Reasons in Weighted Argumentation Graphs

David Streit(✉), Vincent de Wit, and Aleks Knoks

Department of Computer Science, University of Luxembourg, Esch-Sur-Alzette, Luxembourg
{david.streit,vincent.dewit,aleks.knoks}@uni.lu

Abstract. The philosophical literature that tackles foundational questions about normativity often appeals to normative reasons—or considerations that count in favor of or against actions—and their interaction. The interaction between normative reasons is usually made sense of by appealing to the metaphor of (normative) weight scales. This paper substitutes an argumentation-theoretic model for this metaphor. The upshot is a general and precise model that is faithful to the philosophical ideas.

Keywords: Argumentation theory · Normative reasons · Weighing

1 Introduction

Philosophers who explore normative matters often appeal to *normative (practical) reasons*, understanding them as considerations that count in favor of or against actions. When discussing the interaction between reasons, they often use such phrases as "the action supported on the balance of reasons" and "reasons in favor outweigh the reasons against", inviting an image of *weight scales for reasons*. The simplest model of these (normative) weight scales works, roughly, as follows. Reasons speaking in favor of φ-ing go in one pan of the scales, and reasons against go in the other. If the weight of the reasons in the first pan is greater, 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.

While philosophers have explored various ideas about the exact workings of the weight scales and also looked at some alternatives, their investigations have mostly been carried out in informal terms. The goal of this paper is to develop a *formal* model of (normative) weight scales, drawing on formal argumentation. Instead of starting from scratch, we repurpose Gordon and Walton's model of "balancing arguments" [4].

This paper is structured as follows. Section 2 sketches the philosophical ideas on weighing reasons. Sections 3 and 4 set up the model. Section 5 discusses our main results and some of the work that is most closely related to ours.

N. Alechina et al. (Eds.): LORI 2023, LNCS 14329, pp. 251–259, 2023.
https://doi.org/10.1007/978-3-031-45558-2_19

2 (Normative) Weight Scales

This section provides a bird's-eye view summary of the main ideas from the philosophical literature on weighing reasons. Note that it is an opinionated sketch: we simplify where possible and bracket a whole plethora of important and complex questions. (For more thorough overviews, see, e.g., [6] and [9, pp. 1–7].)

Normative reasons are typically taken to be *facts* that are not subject to debate. Thus, the fact that the person next door is in need of help is a reason for you to help them, regardless of your values and preferences, as well as your views on ethics and metaethics. Reasons are always reasons for someone: they favor or speak against someone's action. They are also intimately tied to their *weights*, which are comprised of a *magnitude* and a *polarity*. Magnitude has to do with the relative importance of the reason; polarity with whether it is a reason for or against. Reasons against an action count (either directly, or indirectly) as favoring alternative actions.

Reasons play a core role in determining the deontic statuses of actions. Thus, whether some action is permitted/required/ought to be taken depends on the reasons that count for/against it and their interaction. In staying with the weight scales metaphor, we say that one is permitted to φ just in case the net weight of the reasons for φ-ing is at least as high as the net weight of the reasons for the alternatives. (For a discussion of subtle changes one could make to this definition, see, e.g., [9].)

An important and hotly debated question concerns the effects of context on the weights of reasons. Positions range from extreme *atomist* views on which a reason's weight is context-independent to extreme *holist* views on which a fact that is a reason for φ-ing in one context can be a reason against φ-ing in a different one, or cease to be a reason at all. Most philosophers find positions at both ends of the spectrum implausible, preferring views on which there is both (some) stability in reasons' weights and that allow for (some) context-sensitivity. A common move here is to appeal to what we might call *normatively-relevant considerations that aren't reasons*. Such considerations don't qualify as reasons because they don't count for/against actions. However, they can affect the weights of reasons, and so have an (indirect) effect on an action's deontic status. It's common to distinguish between two types of such considerations: *undercutters* and *modifiers*. An undercutter nullifies the weight of a reason, effectively making it cease to be a reason. Modifiers are of two types: *attenuators* and *amplifiers*. An attenuator reduces the magnitude of a reason, making it less weighty. An amplifier amplifies the magnitude of a reason, making it more weighty. Undercutters and modifiers suggest the view that every reason has a context-independent *default weight* and a context-specific *final weight*, and that any difference between the two can be accounted for by appeal to undercutters and modifiers. This view is common, and it seems up to debate whether it is closer to atomism or holism.

3 Normative Graphs

In this section, we explain how to represent the structural relations that can obtain between normative reasons, other normatively-relevant considerations, and alternative actions (or options). (In the next one, we focus on modeling the normative effects that these can exert on each other.)

