice* is permissible in the presence of *Keep Promise*. (intuition)

4. *Walk Away* is permissible in the presence of *Keep Promise*. (from 1–3)

5. But, clearly, *Walk Away* is not permissible in the presence of *Keep Promise*. (intuition)
Kamm’s Intransitivity Paradox

\[
A = \text{Save Alice;} \\
K = \text{Keep Promise;} \\
L = \text{Lose legs;} \\
A \vdash L; A \vdash \neg K
\]

1. If \( \text{Can}(X|X \lor Y) \) and \( \text{Can}(Y|Y \lor Z) \), then \( \text{Can}(X|X \lor Z) \) (transitivity principle)

2. \( \text{Can}([\neg K \& \neg L]|[\neg K \& \neg L] \lor A) \) (intuition)

3. \( \text{Can}(A|A \lor K) \) (intuition)

4. \( \text{Can}([\neg K \& \neg L]|[\neg K \& \neg L] \lor K) \) (from 1–3)

5. But, clearly, \( \neg \text{Can}([\neg K \& \neg L]|[\neg K \& \neg L] \lor K) \) (intuition)

Kamm: (1) must be false. My account supports this verdict.
Dual-role views, core insight

Dual-role views, core idea: there are **two completely different types of considerations** / **two roles** reasons can play:

- **Requiring reasons** tend to make actions required;
- **Justifying reasons** tend to make actions permissible.

See Gert (2004, 2007), Greenspan (2005), Muñoz (2021), Tucker (2022), and many others.

The accounts differ in detail, which we don’t need to care about. Also, they are **motivated on independent grounds**.

Now I’m going to modify the baseline model to capture this idea..
Capturing the core idea in the model

Step #1: Split the single set of reasons $\mathcal{R}$ into two: $\mathcal{R}$ and $\mathcal{I}$
- Henceforth, we work with dual-role reason theories, or structures of the form $\langle \mathcal{R}, \mathcal{I}, < \rangle$, with the constraint that $\mathcal{R} \subseteq \mathcal{I}$

Step #2: Define the notions of stable scenarios and restricted theories:
- $\mathcal{D}$ is stable just in case $\mathcal{R} \subseteq \mathcal{D} \subseteq \mathcal{I}$
- from $\Delta = \langle \mathcal{R}, \mathcal{I}, < \rangle$ to $\Delta^\mathcal{D} = \langle \mathcal{R}, \mathcal{D}, <^\mathcal{D} \rangle$

Step #3: Substitute the notion $\text{Undefeated}_{\langle \mathcal{R}, \mathcal{I}, < \rangle}$ for $\text{Undefeated}_{\langle \mathcal{R}, < \rangle}$

Step #4: Let the oughts follow from all undefeated requiring reasons:
- Given $\Delta = \langle \mathcal{R}, \mathcal{I}, < \rangle$, we have $\Delta \models \text{Ought}(X)$ just in case $\mathcal{M} \models X$ for every $\mathcal{M} \subseteq \text{Conclusion}(\text{Undefeated}_{\Delta^\mathcal{R}})$ that is maximally consistent.
- Notice that $\Delta^\mathcal{R} = \langle \mathcal{R}, \mathcal{R}, <^\mathcal{R} \rangle$
Capturing the core idea in the model

Step #5: Let the requirements follow from all undefeated reasons:

▶ Given $\Delta = \langle R, J, < \rangle$, we have $\Delta \vdash \text{Must}(X)$ just in case, for every stable scenario $D$:
\[ \mathcal{M} \vdash X \text{ for every } \mathcal{M} \subseteq \text{Conclusion}(\text{Undefeated}_{\Delta_D}) \] that is maximally consistent.

Step #6: Let the permissions follow from some undefeated reasons:

▶ Given $\Delta = \langle R, J, < \rangle$, we have $\Delta \vdash \text{Can}(X)$ just in case, for some stable scenario $D$:
\[ \mathcal{M} \vdash X \text{ for some } \mathcal{M} \subseteq \text{Conclusion}(\text{Undefeated}_{\Delta_D}) \] that is maximally consistent.

