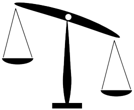
Reason-based detachment

LNGAI 2023, Hangzhou

Aleks Knoks
aleks.knoks@uni.lu
(joint work with Leendert van der Torre)

September 8, 2023

- Normative reasons = considerations that speak in favor of or against actions, usually taken to be facts
- Schroeder's ('21) "three marks": competition, act-orientation, and acting for
- An action's deontic status is determined by weighing reasons on weight scales



- Reasons are really important & "reasons-first" (Parfit '11, Raz '90, Scanlon '98, Schroeder '21, a.o.)
- Our approach: think of weighing reasons on scales as an inference pattern—this is the titular *reason-based detachment*.
  Cf. Makinson & van der Torre ('01).
- **Main goal**: set up a general formal framework that puts one in a position to study reason-based detachment from first principles.

- Core formal concepts
- Detachment systems and their properties
- Balancing operations
- Principle-based analysis
- Relations to logic
- Future work

Our approach moves from abstract to more concrete..

Let $\mathcal{A}$ be an infinite set, called the universe of discourse, and $\mathcal{V}$ be a set, called values. Examples:

- $\mathcal{A}$ = propositional atoms; $\mathcal{V} = \{+, 0, -\}$; $\mathcal{V}' = \mathbb{Z}$; $\mathcal{V}'' = \{\mathsf{f}, \mathsf{j}, \mathsf{e}, \mathsf{p}\}$

Core notions:

(1) A **reason** is a triple of the form (x, y, v) where x and y are elements of $\mathcal{A}$ and v is an element of $\mathcal{V}$. $(a, b, +)$, $(a, b, 8)$

(2) A **context** C is a pair of the form (R, y) where R is a finite set of reasons and y, called an *issue*, is an element from the universe of discourse $\mathcal{A}$. $(\{(a, y, v), (b, z, v')\}, y)$

(3) A **detachment system** $\mathcal{D}$ is a two-place relation between contexts and values from $\mathcal{V}$. $((R, y), v) \in \mathcal{D}$

We use the terms properties, principles, and axioms interchangeably..

Some principles for detachment systems:

- `Universal Domain`: for every context C, there is a value v such that $(C, v) \in \mathcal{D}$.
- `Relevance`: only the "y-reasons" matter for which value gets detached from (R, y).
- `Reason Univ. Domain`: if $((R', y), v') \in \mathcal{D}$ and $((R^*, y), v^*) \in \mathcal{D}$, then we have $((R, y), v) \in \mathcal{D}$ for some v, for every $R \subseteq R \cup R^*$.

  (Notice that this is a natural weakening of `Universal Domain`.)
- `Fixed Value`: reasons don't switch values from context to context.
- `Anonymity`: all elements of $\mathcal{A}$ are treated equally.
- `Unanimity`: if *all* "y-reasons" in R have value v, then we have $((R, y), v) \in \mathcal{D}$.

Intuitively, balancing operation are relatively simple detachment systems that are closely related to the informal idea of weight scales.

Formally, let $\mathcal{A}$ = propositional atoms and $\mathcal{V} = \{+, 0, -\}$. Then a detachment system $\mathcal{D}$ qualifies as a balancing operation just in case it is a function from $2^{\mathcal{A}\times\mathcal{A}\times\{+,-\}} \times \mathcal{A}$ to $\mathcal{V}$.

Some principles for balancing operations:

- `Neutrality`: if $+$ gets detached and you switch the polarities of reasons, then the new detached value is $-$. (Ditto for $-$.)
- `Monotony`: additional reasons have no effect on the detached value.
- `Polarity Monotony`: if $+$ gets detached for (R, y), then adding a positive reason to R has no effect. (Ditto for $-$.)
- `Polarity Cut`: if $+$ gets detached for (R, y), then removing a negative reason from R has no effect. (Ditto for $-$.)
- `Polarity Switching`: adding enough reasons of the opposite polarity will force the detached value to switch.

A detachment system $\mathcal{D}$ is an anonymous balancing operation just in case it is a balancing operation that satisfies both `RUd` and `An`.

