Reason-Based Detachment

Aleks Knoks

University of Luxembourg
Maison du Nombre, 6, Av. de la Fonte
4365 Esch-sur-Alzette, Luxembourg

Leendert van der Torre

University of Luxembourg
Maison du Nombre, 6, Av. de la Fonte
4365 Esch-sur-Alzette, Luxembourg

Abstract

The more recent philosophical literature on foundational questions about normativity relies heavily on the notion of normative reasons, understood as considerations that count in favor or against actions: the notion is used when answering various kinds of normative and metanormative questions and when analyzing other normative notions. The interaction between normative reasons is often made sense of by analogy with weight scales. This paper, by contrast, construes it as a type of inference pattern—titular reason-based detachment—and analyzes it from first principles. While very abstract and exploratory, the approach offers a novel perspective on the (philosophical) idea of weighing normative reasons, and promises to let us relate it to the broader concerns of nonmonotonic logic and related disciplines.

Keywords: detachment, principles, reasons, weighing.

1 Introduction

When philosophers talk about normative matters—about what is right, obligatory, permitted, and so on—they tend to rely on the notion of *normative reasons*, understanding them as considerations that count in favor of or against actions (or attitudes).[1] The notion has become a mainstay of practical philosophy, where it is routinely used when answering various normative and metanormative questions. This is taken to the extreme in the *reasons-first program* which holds, roughly, that the notion of reason is basic, and that all other normative notions should be analyzed in terms of it.[2] When discussing the interaction between reasons, philosophers often use phrases such as "the

[1] The philosophical literature distinguishes between normative, motivating, and explanatory reasons—see [2]. We restrict our attention to normative reasons here.

[2] The locus classicus here is Scanlon [24]. But see also, e.g., [21], [23], [25].

action supported on the balance of reasons" and "the reasons for outweigh the reasons against", inviting an image of *weight scales*. The simplest version of these normative scales is meant to work roughly as follows. [3] The reasons in favor of φ-ing go in one pan of the scales, the reasons against φ-ing go in the other. If the weight of the reasons in the first pan is greater than the weight of the reasons in the second pan, φ ought to be carried out. If the weight of the reasons in the second pan is greater, φ ought not to be carried out. [4]

Philosophers have explored various ideas about the exact workings of normative weight scales and have looked at some alternatives. [5] However, with few exceptions, these investigations have been carried out informally, and the more formal investigations have focused on exploring particular models. [6] In this paper, we propose to think of the weight scales as a kind of inference pattern. We call this pattern *reason-based detachment*, and the goal we set ourselves here is to set up and begin to explore a general formal framework built around it. [7] We start with the general notion of *detachment systems*—which can be thought of as structures in which reason-based detachment is guaranteed to be valid—and we formulate a number of principles or properties that a detachment system can satisfy. Then we focus on a class of detachment systems called *balancing operations*, and formulate and discuss a handful of further principles specific to them. For instance, the principle we call *Neutrality* requires, roughly, that reasons of opposing polarity—reasons for and against—are treated equally, while the principle we call *Fixed Value* requires that a reason's polarity always stays the same. We also define several concrete balancing operations, or, roughly, methods specifying how to determine whether φ is supported on the balance of reasons.

The rest of this paper is structured as follows. Section 2 introduces the core formal concepts—including detachment system—and principles that detachment systems can satisfy. Section 3 defines the concept of balancing operation and discusses principles that balancing operations can satisfy. Sections 4–5 discuss two different types of concrete balancing operations. Section 6 presents our principle-based analysis of reason-based detachment. Section 7 takes a first step towards relating reason-based detachment to logical consequence. Section 8 explains where we plan to take the project we started here in the future. Finally, the rather brief Section 9 presents our conclusions.

---

[3] Cf., e.g., [14] and [27].

[4] Cf., e.g., [6] and [27].

[5] While the scales model has its detractors, it is fair to say that it is the dominant model, and that it is often simply taken for granted—see, e.g., Broome's inquiry into the normativity of rationality [4]. For detractors, see [6, 8, 10, 12, 26].

[6] For the latter, see [7, 9, 12].

[7] It pays noting that our approach is similar to the methodology underlying input/output logic [16, 17] which is built around factual detachment—see [20, pp. 502–5] for a discussion.

2 Detachment systems

2.1 Core formal notions

In general, a detachment system is a two-place relation between, on the one hand, an issue (an element of the universe of discourse) with a set of reasons (other elements of the universe of discourse with a value) and, on the other hand, a value. We call an issue together with a set of reasons a *context*. Thus, a detachment system is a relation between contexts and values.

To facilitate the formal presentation, and to be more flexible, we represent reasons as follows:

Definition 2.1 [Reasons] Let $\mathcal{A}$ be an infinite set called the *universe of discourse*, and let $\mathcal{V}$ be a set called *values*. 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 $\mathcal{V}$.[8]

Our formal definition of context is as follows:

Definition 2.2 [Contexts] 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 $\mathcal{A}$.

Note that this general representation of contexts and reasons allows for such contexts as $(\{(a, y, v), (b, z, v')\}, y)$. One may wonder whether the latter reason is not superfluous in this context, and whether this context isn't the same as $(\{(a, y, v)\}, y)$. It is exactly these kinds of general considerations that we want to be explicit about in our formal framework. Below, we call this particular property *Relevance*.

Definition 2.3 [Detachment systems] A detachment system $\mathcal{D}$ is a two-place relation between contexts and values from $\mathcal{V}$.

We call elements that comprise a detachment system *detachments*.