As background, we assume a propositional language ($\mathcal{L}$) with the standard connectives. We use the term 'normatively-relevant consideration' (or simply 'consideration') to refer to reasons, undercutters, and modifiers. Following the philosophical literature, we assume that every consideration (a) has a default weight; (b) can be undercut; and (c) can have its weight changed by modifiers. Jointly, the relevant considerations are meant to determine the deontic status of actions available to an agent. Adapting the notion of *issue* from [4], we think of options as a finite subset of $\mathcal{L}$ representing a set of mutually exclusive and exhaustive actions available to the agent. Adapting the notion of *argument* from [4], we define considerations as follows:

Definition 1 (Normatively-relevant consideration). *A* consideration *is a tuple* $C = (p, c, u, a^-, a^+, w)$*, where the first five elements are formulas of* $\mathcal{L}$ *called, respectively,* premise*,* conclusion*,* undercutter*,* attenuator*, and* amplifier*, while the sixth element is a positive real number called* default weight*.*

We will use the following scenario to illustrate this and future definitions—note that the expressions in brackets are atomic sentences of $\mathcal{L}$:

Example 1. You are to choose between two options: to go to the movies with me ($Movies$), or to have dinner with your mom at her favorite restaurant ($Dinner$). You have made a promise to me ($Promise$). What's more, you were very insistent when making the promise: you said that you would keep it no matter what ($Insist$). Dining with mom would make her happy ($MomHappy$). The restaurant also happens to serve your favorite cake ($Cake$). Also, it is Mother's Day ($MothersDay$) and you haven't seen your mom in a while ($LongTime$).

Notice that the options here are $\{Movies, Dinner\}$. The intuitive idea that your promise is a reason to go to the movies is captured by the consideration $C_1 = (Promise, Movies, u_1, a_1^-, a_1^+, w_1)$. Similarly, the idea that going to the restaurant will make your mom happy and that the restaurant serves your favorite cake are reasons to have dinner with mom is represented by $C_2 = (MomHappy, Dinner, u_2, a_2^-, a_2^+, w_2)$ and $C_3 = (Cake, Dinner, u_3, a_3^-, a_3^+, w_3)$ respectively. The idea that your being very insistent when making the promise *amplifies* C_1 is captured by consideration $C_4 = (Insist, a_1^+, u_4, a_4^-, a_4^+, w_4)$. Notice that the conclusion of C_4 (a_1^+) corresponds to the amplifier of C_1, which means that C_4 *amplifies* C_1 or that C_4 is an amplifier of C_1. The rest of the example is captured by considerations $C_5 = (MothersDay, a_2^+, u_5, a_5^-, a_5^+, w_5)$, $C_6 = (LongTime, a_2^+, u_6, a_6^-, a_6^+, w_6)$, and $C_7 = (Release, u_1, a_7^-, a_7^+, w_7)$.

The considerations that are in force in a given context form a graph structure. This is captured by our next definition—given a consideration $C = (p, c, u, a^-, a^+, w)$, we let $p_C = p$; $c_C = c$; $u_C = u$; $a_C^- = a^-$; $a_C^+ = a^+$, and $w_C = w$:[1]

[1] We follow [4] in calling the structures specified in Definition 2 *graphs* as they can be mapped straightforwardly to directed graphs.

Definition 2 (Normative graph). *A* normative graph *is a triple of the form* $N = (S, O, R)$, *where* S *is a finite subset of* $\mathcal{L}$; $O \subseteq S$ *called the* set of options; *and* R *is a finite set of considerations, where for every* $C \in R$, p_C, c_C, u_C, a^-_C, a^+_C *are all members of* S *and* $w_C \in \mathbb{R}_{>0}$.

