A Principle-based Analysis for Numerical Balancing

Aleks Knoks[a,b] Muyun Shao[c] Leendert van der Torre[b,c]
Vincent de Wit[b] Liuwen Yu[b]

[a] *University of Luxembourg*
Institute of Philosophy
Esch-sur-Alzette, Luxembourg

[b] *University of Luxembourg*
Department of Computer Science
Esch-sur-Alzette, Luxembourg

[c] *Zhejiang University*
School of Philosophy
Hangzhou, China

Abstract

The more recent philosophical literature concerned with foundational questions about normativity often appeals to the notion of normative reasons, or considerations that count in favor or against actions, and their interaction. The interaction between reasons is standardly conceived of in terms of weighing reasons on (normative) weight scales. Knoks and van der Torre [8] have recently proposed a formal framework that allows one to think about the interaction between reasons as a kind of inference pattern. This paper extends that framework by introducing and exploring what we call *numerical balancing operators*. These operators represent the weights or magnitudes of reasons by means of numbers, and they are particularly well-suited for capturing the intuition of aggregating and weighing reasons. We define a number of concrete classes of balancing operators and explore them using a principle-based analysis.

Keywords: reasons, weighing, detachment, principle-based analysis.

1 Introduction

The notion of *normative reasons* has been playing an increasingly important role in the philosophical literature tackling foundational questions about normativity. In the practical domain, normative reasons are standardly understood to be facts that speak in favor of or against actions. [1] Thus, the fact that you have made a promise to a friend is a reason that speaks in favor of your keeping the promise, and the fact that throwing this punch would result in harming

[1] The locus classicus is Scanlon [13, p. 17]. See also [11], [12], [19], among many others.

someone is a reason that speaks against throwing the punch. The interaction between normative reasons is standardly taken to determine the deontic statuses of actions—whether they are permissible, obligatory, or forbidden—and this interaction itself is usually made sense of by analogy with weight scales. [2] On the simplest construction, these weight scales work roughly as follows. The reasons that speak in favor of some action φ (positive reasons) go in one pan of the scales, while those that speak against φ (negative reasons) go in the other. If the overall weight (or magnitude) of reasons in the first pan is greater than the overall weight of reasons in the second, φ is obligatory. If the overall weight of reasons in the second pan is greater, φ is forbidden. If the pans are equally balanced, φ is optional, that is, both φ and not-φ are permissible. [3]

While most of the work theorizing about the interaction between normative reasons and their relation to the overall deontic statuses of actions has been carried out informally, there are some exceptions. One such is a recent paper of Knoks and van der Torre [8]. [4] Our main goal in this paper is to adjust the approach of Knoks and van der Torre and apply it to (richer) structures in which reasons are associated with numerical weights and deontic statuses are assigned to actions on the basis of these weights—with this, the approach is steered closer to the way the interaction between reasons is conceived of in the informal (philosophical) literature. To reach our goal, we introduce the formal notion of *numerical balancing operators*, formulate some concrete classes of such operators, and carry out a principle-based analysis of them. The results we present in this paper show that adding numerical weights to the picture makes a huge difference: some of the core principles formulated in [8] no longer hold in general, and new principles need to be formulated to distinguish the operators.

The rest of this paper is structured as follows. In Section 2, we recall some basic notions from Knoks and van der Torre [8]. In Section 3, we extend the framework with numerical weights and introduce the core notion of numerical balancing operators. In Section 4, we introduce six concrete classes of balancing operators, and in Section 5, we present our principle-based analysis. Section 6 clarifies the relationship between the results we present here and the more general framework of Knoks and van der Torre [8]. Finally, Section 7 concludes and hints at some ideas for future research.

2 Preliminaries

In this section, we recall some definitions from [8]. The two basic building blocks in that paper are an infinite set $\mathcal{A}$ and an abstract set of values $\mathcal{V}$. Given that we will be interested in what Knoks and van der Torre call *balancing operations*,

[2] See, for instance, [1], [2], [9], [15], [17], [18].

[3] For the most careful (informal) analysis of the weight scales metaphor, see [18], for a good introduction, see [9].

[4] Other notable exceptions include Horty's [5], [6] default logic-based framework and the recent approaches that draw on decision and probabily theory [3], [10], [14].

we will work with a concrete set of values, namely, $\{+, -, 0\}$. Our formal notion of a *reason* is then defined thus:

Definition 2.1 [Reasons] Let $\mathcal{A}$ be an infinite set, called the *universe of discourse* and let $\mathcal{V}$ be the set $\{+, -, 0\}$, called *values*. A reason r is a triple of the from (x, y, v) where x and y are elements of $\mathcal{A}$ and $v \in \{+, -\}$ is the value associated with a reason, also called the *polarity* of r. [5]

The next important notion is that of a *context*:

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 is an element of $\mathcal{A}$, called the *issue*.

Contexts are meant to represent particular scenarios or situations. Each context can be thought of as asking a question about some action—this is why we call y an *issue*: is it the case that y ought to be taken, that y ought not to be taken, or that it is permissible to take y and also not to take it (that is, y is optional)? The set of reasons R of a context, in turn, is comprised of the considerations that are relevant for answering this question. We use $\mathcal{U}$ to denote the set of all possible contexts, that is, the set of contexts that can be constructed by Definitions 2.1–2.2.