2.2 Principles for detachment systems

In general, we identify properties of detachment systems. While we call them *principles*, we could also have called them *axioms*. In the context of this paper, 'properties', 'principles' and 'axioms' are used synonymously. They can be used to classify and distinguish between different detachment systems. Some of these properties may be seen as desirable and, therefore, could be called *postulates* or *desiderata*. However, it is important to note that not all of our principles have the status of a desideratum. In fact, some of them, like the Monotony Principle, are clearly undesirable. Nevertheless, it is useful to make such undesirable properties as Monotony explicit and formal as well. This is why we prefer to refer to the properties as *principles*. They can be used in a principle-based analysis of reason-based detachment, as shown in Section 6.

[8] The reader familiar with the philosophical literature on reasons will notice that our formal notion of reason corresponds more closely to what is often called *reason relation*. This is hardly a problem, since the two are closely related and reasons can be read from the relation.

It is natural to think that detachment systems should be *complete*, or that we should be able to detach a value for every context. Note, however, that there are at least two ways to make the intuitive notion of completeness more precise, as Principles 2.4 and 2.6 make clear. The general completeness property, as expressed in Principle 2.4 (Universal Domain), is quite strong. If this property of completeness is considered to be too strong, we also consider the notion of completeness with respect to a set of reasons, as expressed in Principle 2.6 (Reason Universal Domain). This notion is more complicated because the set of reasons depends on the issue under discussion.

Principle 2.4 (Universal Domain) *A detachment system $\mathcal{D}$ is said to satisfy Universal Domain,* `Ud`*, just in case, it is total, that is, for every context C, there is a value v such that $(C, v) \in \mathcal{D}$.*

Our second principle states that the assignment of a value to an issue (in a context) is determined solely on the basis of the reasons that have to do with that issue. Other reasons can be removed from the context without affecting the result.

Principle 2.5 (Relevance) *A detachment system $\mathcal{D}$ satisfies Relevance,* `Re`*, just in case $((R, y), v) \in \mathcal{D}$ if and only if $((R_y, y), v) \in \mathcal{D}$, where $R_y = \{(x, y, v') \in R : x \in \mathcal{A}$ and $v' \in \mathcal{V}\}$.*

In case the first principle is considered to be too strong, we can use our third principle instead. This principle makes use of the notion of the *universe of reasons $\mathcal{R}_y$ of an issue y* in a detachment system, which is the set of reasons that occur in some context for the issue for which the detachment system is defined. What the principle requires, then, is that a value can be detached for every context as long as its set of reasons is a subset of the universe of reasons of its issue.

Principle 2.6 (Reason Universal Domain) *Let $\mathcal{D}$ be a detachment system and y an element of $\mathcal{A}$. Let $\mathcal{R}_y = \bigcup\{R : ((R, y), v) \in \mathcal{D}\}$, called the* universe of reasons *of y. Then $\mathcal{D}$ is said to satisfy Reason Universal Domain,* `RUd`*, just in case, for any $y \in \mathcal{A}$ and for any $R \subseteq \mathcal{R}_y$, there is a value v such that $((R, y), v) \in \mathcal{D}$.*

This principle may sound circular at first, but it is not. One can check whether a detachment system satisfies Reason Universal Domain by first determining the universe of reasons for every element, and then checking whether for that element all other combinations are also present. In what follows, we will often talk about the universe of reasons $\mathcal{R}$ associated with a detachment system without qualification: it is but the union of the universes of reasons of all issues.

If a detachment relation satisfies Reason Universal Domain but not Universal Domain, then various other principles can be defined. Our fourth principle, Fixed Value, is a case in point. It states that if x is a v type of reason for y in an universe of reasons, it cannot occur as another type of reason for y in this universe of reasons.

Principle 2.7 (Fixed Value) *Let $\mathcal{D}$ be a detachment system and let $\mathcal{R} = \bigcup\{R : ((R, y), v) \in \mathcal{D}\}$. Then $\mathcal{D}$ is said to satisfy Fixed Value,* `FiVa`*, just in case, for any two $(x, y, v), (x, y, v') \in \mathcal{R}$, we have $v = v'$.*

The idea that reasons never change their polarities is one of the core tenets of a philosophical view called *atomism*—we will say a little more about this view in Section 6.

While we could formulate more principles strengthening Reason Universal Domain, for reasons of space, we move on to principles of a different kind. And the fifth principle we introduce is Anonymity—we could also have called it *Syntax Independence*. Intuitively, a detachment system satisfies Anonymity when all elements in the universe are treated equally. (In Section 4, we illustrate this property using six balancing operations while in Section 5, we discuss three balancing operations that do not satisfy it.)

Principle 2.8 (Anonymity) *A detachment system $\mathcal{D}$ satisfies Anonymity,* `An`*, just in case, for every $((R, y), v) \in \mathcal{D}$ and any bijection $\pi : \mathcal{A} \mapsto \mathcal{A}$, if we have $((\{(\pi(x), \pi(z), v') : (x, z, v') \in R\}, \pi(y)), v'') \in \mathcal{D}$, then $v'' = v$.*

Our sixth principle is Unanimity. It states that if all the reasons for an issue are of v type, then the assignment should also be of the corresponding type.

Principle 2.9 (Unanimity) *A detachment system $\mathcal{D}$ is said to satisfy Unanimity,* `Ua`*, just in case, for any context $C = (R, y)$, if there is some $(x, y, v) \in R$ and, for all other $(z, y, v') \in R$, we have $v = v'$, then $((R, y), v) \in D$.*

Our seventh principle is Groundedness. It can be seen as the inverse of Unanimity. It states that if a context is assigned some value v, then its set of reasons should contain at least one reason of the corresponding type.