(Notice that maximal consistent subsets are just a device that lets us derive consequences.)
Save One or Two:

Consider $\Delta_3 = \langle \mathcal{R}, \mathcal{J}, < \rangle$ where

$\mathcal{R} = \{!A, !B\}$

$\mathcal{J} = \{!A, !B, !\neg L\}$

$< \text{ is empty}$

$A = \text{Save Alice,}$

$B = \text{Save Bob,}$

$L = \text{Lose legs,}$

$A \& B \vdash L$

$\textbf{Undefeated}_{\Delta_3} = \{!A, !B\}$

Hence, $\Delta_3 \mid \sim \text{Ought}(A \land B)$

$\textbf{Two stable scenarios: } \{!A, !B\} \text{ and } \{!A, !B, !\neg L\}$

Three maxiconsistent sets (of undefeated reasons):

$\{A, B\}, \{A, \neg L\}, \text{ and } \{B, \neg L\}$

Hence, $\Delta_3 \mid \sim \text{Must}(A \lor B), \text{Can}(\neg L), \text{etc.}$
Some cases

All or Nothing:

Consider $\Delta_4 = \langle \mathcal{R}, \mathcal{J}, < \rangle$ where

$\mathcal{R} = \{!A, !B\}$

$\mathcal{J} = \{!A, !B, !\neg L\}$

$< \text{ is empty}$

$A = \text{Save Alice},$

$B = \text{Save Bob},$

$L = \text{Lose legs},$

$A \vdash L$

$B \vdash L$

- **Undefeated** $\Delta_4^\mathcal{R} = \{!A, !B\}$

Hence, $\Delta_4 \not\models \text{Ought}(A \land B)$

- Two stable scenarios: $\{!A, !B\}, \{!A, !B, !\neg L\}$

Two maxiconsistent sets (of undefeated reasons): $\{A, B\}, \{\neg L\}$

Hence, we don’t get either $\Delta_4 \not\models \text{Can}(A \land \neg B)$, or

$\Delta_4 \not\models \text{Can}(\neg A \land B)$

- Ideally, we’d also have $\Delta_4 \not\models \text{Must} \neg[(A \land \neg B) \lor (\neg A \land B)]$

(A requiring reason speaking against letting people die gratuitously should do the trick, but I haven’t found a natural way to express it..)
Kamm’s case:

Consider $\Delta_5 = \langle \mathcal{R}, \mathcal{J}, < \rangle$ where

$\mathcal{R} = \{ !A, !K \}$

$\mathcal{J} = \{ !A, !K, !\neg L \}$

$!K < !A$

$\blacktriangleright$ Undefeated$_{\Delta_5}^{\mathcal{R}} = \{ !A \}$

Hence, $\Delta_5 \models \text{Ought}(A)$

$\blacktriangleright$ Two stable scenarios: $\{ !A, !K \}$ and $\{ !A, !K, !\neg L \}$

Two maxiconsistent sets (of undefeated reasons):

$\{ A \}, \{ K, \neg L \}$

Hence, $\Delta_5 \models \text{Must}(A \lor K)$

(Actually, I’m cheating, but these are the “Brewka scenarios”)

$A = \text{Save Alice,}$

$K = \text{Keep promise,}$

$L = \text{Lose legs,}$

$A \vdash \neg K$

$A \vdash L$
Some facts: For any dual-role theory $\Delta = \langle R, J, < \rangle$:

1. if $\Delta \searrow Must(X)$, then $\Delta \searrow Ought(X)$; and
   if $\Delta \searrow Ought(X)$, then $\Delta \searrow Can(X)$;

2. neither $\Delta \searrow Ought(\bot)$; nor $\Delta \searrow Must(X)$; nor $\Delta \searrow Can(X)$;

3. $\Delta \searrow Ought(X \& Y)$ iff $\Delta \searrow Ought(X)$ and $\Delta \searrow Ought(Y)$;
   $\Delta \searrow Must(X \& Y)$ iff $\Delta \searrow Must(X)$ and $\Delta \searrow Must(Y)$.

Response to the “Classic Paradox”:
Supererogatory action is superior because it maximizes requiring reasons;
it’s not obligatory because there can be counterbalancing justifying reasons.