Formally, balancing operations are functional relations between contexts and values. Intuitively, they can be thought of as answers to questions posed by contexts. If the context (R, y) is assigned a $+$, then y ought to be taken. If it is assigned a $-$, then y ought not to be taken. And if it is assigned a 0, then y is optional.

Definition 2.3 [Balancing operations] A *balancing operation*, denoted by $\mathcal{B}$, is a functional relation between contexts and values, that is, $\mathcal{B} \subseteq \mathcal{U} \times \mathcal{V}$ such that, for any (c, v), $(c', v') \in \mathcal{B}$, if $c = c'$, then $v = v'$.

With Definition 2.3 on the table, we are in a position to formulate principles that balancing operations might satisfy. Before we recall some important principles from [8], however, let us introduce some useful 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 $action(r) = y$ 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 $pos(R, y) = \{r \in R : r = (x, y, +)\}$; the set of reasons from R that *speak against* y is the set $neg(R, y) = \{r \in R : r = (x, y, -)\}$; and the set of reasons from R that are relevant to y is the set $R_y = pos(R, y) \cup neg(R, y)$.
- When talking about sets of contexts, we can distinguish between the set of all possible contexts, denoted by $\mathcal{U}$, and the set of contexts under consideration,

[5] The reader familiar with the philosophical literature on reasons may notice that our technical concept corresponds to what is often called the *reason relation*.

denoted by $\mathcal{C}$. The latter is the set of contexts for which the balancing operation that we are discussing at a given point is defined.

While Knoks and van der Torre formulate a handful of principles, here we recall the two that, they claim, are particularly basic, intuitive, and important because they formalize properties that seem to be inherent in the metaphor of weighing reasons on scales. These principles are Relevance and Neutrality. The intuitive idea behind Relevance is that the values assigned to an issue y within a context must be based only on the reasons that are directly related to y, and, thus, that reasons that are not related to y can be removed from the context without affecting the result.

Principle 2.4 (Relevance) *A balancing operation $\mathcal{B}$ satisfies* Relevance *just in case if $((R, y), v) \in \mathcal{B}$ and $((R_y, y), v') \in \mathcal{B}$, then $v = v'$.*[6]

Turning to Neutrality, it is meant to capture the intuition that the values $+$ and $-$ should be treated equally: if we switch the polarities of all reasons in a given context, then the value that is assigned to the context should also switch.

Principle 2.5 (Neutrality) *Given a set of reasons R, let $R' = \{(x, y, \overline{v}) : (x, y, v) \in R\}$. A balancing operation $\mathcal{B}$ satisfies* Neutrality *just in case if $((R, y), v) \in \mathcal{B}$ and $((R', y), v') \in \mathcal{B}$, then $v' = \overline{v}$.*

3 Numerical balancing operators

An important part of the intuitive picture of weighing reasons on scales is that one reason can have more weight than another, and that the weights of multiple reasons can add up. The formal notion of balancing operations does not allow us to represent this idea explicitly. The main goal of this section then is to formulate an analogous notion—that of (numerical) *balancing operators*—that will allow us to do that.

As a first step, we introduce the notion of a *weight system*.

Definition 3.1 [Weight systems] Let $\mathcal{C}$ be a set of contexts. The set of reason-context pairs of $\mathcal{C}$, written as $\mathcal{X}_\mathcal{C}$, is the set $\{(r, (R, y)) : r \in R, (R, y) \in \mathcal{C}\}$. Then a *weight system for $\mathcal{C}$*, written as $w_\mathcal{C}$, is a pair (W, f_w) where $W \subseteq \mathbb{R}^+$ is a set of weights and $f_w : \mathcal{X}_\mathcal{C} \to W$ is a total function.

It is natural to wonder about the effects of context shifts on the weights of reasons, or to ask whether any given reason has to have the same weight in every context. The positions that have been explored in the philosophical literature here range from extreme *atomist views*, on which any given reason's

---

[6] Notice that, in general, a balancing operation $\mathcal{B}$ can be such that $((R, y), v) \in \mathcal{B}$, while $((R_y, y), v') \notin \mathcal{B}$. It's not difficult to define a constraint that rules out this possibility. We can think of it as a variation on Relevance. **Principle (Relevance$'$)**: A balancing operation $\mathcal{B}$ satisfies *Relevance$'$* just in case if $((R, y), v) \in \mathcal{B}$, then there exists some value v' such that $((R_y, y), v') \in \mathcal{B}$. It shouldn't be difficult to see that Relevance and Relevance$'$ entail the following stronger principle. **Principle (Relevance$''$)**: A balancing operation $\mathcal{B}$ satisfies *Relevance$''$*, $\texttt{Re}''$, just in case if $((R, y), v) \in \mathcal{B}$, then $((R_y, y), v) \in \mathcal{B}$.

weight and polarity are context-independent, to extreme *holist views*, on which a reason's weight and its polarity can both change from context to context. Since ours is a general and formal exploration, we do not want to commit to any particular view here. However, we also want to be able to express any view lying on the atomism-holism spectrum formally. While the above definition allows reasons to be associated with different weights in different contexts—naturally inviting a holist picture—we can impose further constraints on weight systems to express views that are closer to the atomist side of the spectrum. Thus, our next definition captures one of the core tenets of atomism: that the weights of reasons are context-independent.