Principle 2.10 (Groundedness) *A detachment system $\mathcal{D}$ satisfies Groundedness,* `Gr`*, just in case, for any $((R, y), v) \in \mathcal{D}$ with $v \neq 0$, there is some $r = (x, y, v) \in R$.*

3 Balancing operations

Our main focus in this paper is on a particular type of detachment system that we call *balancing operation*. These are (more) closely related to the informal model of normative weight scales.

3.1 Balancing operations defined

Balancing operations are specific detachment systems (for basic weight scales) with the following properties:

(i) Contexts can only be related to the values $+$, $-$, or 0, reflecting the weight scales metaphor: leaning towards the "for" side, leaning towards the "against" side, or being equally balanced.

(ii) Reasons can only have the value $+$ or $-$, reflecting whether they are reasons for or against a given issue. (Note that an element can be a positive reason for one issue and a negative reason for another.)

(iii) Contexts are related to exactly one value. [9]

More formally:

Definition 3.1 [Balancing operations] Let $\mathcal{A}$ be an infinite set of propositional atoms and $\mathcal{V}$ the set $\{+, 0, -\}$. A detachment system $\mathcal{D}$ is called a *balancing operation* just in case it is a function from $2^{\mathcal{A}\times\mathcal{A}\times\{+,-\}} \times \mathcal{A}$ to $\mathcal{V}$.

The reader may wonder about the difference between how + and − and true and false. This issue is taken up in Section 7, where we discuss the differences between balancing operations and logical relations.

Before we turn to principles specific to balancing operations, we introduce some useful formal notation:

- Where $v \in \{+, 0, -\}$, we let $\overline{v}$ stand for the value that is opposite to v, that is: $\overline{v} = -$ if $v = +$; $\overline{v} = +$ if $v = -$; and $\overline{v} = 0$ if $v = 0$.
- Where $r = (x, y, v)$ is a reason, let $ground(r) = x$, $action(r) = y$, and $polarity(r) = v$.
- Where R is a set of reasons and $y \in \mathcal{A}$, the set of reasons from R that *speak in favor of* y is the set $positive(R, y) = \{r \in R : r = (x, y, +)\}$; the set of reasons from R that *speak against* y is the set $negative(R, y) = \{r \in R : r = (x, y, -)\}$; and the set of reasons relevant to y is the set $relevant(R, y) = positive(R, y) \cup negative(R, y)$. (We follow Raz [23] in calling reasons for *positive* and reasons against *negative*.)

3.2 Principles for balancing operations

The first principle pertaining to balancing operations—and the eighth principle overall—is called *Neutrality*. Where Anonymity states that reasons are to be treated equally, Neutrality states that values are to be treated equally. Roughly, if we switch + and − in the context, and vice versa, then the assignment switches its value too. Of all the principles we discuss, this is perhaps the one that is most characteristic of weight scales. Somewhat surprisingly, to the best of our knowledge, this characteristic principle has not yet been formalized in the literature on reasons.

Principle 3.2 (Neutrality) *Let $\mathcal{D}$ be a detachment system and let $\mathcal{R} = \bigcup\{R : ((R, y), v) \in \mathcal{D}\}$. Then $\mathcal{D}$ is said to satisfy Neutrality,* `Ne`*, just in case, for every $((R, y), v) \in \mathcal{D}$, if $R' = \{(x, y, \overline{v}) : (x, y, v) \in relevant(R, y)\} \subseteq \mathcal{R}$, then $((R', y), \overline{v}) \in \mathcal{D}$.*

The remaining four principles we discuss describe (non)monotonicity properties. Our ninth principle is called Monotony. It states that if a context gets assigned a nonzero value, adding more reasons to it is not going to change the value that gets assigned.

Principle 3.3 (Monotony) *Let $\mathcal{D}$ be a detachment system and let $\mathcal{R} = \bigcup\{R : ((R, y), v) \in \mathcal{D}\}$. Then $\mathcal{D}$ satisfies Monotony,* `Mn`*, just in case,*

[9] Thus, balancing operations are deterministic relations.

if $((R, y), v) \in \mathcal{D}$ *where* $v \neq 0$ *and* $(x, y, v') \in \mathcal{R}$*, then we have* $((R \cup \{(x, y, v')\}, y), v) \in \mathcal{D}$*.*

Clearly Monotony is not a desirable property for balancing operations, and most operations defined in the literature are nonmonotonic. This raises the question of whether there are weaker principles than Monotony that can be defined for balancing operations. As a first response, we formulate a principle called *Polarity Monotony*. If + gets detached, then adding a positive reason will not change the assignment, and this applies also for − and negative reasons. While the principle is not uncontroversial, it seems intuitive, and it is satisfied by all but one of the balancing operations defined in this paper.

Principle 3.4 (Polarity Monotony) *A detachment system* $\mathcal{D}$ *satisfies Polarity Monotony,* `PoMn`*, just in case, if* $((R, y), v) \in \mathcal{D}$ *where* $v \neq 0$ *and* $(x, y, v) \in \mathcal{R}$*, then we have* $((R \cup \{(x, y, v)\}, y), v) \in \mathcal{D}$*.*

The next principle we discuss is Polarity Cut. It can be seen as the inverse of Polarity Monotony. If a positive value gets detached, then removing a negative reason from the context doesn't affect the detachment. This is the case also for detachments of negative values and positive reasons.

Principle 3.5 (Polarity Cut) *A detachment system* $\mathcal{D}$ *is said to satisfy Polarity Cut,* `PoCu`*, just in case, for any* $((R \cup \{(x, y, \overline{v})\}, y), v) \in \mathcal{D}$ *with* $v \neq 0$*, we have* $((R, y), v) \in \mathcal{D}$*.*

The twelfth and final principle we introduce is Polarity Switching. It can be seen as a strong kind of nonmonotonicity. It assumes that the universe of reasons is infinite, and it states that, for every context that gets assigned a positive value, we can extend the context if we have enough negative reasons so that the resulting context gets assigned a negative value, and vice versa.