Some concrete examples:

$$\mathtt{SiCo}(\langle R, y\rangle) = \begin{cases} + & \text{if } |\textit{positive}(R, y)| > |\textit{negative}(R, y)| \\ - & \text{if } |\textit{negative}(R, y)| > |\textit{positive}(R, y)| \\ 0 & \text{otherwise} \end{cases}$$

$$\mathtt{AllNot}(\langle R, y\rangle) = \begin{cases} + & \text{if } \textit{positive}(R, y) = \textit{relevant}(R, y) \neq \emptyset \\ - & \text{if } \textit{negative}(R, y) = \textit{relevant}(R, y) \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

$$\mathtt{DefNeg}(\langle R, y\rangle) = \begin{cases} + & \text{if } |\textit{positive}(R, y)| > |\textit{negative}(R, y)| \\ - & \text{otherwise} \end{cases}$$

There's two more: `Most` and `Threshold`.

Notice that each of the above specifies many balancing operations..

A detachment system $\mathcal{D}$ is a relational balancing operation just in case (roughly) (i) $\mathcal{D}$ satisfies $\texttt{RUd}$ and (ii) $\mathcal{D}$ is equipped with a binary anti-symmetric relation $\prec$ over its corresponding set of reasons $\mathcal{R}$.

Some concrete examples:

$$\forall\exists(\langle R, y\rangle) = \begin{cases} + & \text{if for every negative } r\text{, there is a positive } r' \text{ s.t. } r \prec r' \\ - & \text{if for every positive } r\text{, there is a negative } r' \text{ s.t. } r \prec r' \\ 0 & \text{otherwise} \end{cases}$$

$$\exists\forall(\langle R, y\rangle) = \begin{cases} + & \text{if there is a positive } r' \text{ s.t., for every negative } r\text{, } r \prec r' \\ - & \text{if there is a negative } r' \text{ s.t., for every positive } r\text{, } r \prec r' \\ 0 & \text{otherwise} \end{cases}$$

$$\forall\forall(\langle R, y\rangle) = \begin{cases} + & \text{if for any positive } r' \text{ we have } r \prec r' \text{ for all negative } r \\ - & \text{if for any negative } r' \text{ we have } r \prec r' \text{ for all positive } r \\ 0 & \text{otherwise} \end{cases}$$

	SiCo	AllNot	Most	DefNeg	Thresh	$\forall\exists$	$\exists\forall$	$\forall\forall$
1. Ud	—	—	—	—	—	—	—	—
2. Re	✓	✓	✓	✓	✓	✓	✓	✓
3. RUd	✓	✓	✓	✓	✓	✓	✓	✓
4. FiVa	—	—	—	—	—	—	—	—
5. An	✓	✓	✓	✓	✓	—	—	—
6. Ua	✓	✓	✓	✓	✓	✓	✓	✓
7. Gr	✓	✓	✓	—	✓	✓	✓	✓
8. Ne	✓	✓	✓	—	—	✓	✓	✓
9. Mn	—	—	—	—	—	—	—	—
10. PoMn	✓	✓	✓	✓	✓	✓	✓	—
11. PoCu	✓	✓	✓	✓	✓	✓	✓	✓
12. PoSw	✓	—	✓	✓	—	—	—	—

	SiCo	AllNot	Most	DefNeg	Thresh	∀∃	∃∀	∀∀
1. Ud	—	—	—	—	—	—	—	—
2. Re	✓	✓	✓	✓	✓	✓	✓	✓
3. RUd	✓	✓	✓	✓	✓	✓	✓	✓
4. FiVa	—	—	—	—	—	—	—	—
5. An	✓	✓	✓	✓	✓	—	—	—
6. Ua	✓	✓	✓	✓	✓	✓	✓	✓
7. Gr	✓	✓	✓	—	✓	✓	✓	✓
8. Ne	✓	✓	✓	—	—	✓	✓	✓
9. Mn	—	—	—	—	—	—	—	—
10. PoMn	✓	✓	✓	✓	✓	✓	✓	—
11. PoCu	✓	✓	✓	✓	✓	✓	✓	✓
12. PoSw	✓	—	✓	✓	—	—	—	—

Against Dancy ('04)