For the remaining puzzles, we need conditional normative statements:

To see if $Ought(Y|X)$ follows from $\Delta = \langle R, J, < \rangle$, determine whether
$Ough(Y)$ follows from $\Delta[X]$, or $\Delta$ on the assumption that $\vdash X$. 

Aleks Knoks
Recall the principle that gave rise to the problem:
If \( \text{Can}(X) \) and \( \text{Must}(\neg Y) \), then \( \text{Must}(X \mid X \lor Y) \)

The corresponding principle
If \( \Delta \models \sim \text{Can}(X) \) and \( \Delta \models \sim \text{Must}(\neg Y) \), then \( \Delta \models \sim \text{Must}(X \mid X \lor Y) \)
is demonstrably false.

But isn’t this ad hoc? Why does the principle seem so intuitive?

**Answer:** There are two weaker principles in the vicinity that are true.

For any dual-role theory \( \Delta = \langle \mathcal{R}, \mathcal{J}, \prec \rangle \):

1. if \( \Delta \models \sim \text{Ought}(X) \) and \( \Delta \models \sim \text{Must}(\neg Y) \), then \( \Delta \models \sim \text{Ought}(X \mid X \lor Y) \);
2. if \( \Delta \models \sim \text{Can}(X) \) and \( \Delta \models \sim \text{Must}(\neg Y) \), then \( \Delta \models \sim \text{Can}(X \mid X \lor Y) \).
Recall the Transitivity Principle:
If \( \text{Can}(X|X \lor Y) \) and \( \text{Can}(Y|Y \lor Z) \), then \( \text{Can}(X|X \lor Z) \)

\( \Delta_4 \) (Kamm’s case) shows that the corresponding principle
If \( \Delta \not\vdash \text{Can}(X|X \lor Y) \) & \( \Delta \not\vdash \text{Can}(Y|Y \lor Z) \), then \( \Delta \not\vdash \text{Can}(X|X \lor Z) \) is false.

Why does the Transitivity Principle seem so intuitive then?

**Answer:** (1) the assumption that reasons function in one way only; and (2) the fact that transitivity holds for oughts:

For any dual-role theory \( \Delta = \langle R, J, < \rangle \):

If \( \Delta \not\vdash Ought(X|X \lor Y) \) and \( \Delta \not\vdash Ought(Y|Y \lor Z) \), then
\( \Delta \not\vdash Ought(X|X \lor Z) \).

(I’m convinced that transitivity holds of purely justifying reasons. Unfortunately, it’s difficult to state a nice principle expressing it.)
The loose end from the APA paper is not a loose end. The overly simplistic notion of defeat was causing all the problems.

Rob Mullin’s (2021) has proposed a formal account that looks similar to ours:

- $\Delta = \langle R, J, < \rangle \models Must(X)$ just in case $M \vdash X$ for every maxiconsistent $M \subseteq Conclusion(Undefeated_{\Delta R})$
- $\Delta = \langle R, J, < \rangle \models Ought(X)$ just in case $M \vdash X$ for some “maximally requirement consistent” subset $M$ of $Conclusion(Undefeated_{\Delta J})$

This account falls flat when applied to supererogation..

A lot of informal literature seems to pursue a similar strategy. (It’s probably where Mullin’s got the idea from..)
Conclusion

My goal was to account for supererogation and solve the puzzles associated with it.

To reach the goal, I set up a formal dual-role account of reasons.

This account:

1. explains supererogation (= resolves the first puzzle);
2. captures the intuitive relation between *musts*, *oughts*, and *permissions*;
3. supports appealing responses to Horton’s All or Nothing Problem and Kamm’s Intranisitivity Paradox;
4. fares much better than competing formal accounts;
5. is sufficiently different from informal accounts of supererogation.
6. (Bonus: it (1) can account for Raz’s “basic belief” and (2) distinguish dilemmas from cases involving abundant choice)

All in all, we have a(nother) killer application of logics for defeasible in ethics.
Thanks!


W. D. Ross (1930), *The Right and the Good*, Oxford University Press.