Principle 3.6 (Polarity Switching) *Let* $\mathcal{D}$ *be a detachment system and let* $\mathcal{R} = \bigcup\{R : ((R, y), v) \in \mathcal{D}\}$*. Then* $\mathcal{D}$ *satisfies Polarity Switching,* `PoSw`*, just in case, for any* $((R, y), v) \in \mathcal{D}$*, there is an* $R' = \{(x, y, \overline{v}) : r \in \mathcal{R}\}$ *such that* $((R \cup R', y), \overline{v}) \in \mathcal{D}$*.*

Having introduced the principles, we turn to concrete balancing operations.

4 Anonymous balancing operations

Over the course of this section and the next, we introduce a handful of balancing operations. All of them are defined with respect to a universe of reasons. In this section, we discuss *anonymous* balancing operations.

Definition 4.1 [Anonymous balancing operations] Let $\mathcal{D}$ be a detachment system and let $\mathcal{R} = \bigcup\{R : ((R, y), v) \in \mathcal{D}\}$. Then $\mathcal{D}$ is an *anonymous balancing operation* just in case $\mathcal{D}$ satisfies:

(i) the Reason Universal Domain principle (with respect to $\mathcal{R}$); and

(ii) the Anonymity principle.

According to our first sample balancing operation, the context (R, y) is assigned the value $+$ in case the sheer number of reasons speaking in favor of y is greater than the number of reasons speaking against y; it gets assigned the value $-$ in case the number of reasons against y is greater than the number of reasons for y; and it gets assigned 0 otherwise. More formally:

Definition 4.2 [Simple Counting] Let $\mathcal{D}$ be an anonymous balancing operation. Then $\mathcal{D}$ is called *Simple Counting* just in case:

- $((R, y), +) \in \mathcal{D}$, if $|positive(R, y)| > |negative(R, y)|$;
- $((R, y), -) \in \mathcal{D}$, if $|negative(R, y)| > |positive(R, y)|$;
- $((R, y), 0) \in \mathcal{D}$, otherwise.

Admittedly, Simple Counting—as well as the other balancing operations we are about to define—is, well, very simple and inadequate for most practical purposes. That is, if we think back to the balancing scales metaphor from the philosophical literature and use Simple Counting as a concrete proposal regarding how to assign deontic statuses to actions—with the assignment of $+$ $(-)$ standing for the conclusion that the action ought (not) to be carried out—then we would surely get many cases wrong. That being said, Simple Counting does justice to at least two important features that are inherent in the idea of normative weight scales. First, it treats positive and negative reasons in a symmetric fashion. Second, it can be seen as adding up the weights of reasons while relying on the assumption that the magnitude (or "weightiness") of all reasons is the same.

It is worth making it explicit that Definition 4.2 does not define a single balancing operation but a class of balancing operations: one for every different universe of reasons $\mathcal{R}$. The same applies to the other balancing operations defined in this section.

Our second balancing operation, called *All or Nothing*, assigns the value $+$ to the context (R, y) if all the reasons that concern y in R are positive, and the value $-$ if all such reasons are negative. In case neither of these conditions obtain, the context gets assigned 0.

Definition 4.3 [All or Nothing] Let $\mathcal{D}$ be an anonymous balancing operation. Then $\mathcal{D}$ is called *All or Nothing* just in case:

- $((R, y), +) \in \mathcal{D}$ if $positive(R, y) = relevant(R, y) \neq \emptyset$;
- $((R, y), -) \in \mathcal{D}$ if $negative(R, y) = relevant(R, y) \neq \emptyset$;
- $((R, y), 0) \in \mathcal{D}$ otherwise.

Our third balancing operation can be thought of as lying in between Simple Counting and All or Nothing. The intuitive idea behind it is that the context (R, y) gets assigned the value $+$ $(-)$ where *most* reasons that are relevant to y argue in favor of (or against) y. For simplicity, we assume that 'most reasons'

translates into at least four times as many reasons. [10]

Definition 4.4 [Most Reasons] Let $\mathcal{D}$ be an anonymous balancing operation. Then $\mathcal{D}$ is called *Most Reasons* just in case:

- $((R, y), +) \in \mathcal{D}$ if $|positive(R, y)| \geq 4 \times |negative(R, y)|$ and $relevant(R, y) \neq \emptyset$;
- $((R, y), -) \in \mathcal{D}$ if $|negative(R, y)| \geq 4 \times |positive(R, y)|$ and $relevant(R, y) \neq \emptyset$;
- $((R, y), 0) \in \mathcal{D}$ otherwise.

The next operation assigns $-$ to a context, as long as it is not the case that there are more positive than negative reasons for y. (In the latter case, the context gets assigned a $+$.)

Definition 4.5 [Default Negative] Let $\mathcal{D}$ be an anonymous balancing operation. Then $\mathcal{D}$ is called *Default Negative* just in case:

- $((R, y), +) \in \mathcal{D}$, if $|positive(R, y)| > |negative(R, y)|$;
- $((R, y), -) \in \mathcal{D}$, otherwise.