	SiCo	AllNot	Most	DefNeg	Thresh	∀∃	∃∀	∀∀
1. Ud	—	—	—	—	—	—	—	—
2. Re	✓	✓	✓	✓	✓	✓	✓	✓
3. RUd	✓	✓	✓	✓	✓	✓	✓	✓
4. FiVa	—	—	—	—	—	—	—	—
5. An	✓	✓	✓	✓	✓	—	—	—
6. Ua	✓	✓	✓	✓	✓	✓	✓	✓
7. Gr	✓	✓	✓	—	✓	✓	✓	✓
8. Ne	✓	✓	✓	—	—	✓	✓	✓
9. Mn	—	—	—	—	—	—	—	—
10. PoMn	✓	✓	✓	✓	✓	✓	✓	—
11. PoCu	✓	✓	✓	✓	✓	✓	✓	✓
12. PoSw	✓	—	✓	✓	—	—	—	—

Not discussed by Horty ('12), Lord & Maguire ('16), Tucker ('22), a.o.

	SiCo	AllNot	Most	DefNeg	Thresh	∀∃	∃∀	∀∀
1. Ud	—	—	—	—	—	—	—	—
2. Re	✓	✓	✓	✓	✓	✓	✓	✓
3. RUd	✓	✓	✓	✓	✓	✓	✓	✓
4. FiVa	—	—	—	—	—	—	—	—
5. An	✓	✓	✓	✓	✓	—	—	—
6. Ua	✓	✓	✓	✓	✓	✓	✓	✓
7. Gr	✓	✓	✓	—	✓	✓	✓	✓
8. Ne	✓	✓	✓	—	—	✓	✓	✓
9. Mn	—	—	—	—	—	—	—	—
10. PoMn	✓	✓	✓	✓	✓	✓	✓	—
11. PoCu	✓	✓	✓	✓	✓	✓	✓	✓
12. PoSw	✓	—	✓	✓	—	—	—	—

Prakken & Sartor ('98) vs., e.g., Bader ('15) and Tucker ('22)

Common language for balancing operations and logical consequence:

- Let L, L_1, …, L_n be the elements of $\mathcal{A}$ or their negations.

Starting with $((R, y), v)$, we write x if $(x, y, +) \in R$ and $\neg x$ if $(x, y, -) \in R$. Depending on the value of v, we can have:

$+$: $L_1, \ldots, L_n \vDash y$

$-$: $L_1, \ldots, L_n \vDash \neg y$

0: Neither of the above

Notice that:

- "x is a reason against y" is very different from "the negation of x implies y";
- `Polarity Monotony` makes sense in reason-based detachment, but not for nonmonotonic inference;
- *Reasoning by cases* is a hallmark of logical inference, but highly controversial for reason-based detachment.

In future work we want to explore:

(1) richer contexts that include considerations that have indirect effects;

(2) contexts that allow reasons with numerical values, representing different "magnitudes"—cf. Keeney & Raiffa ('93);*

(3) incomplete detachment systems and case-based reasoning;*

(4) detachment systems built around richer universes of discourse.

(5) We also want to extend our principle-based analysis further.

Ultimately, our aim is a framework within which one can

(i) define, relate, and compare various different (and possibly complex) accounts of the way reasons interact to support actions; as well as

(ii) relate these accounts to the ideas from MCDM, nonmonotonic reasoning, argumentation, and related disciplines.

Thanks!

R. Bader (2016), "Conditions, modifiers, and holism", In: Lord E. & Maguire B. (eds.) *Weighing Reasons*, Oxford University Press.

J. Dancy (2004), *Ethics without Principles*, Oxford University Press.

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

R. Keeney & H. Raiffa (1993), *Decisions with Multiple Objectives: Preferences and Value*, Oxford University Press.

E. Lord & B. Maguire (2016), "An opinionated guide to the weight of reasons", In: Lord E. & Maguire B. (eds.) *Weighing Reasons*, Oxford University Press.

D. Makinson & L. van der Torre (2001), "Input/output logics", In: *Journal of Philosophical Logic*, 29.

D. Parfit (2011), *On What Matters*, vol. 1, Oxford University Press.

H. Prakken & G. Sartor (1998), "Modelling reasoning with precedents in a formal dialogue game", In: *Artificial Intelligence and Law*, 6.

J. Raz (1990), *Practical Reason and Norms*, Oxford University Press.

T. M. Scanlon (1998), *What We Owe to Each Other*, Oxford University Press.

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

C. Tucker (2022), "A holist balance scale", In: *Journal of the American Philosophical Association*, First View, pp. 1–21.