Definition 3.2 [Fixed weight systems] Let $\mathcal{C}$ be a set of contexts and $w_{\mathcal{C}} = (W, f_w)$ a weight system for $\mathcal{C}$. Then $w_{\mathcal{C}}$ is called a *fixed weight system* just in case, for any reason r and any pair of contexts c, $c' \in \mathcal{C}$, we have $f_w(r, c) = f_w(r, c')$.

The scales metaphor has it that the weights of individual reasons with the same polarity get aggregated into a collective weight, and that the collective weights of positive and negative reasons determine the final position of scales. Since in our framework this final position corresponds to the value associated with a context, we need a bridge from contexts supplemented with weight systems to values. This bridge is provided by what we call *procedures*:

Definition 3.3 [Procedures] Let $\mathcal{U}$ be the set of all contexts and $\mathcal{W}$ the of set of all weight systems for $\mathcal{U}$. A *procedure* is a function $\mathcal{P} : \mathcal{U} \times \mathcal{W} \to \mathcal{V}$ associating contexts and weight systems with values.

Notice that procedures are independent of weight systems: we can apply the same procedure to contexts with different weight systems, or different procedures to contexts with the same weight system.

Now we have all the ingredients we need to define balancing operators (our substitute for balancing operations from [8]). These, in effect, combine weight systems and procedures:

Definition 3.4 [Balancing operators] A *balancing operator*, denoted by $\mathcal{B}_o$, is a triple ($\mathcal{C}$, $w_{\mathcal{C}}$, $\mathcal{P}$) where $\mathcal{C}$ is a set of contexts, $w_{\mathcal{C}}$ a weight system for $\mathcal{C}$, and $\mathcal{P}$ a procedure.

In the next section, we introduce several concrete (classes of) balancing operators. Before we turn to them, however, let's formulate some general principles that balancing operators can satisfy, and we start by restating Relevance and Neutrality from Section 2 as principles for balancing operators:

Principle 3.5 (Relevance) *A balancing operator $\mathcal{B}_o = (\mathcal{C}, w_{\mathcal{C}}, \mathcal{P})$ satisfies* Relevance*,* `Re`*, just in case if there are v and v' such that $\mathcal{P}((R, y), w_{\mathcal{C}}) = v$ and $\mathcal{P}((R_y, y), w_{\mathcal{C}}) = v'$, then $v = v'$.*[7]

[7] The counterpart of the stronger version of Relevance discussed in footnote 6 would run as follows. **Principle (Relevance$''$)**: A balancing operator $\mathcal{B}_o = (\mathcal{C}, w_{\mathcal{C}}, \mathcal{P})$ satisfies *Relevance$''$*, `Re''`, just in case if $\mathcal{P}((R, y), w_{\mathcal{C}}) = v$, then $\mathcal{P}((R_y, y), w_{\mathcal{C}}) = v$.

Principle 3.6 (Neutrality) *Given a set of reasons* R*, let* $R' = \{(x, y, \overline{v}) : (x, y, v) \in R\}$*. A balancing operator* $(\mathcal{C},\, w_{\mathcal{C}},\, \mathcal{P})$ *satisfies* Neutrality*,* `Ne`*, just in case if* $\mathcal{P}((R, y), w_{\mathcal{C}}) = v$ *and* $\mathcal{P}((R', y), w_{\mathcal{C}}) = v'$*, then* $v' = \overline{v}$*.*

Recall our definition of fixed weight systems. We can use it to formulate another principle or constraint on balancing operators:

Principle 3.7 (Fixed Weight) *A balancing operator* $\mathcal{B}_o = (\mathcal{C},\, w_{\mathcal{C}},\, \mathcal{P})$ *satisfies* Fixed Weight*,* `FiWe`*, just in case* $w_{\mathcal{C}}$ *is a fixed weight system.*

According to atomism, not only the weights of reasons are fixed, but also their polarity. This idea can, again, be expressed in the form of a principle:

Principle 3.8 (Fixed Polarity) *A balancing operator* $\mathcal{B}_o = (\mathcal{C},\, w_{\mathcal{C}},\, \mathcal{P})$ *satisfies* Fixed Polarity*,* `FiPo`*, just in case, for any reason* $r = (x, y, v)$*, if there is a context* $(R, y) \in \mathcal{C}$ *such that* $r \in R$*, then there is no* $(R', y) \in \mathcal{C}$ *such that* $(x, y, \overline{v}) \in R'$*.*

With these two principles, we can formulate extreme atomism as a class of balancing operators.

Definition 3.9 [Atomist balancing operators] Let $\mathcal{B}_o$ be a balancing operator. We call $\mathcal{B}_o$ *atomist* just in case $\mathcal{B}_o$ satisfies both Fixed Polarity and Fixed Weight.

And given that holism is defined in opposition to atomism, it is also straightforward to formulate.

Definition 3.10 [Holist balancing operators] Let $\mathcal{B}_o$ be a balancing operator. We call $\mathcal{B}_o$ *holist* just in case it is not *atomist*.