Our final balancing operation is similar to Simple Counting, except now there is a threshold that changes the rules of the game: once there are enough positive reasons (the threshold is met), the existence of further negative reasons to the contrary ceases to matter. The idea behind this operation comes from the literature on *threshold deontology*. Advocates of threshold deontology hold, roughly, that deontological norms are to be followed up to a point even if there are adverse consequences, but when the consequences become so dreadful that they cross some threshold, consequentialism takes over. [11]

Definition 4.6 [Threshold] Let $\mathcal{D}$ be an anonymous balancing operation. Then $\mathcal{D}$ is called *Threshold* just in case:

- $((R, y), +) \in \mathcal{D}$, if $|positive(R, y)| \geq 100$ or $|positive(R, y)| > |negative(R, y)|$;
- $((R, y), -) \in \mathcal{D}$, if $|positive(R, y)| < 100$ and $|negative(R, y)| > |positive(R, y)|$;
- $((R, y), 0) \in \mathcal{D}$, otherwise.

For all the balancing operations defined so far, given a context (R, y), one does not need to look beyond R to determine which value to assign to the context. What's more, it is not difficult to see that all of these operations satisfy Anonymity (Principle 2.8). But, to anticipate the discussion in Section 6, only the first three of them satisfy Neutrality (Principle 3.2).

[10] This simple proposal is meant to serve as an illustration of a more general idea or scheme for specifying 'most reasons'.

[11] See, e.g., [1, Sec. 4] or the more recent [5, 18]. Note that the balancing operation is only inspired by the literature on threshold deontology and is not meant to capture any particular account.

5 Relational balancing operations

In this section, we turn to a different class of balancing operations. These assign values to contexts on the basis of the reasons within them, along with a binary anti-symmetric relation $\prec$ on the reasons:

Definition 5.1 [Relation $\prec$] Given a detachment system $\mathcal{D}$ with its underlying set of reasons $\mathcal{R} = \bigcup\{R : ((R, x), v) \in \mathcal{D}\}$, an anti-symmetric relation $\prec$ on $\mathcal{R}$ is a subset of $\mathcal{R} \times \mathcal{R}$ such that $(r, r') \in \prec$ only if $\mathit{polarity}(r) = \overline{\mathit{polarity}(r')}$.

Notice that two reasons can stand in the $\prec$ relation only if one of them is positive and the other negative. Instead of $(r, r') \in \prec$, we will write $r \prec r'$. An expression of the form $r \prec r'$ can be thought of in terms of r' having strictly more weight than r, or r' defeating r.

With this, we can state the general definition of balancing operations discussed in this section.

Definition 5.2 [Relational balancing operations] Let $\mathcal{D}$ be a detachment system, let $\mathcal{R} = \bigcup\{R : ((R, y), v) \in \mathcal{D}\}$, and let $\prec$ be a binary anti-symmetric relation over $\mathcal{R}$, as in Definition 5.1. Then $\mathcal{D}$ is a *relational balancing operation* (for $\mathcal{R}$ and $\prec$) just in case $\mathcal{D}$ satisfies:

(i) the Reason Universal Domain principle (with respect to $\mathcal{R}$); and

(ii) for all $((R, y), v) \in \mathcal{D}$, there is no $r = (x, y, \overline{v}) \in R$ such that $r' \prec r$ for every $r = (z, y, v) \in R$.

Notice that Clause (ii) states that $+$ cannot be detached from (R, y) in case there is some positive reason r for y that stands in the $\prec$ relation to—or is better than, or defeats—every reason against y; and similarly for $-$. This is a very weak property.

We proceed to define some concrete relational balancing operations, or, rather classes of them: much like in the previous section, we get different balancing operations for different $\mathcal{R}$ and $\prec$. We call the first class *Exists Better Reason*. It assigns $+$ to (R, y) in case, for every reason against y, there is a stronger reason for y; and it assigns $-$ to (R, y) in case, for every reason for y, there is a stronger reason against y.

Definition 5.3 [Exists Better Reason, $\forall\exists$] Let $\mathcal{D}$ be a relational balancing operation. Then $\mathcal{D}$ is called *Exists Better Reason* just in case:

- $((R, y), +) \in \mathcal{D}$, if $\mathit{relevant}(R, y) \neq \emptyset$ and, for every $r = (x, y, -) \in R$, there is an $r' = (z, y, +) \in R$ such that $r \prec r'$;
- $((R, y), -) \in \mathcal{D}$, if $\mathit{relevant}(R, y) \neq \emptyset$ and, for every $r = (x, y, +) \in R$, there is an $r' = (z, y, -) \in R$ such that $r \prec r'$;
- $((R, y), 0) \in \mathcal{D}$, otherwise.

Our next balancing operation, Decisive Reason, is more demanding: it assigns $+$ ($-$) to a context (R, y) just in case there exists a reason for (or against)

y that is stronger than all reasons to the contrary. [12]

Definition 5.4 [Decisive Reason, $\exists\forall$] Let $\mathcal{D}$ be a relational balancing operation. Then $\mathcal{D}$ is called *Decisive Reasons* just in case:

- $((R, y), +) \in \mathcal{D}$, if there is an $r' = (x, y, +) \in R$ such that $r \prec r'$ for every $r \in R$ with $action(r) = y$ and $polarity(r) = -$;
- $((R, y), -) \in \mathcal{D}$, if there is an $r' = (x, y, -) \in R$ such that $r \prec r'$ for every $r \in R$ with $action(r) = y$ and $polarity(r) = +$;
- $((R, y), 0) \in \mathcal{D}$, otherwise.

The third operation, All Reasons Better, is even more demanding than Decisive Reason: it assigns the value $+$ $(-)$ to a context (R, y) just in case *all* reasons for (against) y are stronger than *all* reasons against (for) y. Otherwise, it assigns the value 0.

