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

Maryland Logic Seminar

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

UNIVERSITÉ DU LUXEMBOURG

Luxembourg National Research Fund

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, $\vdash$
- $!X$ = "there is a reason to do X"
  $Conclusion(!X) = X$ cf. van Fraassen (1973), Nair (2016)
- A *reason theory* Δ is a pair of the form $\langle \mathcal{R}, < \rangle$, where
  $\mathcal{R}$ is a set of !-formulas (roughly, a set of reasons)
  $<$ is a preorder on $\mathcal{R}$
- Example (Drowning Child): $\Delta_1 = \langle \mathcal{R}, < \rangle$ where
  K = Keep the promise
  $\mathcal{R} = \{!K, !R\}$
  R = Rescue the child
  $!K < !R$
  $R \vdash \neg K$
- $Undefeated_{\langle \mathcal{R}, < \rangle} = \{r \in \mathcal{R} :$ no $r' \in \mathcal{R}$ with $r < r'$ and $Conclusion(r') \vdash \neg Conclusion(r)\}$

  $Undefeated_{\Delta_1} = \{!R\}$ ☹

- Let $\Delta = \langle \mathcal{R}, < \rangle$ be a reason theory. Then *Ought*(X) follows from it, $\Delta \mathrel{|\!\sim} \textit{Ought}(X)$, just in case $\mathcal{M} \vdash X$ for every $\mathcal{M}$, where
  (i) $\mathcal{M} \subseteq \textit{Conclusion}(\textit{Undefeated}_{\langle \mathcal{R}, < \rangle})$, and
  (ii) $\mathcal{M}$ is maximally consistent.

  Notice that *Ought*(R) follows from Δ_1 (Drowning Child)

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

  $\mathcal{R} = \{!A, !B\}$
  $<$ is empty

  A = Save Alice
  B = Save Bob
  $A \vdash \neg B$

- We have $\textit{Undefeated}_{\Delta_2} = \{!A, !B\}$; *Ought*$(A)$ and *Ought*(B) do not, while *Ought*$(A \vee B)$ does follow from Δ_2

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

- I'll be interested in permissions, *Can*(X), as well as requirements, *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 Horty's (2012) default logic-based model which is much more expressive..

$$
\begin{tikzcd}
K \arrow[rr, Leftrightarrow, "|" marking] & & R \\
P \arrow[u, "\delta_1"] & & N \arrow[u, "\delta_2"'] \\
& \top \arrow[ul, Rightarrow] \arrow[ur, Rightarrow] &
\end{tikzcd}
\qquad \delta_1 < \delta_2
$$

$$
\begin{tikzcd}
C \\
B \arrow[u, "\delta'"] \\
A \arrow[u, "\delta"] \\
\top \arrow[u, Rightarrow]
\end{tikzcd}
$$

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..

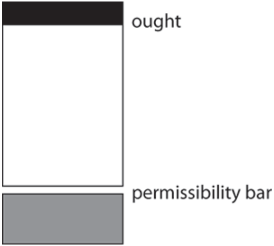
Upshot: these views lack the resources to (i) account for supererogation and (ii) capture the distinction between oughts and requirements

cf. Bedke (2011), Snedegar (2016)

(More problems in the vicinity: (1) Raz's "basic belief"; (2) tragic dilemmas vs. optionality—cf. Mullins (2021))

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

McNamara's (1996) Doing Well Enough framework



A, B, L	
$A, \neg B, \neg L$	$\neg A, B, \neg L$
$\neg A, \neg B, \neg L$	

A = Save Alice
B = Save Bob
L = Lose legs

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)

A, B, L	
$A, \neg B, L$	$\neg A, B, L$
$\neg A, \neg B, \neg L$	

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; $A \vdash L$; $B \vdash L$

1. $Can(\neg L)$ (intuition)
2. $Must(\neg([A\&\neg B] \vee [\neg A\&B]))$ (intuition)