It's worth noting that Fixed Weight and Fixed Polarity illustrate the flexibility of the formal notion of a balancing operator: we can formulate different principles some of which have to do with weight systems, others with the structure of contexts, and yet others with procedures.

4 Some concrete balancing operators

In this section, we introduce six classes of balancing operators. The unifying element of each class is the procedure. The first three classes correspond to three simple and intuitive operations on numbers: addition, multiplication and maximum. The forth class supplements the first of these with a threshold. Finally, the ideas behind our last two operators come from the discussion of (possible) views one might have about the workings of weight scales in Tucker [16].

The first class of operators is based on simple addition. The context (R, y) gets assigned the value $+$ if the sum weight of reasons for y is strictly greater than the sum weight of reasons against y; it gets assigned $-$ if the sum weight of reasons against y is strictly greater than the sum weight of reasons for y; and it gets assigned 0 otherwise.

Definition 4.1 [Additive Balancing Operators] Let $\mathcal{B}_o = (\mathcal{C}, w_{\mathcal{C}}, \mathcal{P}^+)$ be a balancing operator. Then it is called an *Additive Balancing Operator*, `Add`, just in case:

$$\mathcal{P}^+((R,y),w_\mathcal{C}) = \begin{cases} + & \text{if } \sum_{r\in pos(R,y)} f_w(r,(R,y)) > \sum_{r\in neg(R,y)} f_w(r,(R,y)) \\ - & \text{if } \sum_{r\in pos(R,y)} f_w(r,(R,y)) < \sum_{r\in neg(R,y)} f_w(r,(R,y)) \\ 0 & \text{otherwise} \end{cases}$$

The second class of balancing operators is based on multiplication. A context gets assigned $+$ in case the product of weights of positive reasons (for y) is greater than that of negative reasons; it gets assigned $-$ in case the product of weights of negative reasons is greater than that of positive reasons; and it gets assigned 0 otherwise.

Definition 4.2 [Multiplicative Balancing Operators] Let $\mathcal{B}_o = (\mathcal{C}, w_\mathcal{C}, \mathcal{P}^\times)$ be a balancing operator. Then it is called a *Multiplicative Balancing Operator*, `Mul`, just in case :

$$\mathcal{P}^\times((R,y),w_\mathcal{C}) = \begin{cases} + & \text{if } \prod_{r\in pos(R,y)} f_w(r,(R,y)) > \prod_{r\in neg(R,y)} f_w(r,(R,y)) \\ - & \text{if } \prod_{r\in pos(R,y)} f_w(r,(R,y)) < \prod_{r\in neg(R,y)} f_w(r,(R,y)) \\ 0 & \text{otherwise} \end{cases}$$

The balancing operators belonging to the third class we discuss determine the value of context by comparing the maximal weights of positive and negative reasons. A context gets assigned $+$ if the maximal weight of positive reasons is greater than that of negative reasons; it gets assigned $-$ if the maximal weight of negative reasons is greater than that of positive reasons; and it gets assigned 0 if the weights are equal.

Definition 4.3 [Maximizing Balancing Operators] Let $\mathcal{B}_o = (\mathcal{C}, w_\mathcal{C}, \mathcal{P}^m)$ be a balancing operator. Then it is called a *Maximizing Balancing Operator*, `Max`, just in case:

$$\mathcal{P}^m((R,y),w_\mathcal{C}) = \begin{cases} + & \text{if } \mathbf{Max}(\{f_w(r,(R,y)) : r \in pos(R,y)\}) > \\ & \qquad \mathbf{Max}(\{f_w(r,(R,y)) : r \in neg(R,y)\}) \\ - & \text{if } \mathbf{Max}(\{f_w(r,(R,y)) : r \in pos(R,y)\}) < \\ & \qquad \mathbf{Max}(\{f_w(r,(R,y)) : r \in neg(R,y)\}) \\ 0 & \text{otherwise} \end{cases}$$

The balancing operators that belong to the fourth class work with a *threshold* on the weights of reasons. The basic idea here is that a reason can make a difference for which value gets assigned to a context only in case its weight is above a certain threshold. In the following definition, this idea is combined with the familiar operation of addition:

Definition 4.4 [(Additive) Threshold Balancing Operators] Let $\mathcal{B}_o = (\mathcal{C}, w_{\mathcal{C}}, \mathcal{P}^t)$ be a balancing operator. Then it is called an *(Additive) Threshold Balancing Operator*, `(Add)Thr`, just in case:

$$\mathcal{P}^t((R,y),w_{\mathcal{C}}) = \begin{cases} + & \text{if } \sum_{r\in pos(R,y)\wedge f_w(r,(R,y))>t} f_w(r,(R,y)) > \\ & \qquad \sum_{r\in neg(R,y)\wedge f_w(r,(R,y))>t} f_w(r,(R,y)) \\ - & \text{if } \sum_{r\in pos(R,y)\wedge f_w(r,(R,y))>t} f_w(r,(R,y)) < \\ & \qquad \sum_{r\in neg(R,y)\wedge f_w(r,(R,y))>t} f_w(r,(R,y)) \\ 0 & \text{otherwise} \end{cases}$$

It's worth emphasizing that a threshold is not an operation, but, rather, a gatekeeping device that precludes reasons with (relatively) low weights from having any effect on the value assigned to a context. Definition 4.4 adds a threshold to addition. It should be clear that the operations of multiplication and taking the maximum that we used to define balancing operators above can also be supplemented with a threshold.