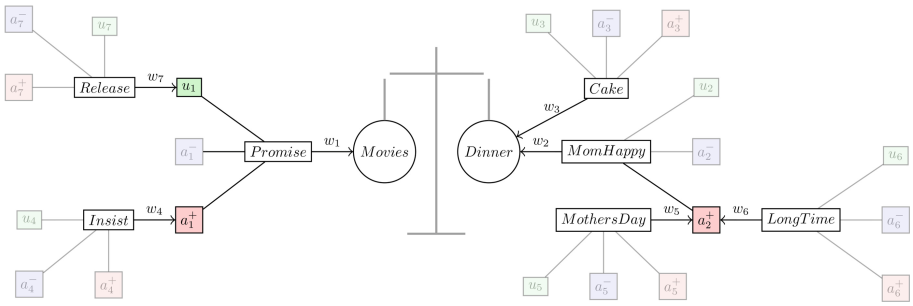

Fig. 1. Movies or Mom's favorite restaurant

Figure 1 depicts the full graph comprising our example visually. The two options are represented as circles in the middle. Each consideration is depicted as comprised of five nodes, some of which are shared between considerations. For example, the node *Insist* is the premise of $C_4 = (\mathit{Insist}, a^+_1, u_4, a^-_4, a^+_4, w_4)$, and the node a^+_1 is the conclusion of C_4 and the amplifier of C_4. The nodes u_4, a^-_4, and a^+_4 stand for the remaining elements of C_4. They are grayed out, since no other considerations affect C_4. The default weight of C_4 is presented on the edge between premise and conclusion.

Notice that the graph structure in Fig. 1 is finite, directed, and acyclic. This is no coincidence. Following the philosophical literature, we allow that graphs representing the structural relations between normatively-relevant considerations and options are complex. However, we also require that they are finite and never contain any cycles. In particular, this rules out intra-consideration circularity, where e. g. the premise and conclusion are identical.

Before moving on, we introduce some terminology. Given a graph $N = (S, O, R)$, we call a consideration $C \in R$ a *reason* (for $o \in O$) if $c_C = o$. We say that C attenuates C' if c_C is the attenuator of some consideration C', we say that it amplifies C' if c_C is an amplifier of C' and that it undercuts C' if c_C is the uncercutter of C'. Overloading the terms, we sometimes call a consideration that undercuts or modifies another reason an undercutter or modifier respectively. Sometimes we may want to add, remove, or replace considerations in a graph. While the addition and removal are straightforward to define, one needs to be careful with replacement. Given a graph $N = (S, O, R)$ and a consideration C, we let $N + C$ denote the graph that results from adding C to N and $N - C$ the graph that results from removing C from N. A graph that results from replacing C by another consideration C' in N is denoted by $N^{+C'}_{-C}$.

It is defined as $(N - C) + C'$ only in case either both C and C' are reasons, or both C and C' modify or undercut the same consideration. This ensures that the replaced consideration occupies the same place in the graph.[2]

4 Weighing Functions

This section explains how we model the effects of normatively-relevant considerations on each other, along with their effects on options.

Suppose that we had a function f^N which, given any normative graph N, would output the *final weights* of all considerations in it. Such a function would put us in a position to determine which options are permitted and which are forbidden in a rather straightforward way, where $\mathcal{C}_s = \{C \in R \mid c_C = s, s \in S\}$:

Definition 3 (Resolution function). *Given a normative graph* $N = (S, O, R)$*, an option* $o \in O$*, and a function* f^N*, let* $r^N(o) =$ `permitted` *if, for every* $o^* \in O$*,* $\sum_{C \in \mathcal{C}_o} f^N(C) \geq \sum_{C \in \mathcal{C}_{o^*}} f^N(C)$*;* $r^N(o) =$ `forbidden` *otherwise.*

And if there is a unique permitted option, we say that it *ought* to be carried out.

Of course, we still need to specify f^N. The way we calculate the final weight of a consideration C is via a two-step process using an additional function g^N. In the first step, we aggregate the (final) weights of C's amplifiers, attenuators, and undercutters. In the second step, we obtain the final weight of C on the basis of these aggregated weights and the default weight of C. This will be the job of f^N. (For readability, we will often omit the superscript where the context makes clear which graph we are talking about.) Note that even if we have a concrete aggregation function g at our disposal, there are many choices we could make for how f calculates the final weight from the output of g and C's default weight. So, before we define any concrete functions, it will be useful to think about some plausible *constraints* or *principles* that they should satisfy. (Many of the principles of weighted argumentation analyzed in [2] can be translated to our setting and make sense here as well.)

Definition 4 (Principles). *Let* $N = (S, O, R)$ *be an arbitrary normative graph,* C *some consideration, and* C' *and* C^* *a modifier and an undercutter of* C*, then:*

[2] It's worth noting two features of our model that might turn out to be either advantages or drawbacks. First, we represent reasons with negative polarity—reasons that speak against actions—only indirectly. In our model, any reason is always a reason *for* an option. So, it is a reason *against* an option only in so far as it adds to the final weight of an alternative option. Second, it is sometimes claimed that reasons can switch their polarity when combined [8]. Thus, in an (in)famous example, Prakken and Sartor [8] describe the effects of heat and rain on your going jogging: taken by themselves, the facts that it is raining and that it is hot constitute reasons for you not to go jogging, but, taken in combination, they speak in favor of going jogging. If these cases exist, our model cannot account for them. However, given that their existence is disputed (see e.g., [3,7,9]), our model may well gives the correct verdict here.