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

A = Save Alice;
K = Keep Promise;
L = Lose legs;
$A \vdash L$; $A \vdash \neg K$

$A, \neg K, L$
$\neg A, K, \neg L$
$\neg A, \neg K, \neg L$

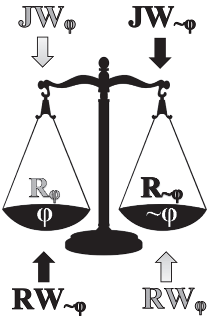
1. If $Can(X|X \vee Y)$ and $Can(Y|Y \vee Z)$, then $Can(X|X \vee Z)$ (transitivity principle)
2. $Can([\neg K \& \neg L]|[\neg K \& \neg L] \vee A)$ (intuition)
3. $Can(A|A \vee K)$ (intuition)
4. $Can([\neg K \& \neg L]|[\neg K \& \neg L] \vee K)$ (from 1–3)
5. But, clearly, $\neg Can([\neg K \& \neg L]|[\neg K \& \neg L] \vee K)$ (intuition)

Kamm: (1) must be false. My account supports this verdict.

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..

Step #1: Split the single set of reasons $\mathcal{R}$ into two: $\mathcal{R}$ and $\mathcal{J}$

- Henceforth, we work with dual-role reason theories, or structures of the form $\langle \mathcal{R}, \mathcal{J}, < \rangle$, with the *constraint* that $\mathcal{R} \subseteq \mathcal{J}$

Step #2: Define the notions of stable scenarios and restricted theories:

- $\mathcal{D}$ is stable just in case $\mathcal{R} \subseteq \mathcal{D} \subseteq \mathcal{J}$
- from $\Delta = \langle \mathcal{R}, \mathcal{J}, < \rangle$ to $\Delta^{\mathcal{D}} = \langle \mathcal{R}, \mathcal{D}, <^{\mathcal{D}} \rangle$

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

Step #4: Let the oughts follow from *all* undefeated requiring reasons:

- Given $\Delta = \langle \mathcal{R}, \mathcal{J}, < \rangle$, we have $\Delta \mathrel{|\!\sim} \mathit{Ought}(X)$ just in case $\mathcal{M} \vdash X$ for every $\mathcal{M} \subseteq \mathit{Conclusion}(\mathit{Undefeated}_{\Delta^{\mathcal{R}}})$ that is maximally consistent.
- Notice that $\Delta^{\mathcal{R}} = \langle \mathcal{R}, \mathcal{R}, <^{\mathcal{R}} \rangle$

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

- Given $\Delta = \langle \mathcal{R}, \mathcal{J}, < \rangle$, we have $\Delta \mid\!\sim \mathit{Must}(X)$ just in case, for every stable scenario $\mathcal{D}$: $\mathcal{M} \vdash X$ for every $\mathcal{M} \subseteq \mathit{Conclusion}(\mathit{Undefeated}_{\Delta^{\mathcal{D}}})$ that is maximally consistent.

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

- Given $\Delta = \langle \mathcal{R}, \mathcal{J}, < \rangle$, we have $\Delta \mid\!\sim \mathit{Can}(X)$ just in case, for some stable scenario $\mathcal{D}$: $\mathcal{M} \vdash X$ for some $\mathcal{M} \subseteq \mathit{Conclusion}(\mathit{Undefeated}_{\Delta^{\mathcal{D}}})$ 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
$\mathcal{R} = \{!A, !B\}$
$\mathcal{J} = \{!A, !B, !\neg L\}$
$<$ is empty

$A =$ Save Alice,
$B =$ Save Bob,
$L =$ Lose legs,
$A \& B \vdash L$

- $Undefeated_{\Delta_3^{\mathcal{R}}} = \{!A, !B\}$
  Hence, $\Delta_3 \mathrel{|\!\!\sim} Ought(A \wedge B)$
- Two stable scenarios: $\{!A, !B\}$ and $\{!A, !B, !\neg L\}$
  Three maxiconsistent sets (of undefeated reasons):
  $\{A, B\}$, $\{A, \neg L\}$, and $\{B, \neg L\}$
  Hence, $\Delta_3 \mathrel{|\!\!\sim} Must(A \vee B)$, $Can(\neg L)$, etc.