Definition 5.5 [All Reasons Better, $\forall\forall$] Let $\mathcal{D}$ be a relational balancing operation. Then $\mathcal{D}$ is called *All Reasons Better* just in case:

- $((R, y), +) \in \mathcal{D}$, if $relevant(R, y) \neq \emptyset$ and, for every $r' \in R$ with $action(r') = y$ and $polarity(r') = +$, $r \prec r'$ for every $r \in R$ with $action(r) = y$ and $polarity(r) = -$;
- $((R, y), -) \in \mathcal{D}$, if $relevant(R, y) \neq \emptyset$ and, for every $r' \in R$ with $action(r') = y$ and $polarity(r') = -$, $r \prec r'$ for every $r \in R$ with $action(r) = y$ and $polarity(r) = +$;
- $((R, y), 0) \in \mathcal{D}$, otherwise.

This operation can be seen as injecting the idea underlying Decisive Reason into the All or Nothing operation described in the previous section.

6 Chart

Now that we have defined a number of principles and a handful of balancing operations, we can analyze and compare the operations by looking at the principles that they satisfy and the ones that they do not. For example:

Proposition 6.1 *Simple Counting (Definition 4.2) satisfies Polarity Monotony (Principle 3.4).*

Proof. Let $\mathcal{D}$ be Simple Counting. Consider an arbitrary context (R, y) and suppose that we have $((R, y), v) \in \mathcal{D}$ with $v \neq 0$. Either $v = +$, or $v = -$. Without loss of generality, we suppose that $v = +$. By Definition 4.2 (Simple Counting), we can be sure that $|positive(R, y)| > |negative(R, y)|$. Now let's consider the context $(R \cup \{r\}, y)$ where $action(r) = y$ and $polarity(r) = +$. Since $action(r) = y$ and $polarity(r) = +$, we have $|positive(R \cup \{r\}, y)| = 1 + |positive(R, y)|$. But $1+|positive(R, y)| > |positive(R, y)| > |negative(R, y)| =$

[12] The terms 'decisive reason' and 'decisive reasons' feature prominently in the philosophical literature—see, e.g., [14].

Table 1
Summary of the principle-based analysis of balancing operations

	SiCount	AllNoth	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	✓	–	✓	✓	–	–	–	–

$|negative(R \cup \{r\}, y)|$, which, by Definition 4.2, is enough for $((R \cup \{r\}, y), +) \in \mathcal{D}$.

□

Proposition 6.2 *All Reasons Better (Definition 5.5) does not satisfy Polarity Monotony (Principle 3.4).*

Proof. Let $\mathcal{D}$ be a detachment system with the universe of reasons $\mathcal{R} = \{(a, d, +), (b, d, +), (c, d, -)\}$ and with $\prec = \{((c, d, -), (a, d, +))\}$, and let $\mathcal{D}$ assign values to contexts in accordance with Definition 5.5. It is not difficult to verify that $D_1 = (\{(a, d, +), (c, d, -)\}, d, +) \in \mathcal{D}$, and that $D_2 = (\{(a, d, +), (c, d, -), (b, d, +)\}, d, 0) \in \mathcal{D}$. This, however, means that the claim that if $((R, y), +) \in \mathcal{D}$, then $((R \cup \{(x, y, +)\}, y), +) \in \mathcal{D}$ does not hold of $\mathcal{D}$, implying that $\mathcal{D}$ does not satisfy Polarity Monotony. □

The proofs of the remaining propositions—that is, the propositions that show which other principles are (not) satisfied by which (other) operations—are about as straightforward as those of Propositions 6.1 and 6.2. For this reason, we omit them here, letting Table 1 summarize the lay of the land: the topmost row lists the balancing operations; the leftmost column lists the principles; the remaining cells state whether the given operation does (✓) or doesn't (–) satisfy the given principle. For example, the third column makes it clear that the balancing operation we called *All or Nothing* (Definition 4.3) satisfies all principles except for Fixed Value, Monotony, and Polarity Switching.

The table provides a way to compare the various operations. It can also be used to get important insights about both operations and principles. In the remainder of this section, we hint at these and flag some other issues of broader significance.

The chart indicates that none of the balancing operations satisfy Universal Domain (Principle 2.4) while all of them satisfy Relevance and Reason

Universal Domain (Principles 2.5 and 2.6). Note that there is an ambiguity here. Recall that each of our definitions from Sections 4–5 specify not a single balancing operation but a class of balancing operations: roughly, we get a different operation for every different universe of reasons. Note that some of these operations do satisfy Universal Domain, namely, those whose universe of reasons $\mathcal{R}$ contains all possible reasons that can be constructed from $\mathcal{A}$ and $\{0, +\}$—call this $\mathcal{R}_{\mathcal{A}}$. However, those operations whose universe of reasons is more restricted do not satisfy the principle. Thus, what the chart indicates is that Universal Domain does not hold true of all operations that belong to the class.

This makes the importance of the universe of reasons underlying a given detachment relation very clear, and the reader may wonder why we allow it to be more restricted than $\mathcal{R}_{\mathcal{A}}$. The motivation here comes from philosophical literature where the standard view is that only some facts (or types of facts) can ever constitute reasons—even if the views on what those facts are are very different. As for Principles 2.5 and 2.6, if we think of balancing operations as serving the same function as the weight scales model serves in the philosophical literature, then we want them to satisfy both of these principles. In the end, it would be very strange if, in certain cases, the deontic status of an action could be genuinely (not epistemically) indeterminate, or if it was determined by something irrelevant.

Recall that Fixed Value (Principle 2.7) states that the universe of reasons underlying a detachment system cannot contain two reasons of the form $(x, y, +)$ and $(x, y, -)$. This idea has a correlate in the philosophical literature, where there is a well-known view called *atomism* that says that a reason can never change its weight, which includes both its polarity and magnitude. Furthermore, there is a well-known argument against the weight scales model that goes roughly as follows. (1) The weight scales model entails atomism. (2) Atomism is false. Hence, (3) the weight scales model has to be wrong. [13] The argument is now widely considered to be flawed—see, e.g., [27]—and our framework provides further support for this. We can see our balancing operations as simplified concrete specifications of the weight scales model. The fact that none of them satisfy Fixed Value suggests that statement (1) must be false.