Now we turn to the final two classes of balancing operators. Both of these are inspired by the discussion in Tucker [16], who works in an informal setting and formulates the counterparts of our balancing operators in terms of *permission*. Since we have been working with obligations above, we re-state Tucker's ideas in terms of obligations.

The first of these two classes formalizes what Tucker calls *relative weight satisficing*: φ is permissible just in case the reasons against φ are no more than twice as weighty as the reasons for φ.[8] Restating this idea in terms of obligations, we get the following: φ is obligatory just in case the reasons against φ are *at most* twice as weighty as the reasons for φ *and* the reasons for φ are (strictly) *more than* twice as weighty as the reasons against φ. Since the second conjunct entails the first, we can simplify: φ is obligatory just in case the reasons for φ are *more than* twice as weighty as the reasons against φ. The formal definition, then, runs as follows:

Definition 4.5 [Relative Weight Satisficing Operators] Let $\mathcal{B}_o = (\mathcal{C}, w_{\mathcal{C}}, \mathcal{P}^R)$ be a balancing operator. Then it is called a *Relative Weight Satisficing Operator*, `RelSat`, just in case:

$$\mathcal{P}^R((R,y),w_{\mathcal{C}}) = \begin{cases} + & \text{if } \sum_{r\in pos(R,y)} f_w(r,(R,y)) > 2\sum_{r\in neg(R,y)} f_w(r,(R,y)) \\ - & \text{if } 2\sum_{r\in pos(R,y)} f_w(r,(R,y)) < \sum_{r\in neg(R,y)} f_w(r,(R,y)) \\ 0 & \text{otherwise} \end{cases}$$

Our final class of balancing operators corresponds to what Tucker calls *absolute weight satisficing*. (This view is meant to be in tension with the idea

[8] See [16, p. 373ff].

of weighing reasons on weight scales.) According to absolute weight satisficing, φ is permissible if the reasons for φ have a weight of at least 100 (no matter how much weight the reasons against φ have), and it is not permissible otherwise. [9] Notice that it is straightforward to define a similar sort of operator—that is, an operator that is sensitive to positive reasons *only*—in terms of obligations: φ is obligatory if the reasons for φ have a weight of at least 100, and it is forbidden (impermissible) otherwise. The formal definition then runs thus:

Definition 4.6 [Absolute Weight Satisficing Operators] Let $\mathcal{B}_o = (\mathcal{C}, w_\mathcal{C}, \mathcal{P}^A)$ be a balancing operator. Then it is called an *Absolute Weight Satisficing Operator*, `AbsSat`, just in case:

$$\mathcal{P}^A((R,y),w_\mathcal{C}) = \begin{cases} + & \text{if } \sum_{r\in pos(R,y)} f_w(r,(R,y)) > 100 \\ - & \text{otherwise} \end{cases}$$

Perhaps, one note about the final two class of operators is in order before we leave this section: we followed Tucker in setting the threshold at 100 in $\mathcal{P}^A$, as well as in requiring that the reasons against φ cannot be more than *two times* as weighty as the reasons for φ to be permissible in $\mathcal{P}^R$. We could define versions of these operators using other numbers.

5 Principle-based analysis

In this section, we formulate four principles and use them to compare the balancing operators defined in Section 4. We start with the formulation of the principles—the first comes from Knoks and van der Torre [8]; the latter three are new. Then we turn to a discussion, of our results and some complications.

The first principle is *Polarity Monotony.* It says that, if a balancing operator assigns + to a context and a positive reason is added, then the operator will still assign + to the context; and similarly, if the operator assigns − to a context and a negative reason is added, then it will still assign − to the context.

Principle 5.1 (Polarity Monotony) *A balancing operator $\mathcal{B}_o = (\mathcal{C}, w_\mathcal{C}, \mathcal{P})$ satisfies* Polarity Monotony*,* `PoMn`*, just in case, for all $\mathcal{P}((R,y),w_\mathcal{C}) = v$ where $v \neq 0$, if $(R \cup \{(x,y,v)\}, y) \in \mathcal{C}$, then $\mathcal{P}((R \cup \{(x,y,v)\}, y), w_\mathcal{C}) = v$.*

Our second principle is called *Commensurate Removal.* It says that for every context, if we remove a pair of opposite reasons with the same weight, then the value assigned to the context doesn't change.

Principle 5.2 (Commensurate Removal) *A balancing operator $\mathcal{B}_o = (\mathcal{C}, w_\mathcal{C}, \mathcal{P})$ satisfies* Commensurate Removal*,* `CoRe`*, just in case, if $\mathcal{P}((R,y),w_\mathcal{C}) = v$, then for each pair of reasons $r, r' \in R$ such that $polarity(r) = \overline{polarity(r')}$ and $f_w(r,(R,y)) = f_w(r',(R,y))$, we have $\mathcal{P}((R \backslash \{r,r'\}, y), w_\mathcal{C}) = v$.*

Our next principle is called *Sensitivity.* It says that, for every equally-balanced context—that is, every context to which 0 is assigned—adding a new

---

[9] See [16, p. 378ff].

reason will cause the new context to be assigned a value that equals the polarity of that reason.