All or Nothing:

Consider $\Delta_4 = \langle \mathcal{R}, \mathcal{J}, < \rangle$ where
$\mathcal{R} = \{!A, !B\}$
$\mathcal{J} = \{!A, !B, !\neg L\}$
$<$ is empty

A = Save Alice,
B = Save Bob,
L = Lose legs,
$A \vdash L$
$B \vdash L$

- $\mathit{Undefeated}_{\Delta_4^{\mathcal{R}}} = \{!A, !B\}$
  Hence, $\Delta_4 \mid\!\sim \mathit{Ought}(A \wedge B)$
- 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 $\Delta_4 \mid\!\sim \mathit{Can}(A \wedge \neg B)$, or $\Delta_4 \mid\!\sim \mathit{Can}(\neg A \wedge B)$
- Ideally, we'd also have $\Delta_4 \mid\!\sim \mathit{Must}\neg[(A \wedge \neg B) \vee (\neg A \wedge B)]$
  (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
$\mathcal{R} = \{!A, !K\}$
$\mathcal{J} = \{!A, !K, !\neg L\}$
$!K < !A$

A = Save Alice,
K = Keep promise,
L = Lose legs,
$A \vdash \neg K$
$A \vdash L$

- $\mathit{Undefeated}_{\Delta_5^{\mathcal{R}}} = \{!A\}$
  Hence, $\Delta_5 \mathrel{|\!\sim} \mathit{Ought}(A)$
- Two stable scenarios: $\{!A, !K\}$ and $\{!A, !K, !\neg L\}$
  Two maxiconsistent sets (of undefeated reasons): ☹
  $\{A\}$, $\{K, \neg L\}$
  Hence, $\Delta_5 \mathrel{|\!\sim} \mathit{Must}(A \vee K)$

  (Actually, I'm cheating, but these are the "Brewka scenarios") ☺

Some facts: For any dual-role theory $\Delta = \langle \mathcal{R}, \mathcal{J}, < \rangle$:

(1) if $\Delta \mathrel{|\!\sim} \mathit{Must}(X)$, then $\Delta \mathrel{|\!\sim} \mathit{Ought}(X)$; and
if $\Delta \mathrel{|\!\sim} \mathit{Ought}(X)$, then $\Delta \mathrel{|\!\sim} \mathit{Can}(X)$;

(2) neither $\Delta \mathrel{|\!\sim} \mathit{Ought}(\bot)$; nor $\Delta \mathrel{|\!\sim} \mathit{Must}(X)$; nor $\Delta \mathrel{|\!\sim} \mathit{Can}(X)$;

(3) $\Delta \mathrel{|\!\sim} \mathit{Ought}(X \& Y)$ iff $\Delta \mathrel{|\!\sim} \mathit{Ought}(X)$ and $\Delta \mathrel{|\!\sim} \mathit{Ought}(Y)$;
$\Delta \mathrel{|\!\sim} \mathit{Must}(X \& Y)$ iff $\Delta \mathrel{|\!\sim} \mathit{Must}(X)$ and $\Delta \mathrel{|\!\sim} \mathit{Must}(Y)$.

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 $\mathit{Ought}(Y|X)$ follows from $\Delta = \langle \mathcal{R}, \mathcal{J}, < \rangle$, determine whether $\mathit{Ough}(Y)$ follows from $\Delta[X]$, or Δ on the assumption that $\vdash X$.