We take Anonymity and Neutrality (Principles 2.8 and 3.2) to be among the most important principles we have identified. Even though we defined the balancing operations in Section 4 using Anonymity, it is natural to see it as their characteristic principle. Neutrality, in its turn, formalizes an idea that seems inherent in the weight scales metaphor: that positive and negative reasons are to be treated symmetrically—this, of course, is not to say that this idea cannot be questioned. We can see at a glance that operations that don't satisfy Neutrality, namely, Default Negative and Threshold (Definitions 4.5 and 4.6), do not treat positive and negative reasons symmetrically.

Unanimity, Polarity Monotony, and Polarity Cut (Principles 2.9, 3.4,

[13] See, e.g., [6].

and 3.5) may look like very natural principles, and one may even be tempted to think that the fact that All Reasons Better doesn't satisfy Polarity Monotony shows that it is a bizarre balancing operation. However, there is a well-known example from the AI & law literature that could make one question all three principles. The example describes the effects of heat and rain on one's decision as to whether or not to go jogging: taken by themselves, the facts that it is raining, and that it is hot constitute reasons for you not to go jogging, but when taken in combination, they make it rational to go jogging [22]. On the face of it, in this example we are dealing with two negative reasons of the form $(x, y, -)$ and $(z, y, -)$ and a detachment system that includes the following three detachments: $((\{(x, y, -)\}, y), -)$, $((\{(z, y, -)\}, y), -)$, and $((\{(x, y, -), (z, y, -)\}, y), +)$. This means that there's a tension between all three of these principles and the most straightforward analysis of the jogging scenario. Those who like the analysis will deny the principles. Those who like the principles will have to argue that the scenario is misdescribed—here see, e.g., [3, 15, 19, 27].

7 Balancing operations and logical consequence relations

Although we have conceptualized detachments as pairs of the form $((R, x), v)$, they can be rewritten either as triples of the form (R, x, v), or as pairs of the form $(R, (x, v))$. This flexibility is an advantage of the framework. In some applications, it is useful to see a detachment system as a binary relation $((R, x), v)$. From a mathematical point of view, it can be seen as a function (if we add some conditions, as we did in Section 3). From an application point of view, it is more like a weight scales for x. But in other applications, it is useful to see a detachment system as a binary relation $(R, (x, v))$. From a mathematical point of view, it should not be seen as a function. From an application point of view, it is more similar to deductive systems and logical relations.

Indeed, given that we used Boolean values when defining balancing operations, it is natural to wonder about their relationship to the logical languages used in propositional logic, logic programming, and nonmonotonic inference relations. Of course, the balancing operations that we considered in Section 4 and 5—as well as other operations from the (informal) literature—are quite different to what can be found in the semantics of propositional logic, logic programming, or nonmonotonic logic. So a syntactic correspondence between the languages may be of interest mainly for technical reasons. Nevertheless, for the definitions of principles, it may be illustrative to define a common language for balancing operations and logical consequence relations. So, here we make such a common language explicit.

To represent balancing operations, we identify the universe of discourse with propositional atoms, and reasons with literals. Let L, L_1, ..., L_n be the elements of the universe of discourse or their negations. A logic programming rule is written as L :- L_1, ..., L_n, stating that L holds if L_1, ..., L_n hold. In propositional logic, this is often written as the rule $L_1 \wedge \ldots \wedge L_n \rightarrow L$.

Alternatively, we could write $L_1, \ldots, L_n \vDash L$. In the latter case, consider an issue y, a set of reasons for y, and a set of reasons against y. If x is a reason for y, we write it as x, and if x is a reason against y we write it as $\neg x$. There are three assignments for y in a context:

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

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

0: neither of the above

In this way, all of the principles we have defined can be rewritten as principles of logical consequence relations.

With this translation, we have a unified language for reason-based entailment and nonmonotonic inference, but the semantics will be very different. The main difference concerns the interpretation of negation. In reason-based detachment, "x is a reason against y" means something completely different to the logical inference "the negation of x implies y". In particular, we must be careful when comparing or importing principles from one area into another. Consider reasoning by cases, one of the hallmarks of logical inference. If "x implies y" and "the absence of x implies y" both hold, then y holds unconditionally. Whether a similar inference pattern holds for reason-based entailment is more controversial. Perhaps even more clearly, the nonstandard reading of negation on the left becomes very clear when we consider Polarity Monotony. While this principle makes a lot of sense for reason-based detachment, it makes little sense for nonmonotonic inference.

Of course, this particular representation does not indicate that there is no other way to represent reason-based detachment in existing logics of nonmonotonic entailment. It does, however, suggest that reason-based entailment is a notion that should be analyzed from first principles. Furthermore, there is an additional drawback to representing balancing as logical consequence relations: it assumes the strong notion of completeness, or Universal Domain.

8 Future work

As future work, we plan to take the framework set up here in a number of different directions. First, in addition to reasons, the philosophical literature talks about considerations which, while not being reasons, can have *indirect* effects on the normative landscape—e.g., on which actions ought to be carried out. In this context, the literature discusses in particular *conditions* (or *undercutters*) which cancel the normative effects of a reason, and *modifiers* which either amplify or attenuate the (default) magnitude (or "weightiness") of a reason.[14] What we want to do, then, is extend the framework so that we can represent these other types of considerations too and explore their effects on detachment. One thing we can do is extend our formal notion of a balancing operation by allowing some of its underlying reasons to also take the value 0. These new reasons—of the form $(x, y, 0)$—could then affect the standard ones—of the form

[14] See, e.g., [27, Sec. 2] for a nice summary.