Principle 5.3 (Sensitivity) *A balancing operator $\mathcal{B}_o = (\mathcal{C}, w_{\mathcal{C}}, \mathcal{P})$ satisfies* Sensitivity*, $\mathtt{Se}$, just in case, if $\mathcal{P}((R, y), w_{\mathcal{C}}) = 0$ and there is a v such that $\mathcal{P}((R \cup \{r\}, y), w_{\mathcal{C}}) = v$, then $v = polarity(r)$.*

Finally, our final principle is *Union Monotony.* It says that if a balancing operator assigns the value v to two contexts, then it will also assign v to the union of these contexts.

Principle 5.4 (Union Monotony) *A balancing operator $\mathcal{B}_o = (\mathcal{C}, w_{\mathcal{C}}, \mathcal{P})$ satisfies* Union Monotony*, $\mathtt{UnMn}$, just in case, if $\mathcal{P}((R_1, y), w_{\mathcal{C}}) = v$, $\mathcal{P}((R_2, y), w_{\mathcal{C}}) = v$, and $(R_1 \cup R_2, y) \in \mathcal{C}$, then $\mathcal{P}((R_1 \cup R_2, y), w_{\mathcal{C}}) = v$.*

Now that we have the principles, they can be used to analyze and compare the operators. However, there is a complication: the framework that we have set up is so unconstrained that, in the general case, (almost) none of the principles are satisfied by any of the operators.[10] This has to do, in particular, with the fact that our formal notion of a weight system (Definition 3.1) allows for unconstrained change of reasons' weights from one context to another. But let's recall our (brief) discussion of atomism and holism from Section 3 here. Atomists say that reasons weights and polarities are the same in all contexts, whereas extreme holists say that the weights of the same reason in two contexts can be wildly different. We wanted to be in a position to express all sorts of views lying on the atomism-holism spectrum in our framework, and, without imposing further constraints, it effectively imposes an extreme holist picture. On reflection, it should be no surprise that the balancing operators from Section 4 do not satisfy any of the principles *if* extreme holism is at work in the background.

What we present below then is a principle-based analysis of those balancing operators from Section 4 which also satisfy Fixed Weight, that is, we restrict attention to balancing operators with fixed weight systems.

Proving that a given operator does (or does not) satisfy some principle is more tedious than difficult. Here are two sample proofs:

Proposition 5.5 *Relative weight satisficing (Definition 4.5) with Fixed Weight (Principle 3.7) does not satisfy Sensitivity (Principle 5.3).*

Proof. Consider a relative weight satisficing operator $(\mathcal{C},\ w_{\mathcal{C}},\ \mathcal{P}^R)$ where $\mathcal{C} = \{c_1, c_2\}$; $c_1 = (\{r_1, r_2\}, y_1)$, $c_2 = (\{r_1, r_2, r_3\}, y_1)$; $r_1 = (x_1, y_1, +)$, $r_2 = (x_2, y_1, -)$, $r_3 = (x_3, y_1, +)$; $f_w(r_1, c_1) = f_w(r_2, c_1) = 5$, and $f_w(r_3, c_2) = 0.5$. Notice that $\sum_{r \in pos(\{r_1,r_2\},y)} f_w(r, c_1) = f_w(r_1, c_1) = 5$, and that $\sum_{r \in neg(\{r_1,r_2\},y)} f_w(r, c_1) = f_w(r_2, c_1) = 5$. From this and Definition 4.5, we get $\mathcal{P}^R(c_1, w_{\mathcal{C}}) = 0$. For Sensitivity to be satisfied, we would have to have $\mathcal{P}^R(c_1 \cup \{r\}, w_{\mathcal{C}}) = +$ for every context $c_1 \cup \{r\}$ where $r = (x, y, +)$. Notice that $\sum_{r \in pos(\{r_1,r_2,r_3\},y)} f_w(r, c_2) = f_w(r_1, c_2) + f_2(r_3, c_2) = 5 + 0.5 = 5.5$, and

[10] The only exception is Absolute Weight Satisficing which (vacuously) satisfies Sensitivity.

	`Add`	`Mul`	`Max`	`(Add)Thr`	`RelSat`	`AbsSat`
3.5 `Re`	✓	✓	✓	✓	✓	✓
3.6 `Ne`	-	-	-	-	-	-
5.1 `PoMn`	✓	-	✓	✓	✓	✓
5.2 `CoRe`	✓	✓	-	✓	-	-
5.3 `Se`	✓	-	-	-	-	✓
5.4 `UnMn`	-	-	✓	-	-	-

Table 1
Summary of the principle-based analysis, assuming Fixed Weight

that $\sum_{r \in neg(\{r_1,r_2,r_3\},y)} f_w(r, c_2) = f_w(r_2, c_2) = 5$. From this and Definition 4.5, we have $\mathcal{P}^R(c_2, w_\mathcal{C}) = 0$. □

Proposition 5.6 *Maximizing balancing (Definition 4.3) with Fixed Weight (Principle 3.7) satisfies Polarity Monotony (Principle 5.1).*