Recall the principle that gave rise to the problem:
If $\mathit{Can}(X)$ and $\mathit{Must}(\neg Y)$, then $\mathit{Must}(X|X \vee Y)$

The corresponding principle
If $\Delta \mathrel{|\!\sim} \mathit{Can}(X)$ and $\Delta \mathrel{|\!\sim} \mathit{Must}(\neg Y)$, then $\Delta \mathrel{|\!\sim} \mathit{Must}(X|X \vee Y)$
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 $\Delta = \langle \mathcal{R}, \mathcal{J}, < \rangle$:

1. if $\Delta \mathrel{|\!\sim} \mathit{Ought}(X)$ and $\Delta \mathrel{|\!\sim} \mathit{Must}(\neg Y)$, then $\Delta \mathrel{|\!\sim} \mathit{Ought}(X|X \vee Y)$;
2. if $\Delta \mathrel{|\!\sim} \mathit{Can}(X)$ and $\Delta \mathrel{|\!\sim} \mathit{Must}(\neg Y)$, then $\Delta \mathrel{|\!\sim} \mathit{Can}(X|X \vee Y)$.

Recall the Transitivity Principle:
If $Can(X|X \vee Y)$ and $Can(Y|Y \vee Z)$, then $Can(X|X \vee Z)$

Δ_4 (Kamm's case) shows that the corresponding principle
If $\Delta \mathrel{|\!\sim} Can(X|X \vee Y)$ & $\Delta \mathrel{|\!\sim} Can(Y|Y \vee Z)$, then $\Delta \mathrel{|\!\sim} Can(X|X \vee Z)$ 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 $\Delta = \langle \mathcal{R}, \mathcal{J}, < \rangle$:

> If $\Delta \mathrel{|\!\sim} Ought(X|X \vee Y)$ and $\Delta \mathrel{|\!\sim} Ought(Y|Y \vee Z)$, then $\Delta \mathrel{|\!\sim} Ought(X|X \vee Z)$.

(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:
  - $\Delta = \langle \mathcal{R}, \mathcal{J}, < \rangle \mid\!\sim \mathit{Must}(X)$ just in case $\mathcal{M} \vdash X$ for every maxiconsistent $\mathcal{M} \subseteq \mathit{Conclusion}(\mathit{Undefeated}_{\Delta^{\mathcal{R}}})$
  - $\Delta = \langle \mathcal{R}, \mathcal{J}, < \rangle \mid\!\sim \mathit{Ought}(X)$ just in case $\mathcal{M} \vdash X$ for some "maximally requirement consistent" subset $\mathcal{M}$ of $\mathit{Conclusion}(\mathit{Undefeated}_{\Delta^{\mathcal{J}}})$

  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..)

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!

M. Bedke (2011), “Passing the deontic buck”, In: *Oxford Studies in Metaethics*, 6.

J. Gert (2004), *Brute Rationality: Normativity and Human Action*, Oxford University Press.

J. Gert (2007), “Normative strength and the balance of reasons”, In: *Pilosophical Review*, 116(4).

P. Greenspan (2007), “Asymmetrical Practical Reasons”, In: Reicher M. & Marek J. (eds.) *Experience and Analysis*.

J. Horton (2017), “The all or nothing problem”, In: *Journal of Philosophy*, 114.

J. Horty, *Reasons as Defaults*, Oxford University Press.

F. Kamm (1985), “Supererogation and obligation”, In: *Journal of Philosophy*, 82.

P. McNamara (1996), “Doing well enough: Toward a logic for common-sense morality”, In: *Studia Logica*, 57.

R. Mullins (2021), “Formalizing reasons, oughts, and requirements”, In: *Ergo*, 7.

D. Munoz (2021), “Three paradoxes of supererogation”, In: *Nous*, 55.

R. Reiter (1980), “A logic for default reasoning”, In: *Artificial Intelligence*, 13.

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

J. Snedegar (2016), “Reasons, oughts, and requirements”, In: *Oxford Studies in Metaethics*, 11.

M. Schroeder (2021), *Reasons First*, Oxford University Press.

C. Tucker (2022), “Weighing reasons against”, In: *Oxford Studies in Metaethics*, 17.

B. van Fraassen (1973), “Value and the heart’s command”, In: *The Journal of Philosophy*, 70.