$(x, y, +)$ and $(x, y, -)$—and thereby indirectly affect detachment.

Another direction for future research is to explore detachment systems built around numerical values. In fact, detachment systems of this sort may be closer to the way reasons and their interaction are conceptualized in the philosophical literature, where the weights of reasons are standardly taken to be comprised of a polarity and a magnitude (or "weightiness"). It certainly seems worth exploring different balancing operations that assign values to contexts by applying numerical operations to reasons. Also, this does not seem to be too far off from the ideas explored in multi-criteria decision-making—see, e.g., [13].

Yet another promising idea is to explore detachment systems that are not complete, even in the weaker sense of not satisfying the Reason Universal Domain principle: they appear to be fitting for modeling case-based reasoning of the sort that is discussed, for instance, in the context of models of precedential constraint—see, e.g., [11]. Yet another idea is to explore the detachment systems built around a richer domain of discourse: logical formulas, as opposed to abstract elements. Finally, it would be useful to extend the principle-based analysis presented here with further principles and balancing operations.

9 Conclusion

Our main goal in this paper was to set up and start exploring a (general) formal framework built around reason-based detachment. We started by introducing detachment systems, or structures in which reason-based detachment is guaranteed to be valid. After formulating some general principles that detachment systems can satisfy, we focused on a class of detachment systems—which we called *balancing operations*—that can be thought of as regimenting the informal model of the normative weight scales: we formulated further principles specific to balancing operations (Section 3), defined a handful of concrete balancing operations (Sections 4–5), and put the two together in a principle-based analysis (Section 6). We also briefly discussed the relationship between reason-based detachment and logical inference, along with the most immediate directions for future research. Ultimately, we are aiming to provide 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 proposed in the context of case-based reasoning, multi-criteria decision-making, nonmonotonic reasoning, and related disciplines. This short paper is but a first step in this direction.

Acknowledgments

Both authors acknowledge financial support from the Luxembourg National Research Fund (FNR) for the project Deontic Logic for Epistemic Rights (OPEN O20/14776480) L. van der Torre is also supported by the (Horizon 2020 funded) European Coordinated Research on Long-term Challenges in Information and Communication Sciences & Technologies ERA-NET (CHIST-ERA) grant CHIST-ERA19-XAI (G.A. INTER/CHIST/19/14589586).

References

[1] Alexander, L. and M. Moore, *Deontological Ethics*, in: E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University, 2021, Winter 2021 edition .

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

[3] Bader, R., *Conditions, modifiers and holism*, in: E. Lord and B. Maguire, editors, *Weighing Reasons*, Oxford University Press, 2016 pp. 27–55.

[4] Broome, J., "Rationality through Reasoning," Wiley Blackwell Publishing, 2013.

[5] Cole, T., *Real-world criminal law and the norm against punishing the innocent: Two cheers for thershold deontology*, in: H. Hurd, editor, *Moral Puzzles and Legal Perspectives*, Cambridge University Press, 2019 pp. 371–87.

[6] Dancy, J., "Ethics without Principles," Oxford University Press, 2004.

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

[8] Drai, D., *Reasons have no weight*, Philosophical Quarterly **68** (2018), pp. 60–76.

[9] Faroldi, F. and T. Protopopescu, *All-things-considered oughts via reasons in justification logic*, in: J. Maranhao, C. Peterson, C. Strasser and L. van der Torre, editors, *Deontic Logic and Normative Systems: 16th International Conference (DEON2023, Trois-Rivières)*, College Publications, 2023 .

[10] Hawthorne, J. and O. Magidor, *Reflections on reasons*, in: D. Star, editor, *The Oxford Handbook to Reasons and Normativity*, Oxford University Press, 2018 .

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

[12] Horty, J., "Reasons as Defaults," Oxford University Press, 2012.

[13] Keeney, R. and H. Raiffa, "Decisions with Multiple Objectives: Preferences and Value," Cambridge University Press, 1993.

[14] Lord, E. and B. Maguire, *An opinionated guide to the weight of reasons*, in: E. Lord and B. Maguire, editors, *Weighing Reasons*, Oxford University Press, 2016 pp. 3–24.

[15] Maguire, B. and J. Snedegar, *Normative metaphysics for accountants*, Philosophical Studies **178** (2018), pp. 363–84.

[16] Makinson, D. and L. van der Torre, *Input/output logics*, Journal of Philosophical Logic **29** (2000), pp. 383–408.

[17] Makinson, D. and L. van der Torre, *Constraints for input/output logics*, Journal of Philosophical Logic **30** (2001), pp. 155–85.

[18] Moore, M., *The rationality of threshold deontology*, in: H. Hurd, editor, *Moral Puzzles and Legal Perspectives*, Cambridge University Press, 2019 pp. 371–87.

[19] Nair, S., *How do reasons accrue?*, in: E. Lord and B. Maguire, editors, *Weighing Reasons*, Oxford University Press, 2016 pp. 56–73.

[20] Parent, X. and L. van der Torre, *Input/output logic*, in: L. van der Torre, D. Gabbay, J. Horty and R. van der Meyden, editors, *Handbook of Deontic Logic, vol. 1*, College Publications .

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

[22] Prakken, H. and G. Sartor, *Modelling reasoning with precedents in a formal dialogue game*, Artificial Intelligence and Law **6** (1998), pp. 231–87.

[23] Raz, J., "Practical Reason and Norms," Oxford University Press, 1990.

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

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

[26] Snedegar, J., *Reasons for and reasons against*, Philosophical Studies **175** (2018), pp. 725–43.

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