Proof. Let $\mathcal{B}_o = (\mathcal{C}, w_\mathcal{C}, \mathcal{P}^m)$ be a maximizing balancing operator with a fixed weight system. Consider an arbitrary context $c = (R, y) \in \mathcal{C}$ such that $\mathcal{P}^m((R, y), w_\mathcal{C}) = v$ and $v \neq 0$. Assume that there is some reason $r' = (x, y, v)$ and a context $(R \cup \{r'\}, y) \in \mathcal{C}$. To establish that Polarity Monotony is satisfied, we need to show that $\mathcal{P}^m((R \cup \{r'\}, y), w_\mathcal{C}) = v$. Without loss of generality, we assume that $v = +$. From $\mathcal{P}^m((R, y), w_\mathcal{C}) = v$ and Definition 4.3, we know that $\max(\{f_w(r, (R, y)) \mid r \in pos(R, y)\}) = P > N = \max(\{f_w(r, (R, y)) \mid r \in neg(R, y)\})$. Now notice that $f_w(r', (R \cup \{r'\}, y)) > 0$, and that in $(R \cup \{r'\}, y)$ reasons have the same weights that they had in (R, y). From here, $\max(\{f_w(x, (R \cup \{r'\}, y)) \mid r \in pos(R, y) \cup \{r'\}\}) = \max(P, f_w(r', (R \cup \{r'\}, y))) \geq P > N = \max(\{f_w(r, (R, y)) \mid r \in neg(R, y)\})$. And this is enough to conclude that $\mathcal{P}^m((R \cup \{r\}, y), w_\mathcal{C}) = +$.

□

The proofs of other propositions—that is, the propositions that show which of the remaining operators do (or do not) satisfy which propositions—are equally straightforward. We leave them for a technical report and let Table 1 summarize the results that they establish: the topmost row lists the balancing operators; the leftmost column lists the principles; the remaining cells state whether the given operator does (✓) or doesn't (−) satisfy the given principle. For example, the third column makes it clear that the class of multiplicative operators (this is what `Mul` stand for) satisfy only two principles, namely, Relevance (`Rel`) and Commensurate Removal (`CoRe`).

It may be surprising to see that none of the operators satisfy Neutrality. Recall that Knoks and van der Torre [8] thought that both Relevance and Neutrality formalize properties that seem to be inherent in the metaphor of weighing reasons on scales. It turns out that the operators do not, in general, satisfy Neutrality with the assumption of Fixed Weight for the same reason that they do not, in general, satisfy all other principles without the assumption of Fixed Weight: nothing in the definition of fixed weight systems precludes them from assigning $(x, y, +)$ and $(x, y, -)$ different weights in different contexts.

We can, of course, define a notion in the vicinity of fixed weight systems that makes this impossible.

Definition 5.7 [Symmetric weight systems] Let $\mathcal{C}$ be a set of contexts and $w_{\mathcal{C}} = (W, f_w)$ a weight system for $\mathcal{C}$. Then $w_{\mathcal{C}}$ is called a *symmetric weight system* just in case, for any pair of reasons $r = (x, y, +)$, $r' = (x, y, -)$ and any pair of contexts c, $c' \in \mathcal{C}$, we have $f_w(r, c) = f_w(r', c')$.

The counterpart of Fixed Weight then runs thus:

Principle 5.8 (Symmetry) *A balancing operator* $\mathcal{B}_o = (\mathcal{C}, w_{\mathcal{C}}, \mathcal{P})$ *satisfies* Symmetry, ***Sym***, *just in case (i)* $(R, y) \in \mathcal{C}$ *if and only if* $(\{(x, y, \overline{v}) : (x, y, v) \in R\}, y) \in \mathcal{C}$ *and (ii)* $w_{\mathcal{C}}$ *is a symmetric weight system.*

It is not difficult to verify that Symmetry entails Fixed Weight. What's more, it turns out that, with Symmetry in the background, every balancing operator from Section 4 satisfies Neutrality. Here is a sample proof.

Proposition 5.9 *Additive balancing (Definition 4.1) with Symmetry (Principle 5.8) satisfies Neutrality (Principle 3.6).*

Proof. Consider some additive operator $\mathcal{B}_o = (\mathcal{C}, w_{\mathcal{C}}, \mathcal{P}^+)$ that satisfies Symmetry. Consider an arbitrary context $c = (R, y)$ for which we have $\mathcal{P}^+((R, y), w_{\mathcal{C}}) = v$. Assume that $(R', y) \in \mathcal{C}$ where $R' = \{(x, y, \overline{v})$. Without loss of generality, suppose that $v = +$. Then we know that $\sum_{r \in pos(R,y)} f_w(r, (R, y)) = P > N = \sum_{r \in neg(R,y)} f_w(r, (R, y))$. Since $w_{\mathcal{C}}$ is symmetric, we know that $f_w((x, y, \overline{v}), (R', y)) = f_w((x, y, v), (R, y))$ for every $(x, y, \overline{v}) \in R'$. As a consequence, $\sum_{r \in pos(R',y)} f_w(r, (R', y)) = N < P = \sum_{r \in neg(R',y)} f_w(r, (R', y))$, and, thus, $\mathcal{P}^+((R, y), w_{\mathcal{C}}) = -$. □

6 Related work

In this section, we relate our extension of Knoks and van der Torre's framework to the original. First of all, it's worth emphasizing that the original framework with its notion of a detachment systems is more general. For instance, it does not assume that the relation between contexts and values is functional, nor makes any restrictions on the shape of the set of values $\mathcal{V}$. Our notion of balancing operators is grounded in some conceptual choices and so it is more specific. Nevertheless, because of those conceptual choices it is also better suited to capture the informal model of weighing reasons on weight scales from the philosophical literature.

Knoks and van der Torre discuss two different classes of balancing *operations* (which qualify as specific types of detachment systems): what they call *anonymous* and *relational balancing operations*. It wouldn't be difficult to restate all the particular anonymous operations that they define as balancing operators. [11] For instance, Knoks and van der Torre's *Simple Counting* assigns a value to a context of the form (R, y) by comparing the number of positive and negative

[11] This, again, speaks to the flexibility of the formal notion of a balancing operator.

y-reasons in R. In the present framework, Simple Counting corresponds to a special case of Additive Balancing, namely, one the underlying weight system of which assigns the same weight to all reasons. (We leave the proof for the journal version of this paper.) Other anonymous balancing operations are straightforward to redefine as operators. Knoks and van der Torre formulated several principles that were satisfied by all of their anonymous balancing operations, but that do not hold for every operator we defined in Section 4. This includes Relevance and Polarity Monotony. As we saw, Relevance does not hold for all of these operators *unless* Fixed Weight is also assumed, and Polarity Monotony does not hold for Multiplicative Balancing *even* if Fixed Weight is assumed. These observations illustrate that the notion of a balancing operator gives us a grip on richer structures.

Turning to relational balancing operations, these are more difficult to relate to balancing operators, since relational operations come equipped with a relation over reasons. It turns out to be possible to establish a connection between Maximizing Balancing Operators and one particular relational operation: what Knoks and van der Torre call *Decisive Reason*. This operation assigns a value to a context by checking the polarity of the "strongest" reason in it. It shouldn't be difficult to see that the "stronger than" relation can be mapped to the greater than relation of numerical weights, and that, with this mapping, Decisive Reasons has the form of Maximizing Balancing operators. (Again, we leave the proof of this for the journal version.) To be in a position to explore the connections between relational and numerical balancing, it would pay to extend the notion of a balancing operator with a further component, namely, a binary anti-symmetric relation over reasons (in contexts). With this, we would be in a position to formulate principles that have to do with the relation—in addition to principles that have to do with weight systems and procedures—and have a general framework for analyzing anonymous, numerical, and rational balancing, as well as the connections between them.

7 Conclusion and future work

In this paper, we extended Knoks and van der Torre's framework [8] to richer structures in which reasons are associated with numerical weights. We started by introducing the formal notions of weight systems, procedures, and balancing operators. Then we introduce six concrete classes of balancing operators, presented a principle-based analysis of them, and explained how the results presented here go beyond those of [8].

For future work, we plan to set up and explore the more general framework we mentioned at the end of previous section, that is, a framework that would unify numerical and relational balancing. It seems to be clear that what we have done above shows that there is a rich variety of balancing operators available for exploration and formal analysis. We also plan to explore how our numerical balancing framework relates to multi-criteria decision-making [7] and qualitative bipolar decision-making [4], as well as how it might be used to model case-based reasoning.

References

[1] Berker, S., *Particular reasons*, Ethics **118** (2007), pp. 109–139.

[2] Broome, J., *Reasons*, in: R. J. Wallace, P. Petti, S. Scheffler and M. Smith, editors, *Reasons and Value: Themes from the Moral Philosophy of Joseph Raz*, Oxford: Clarendon Press, 2004 pp. 28–55.

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

[4] Dubois, D., G. Fargier and J.-L. Bonnefon, *On the qualitative comparison of decisions having positive and negative features*, Journal of Artificial Intelligence Research **32** (2008), pp. 385–417.

[5] Horty, J., *Reasons as defaults*, Philosophers' Imprint **7** (2007), pp. 1–28.

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

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

[8] Knoks, A. and L. van der Torre, *Reason-based detachment*, in: B. Bentzen, B. Liao, D. Liga, R. Markovich, B. Wei, M. Xiong and T. Xu, editors, *Joint Proceedings of the 3rd International Workshop on Logics for New-Generation Artificial Intelligence and the International Workshop on Logic, AI, and Law (LNGAI/LAIL2023, Hangzhou)* (2023), pp. 49–65.

[9] 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.

[10] Nair, S., *"Adding up" reasons: Lessons for reductive and non-reductive approaches*, Ethics **132** (2021), pp. 38–88.

[11] Parfit, D., "On What Matters (Volume I)," Oxford University Press, 2011.

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

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

[14] Sher, I., *Comparative value and the weight of reasons*, Economics and philosophy **35** (2019), pp. 103–158.

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

[16] Tucker, C., *The dual scale model of weighing reasons*, Noûs **56** (2022), pp. 366–92.

[17] Tucker, C., *A holist balance scale*, Journal of the American Philosophical Association **9** (2023), pp. 533–53.

[18] Tucker, C., "The Weight of Reasons: A Framework for Ethics," Oxford University Press, forthcoming.

[19] Whiting, D., "The Range of Reasons: in Ethics and Epistemology," Oxford University Press, 2021.