1. ***No Effects from Spurious Reasons:*** *If C is a reason in $N + C$, then if $f^{N+C}(C) = 0$, then for all options o, $r^N(o) = r^{N+C}(o)$.*
2. ***No Effects from Spurious Modifiers:*** *For any modifier $C \notin R$ of a consideration $C' \in R$ with $f^{N+C}(C) = 0$, we have $f^N(C') = f^{N+C}(C')$.*
3. ***No Valence Flips for Modifiers:*** *Given an amplifier C of some consideration $C' \in R$, $f^N(C') \geq f^{N-C}(C')$. Analogously, for an attenuator C of some consideration C', $f^N(C') \leq f^{N-C}(C')$.*
4. ***Modifier Reciprocity:*** *For any $C \in R$, if $g(a^+_C) = g(a^-_C)$ and $g(u_C) = 0$, then we have $f(C) = w_C$.*
5. ***Modeler's Delight:*** *For any consideration $C \in R$, we have $f(C) \geq 0$. If $g(u_C) > 0$, then $f(C) = 0$.*
6. ***Normative Parsimony*** *The function governing the weight of different considerations is uniform (and not gerrymandered) for reasons, undercutters, and modifiers.*
7. ***Relativity*** *Given a consideration $C = (p, c, u, a^-, a^+, w)$ and $x > 0$, let $x \times C$ be $(p, c, u, a^-, a^+, x \times w)$. Given a set of considerations R, let $x \times R = \{x \times C : C \in R\}$ and $x \times N = (S, O, x \times R)$. Then $r^N(o) =$ `forbidden` iff $r^{x \times N}(o) =$ `forbidden`.*
8. ***Distinct Roles*** *If $C' \in R$, $C^* \notin R$, $f^{N+C^*}(C^*) \neq 0$ and $f^N(C) \neq 0$, then $f^{N^{+C^*}_{-C'}}(C) \neq f^N(C)$. And if $C^* \in R$, $C' \notin R$, $f^N(C^*) \neq 0$ and $f^{N-C^*}(C) \neq 0$, then $f^{N^{+C'}_{-C^*}}(C) \neq f^N(C)$.*

Principles 1 and 2 state that the addition of both reasons and modifiers with the (final) weight of 0 should have no effects on which options are permissible and the final values of other specific considerations. Principle 3 states that no attenuator (no matter the rest of the graph) should ever help strengthen the weight of the consideration it modifies and that no amplifier should help weaken the weight of the consideration it modifies. Similarly, Principle 4 states that if the (aggregated) weight of attenuators and amplifiers is the same, their contributions cancel out and the final weight of the consideration they modify (if it is not undercut) is simply its default weight. Principles 5 and 6 are meant to ensure that weights get calculated in accordance with the scales metaphor. Principle 5 says that the minimal weight a consideration can have is 0 and that 0 is the weight it has if it is undercut. Principle 6 requires that the final weights of different kinds of normatively-relevant considerations are computed in the same way. Principle 7 states that weights have no meaning outside of the ratio scale they constitute. Principle 8, which one may or may not accept depending on one's metaethical inclinations, states that the roles of modifiers and undercutters in the economy of reasons should be distinct. (In particular, it states that an attenuator is never strong enough to entirely remove the weight of a consideration.)

With these principles in mind, we turn to concrete examples. Recall that we are looking for two functions g and f. The simplest thing to do for g is to add the (final) values of the nodes that "feed into" a consideration:

$$g_+(x) = \sum_{C' \in \mathcal{C}_x} f(C')$$

Notice that g_+ treats undercutters in an intuitive way, returning 0 just in case there either are no undercutters or all of them are themselves undercut. The leaves of the graph get assigned the weight 0. So, given a function f, we can compute the final weight of a consideration, working from the leaves up the tree.

Turning to the function f, we define two concrete functions here. The first of these, f_+, assigns a final weight of zero if either the consideration is undercut or if the combined weight of its attenuators is greater than that of its amplifiers and its default weight. Otherwise, it returns the default weight plus the difference between the combined weight of its amplifiers and attenuators. (Note that $\mathbb{1}$ is the indicator function.)

$$f_+(C) = \mathbb{1}_{\{0\}}(g(u_C)) * max(0, g(a_C^+) - g(a_C^-) + w_C)$$

The second function $f_\times$ uses modifiers as multipliers: amplifiers increase the weight of a consideration by a factor, and attenuators lower it by a factor.

$$f_\times(C) = \mathbb{1}_{\{0\}}(g(u_C)) * \frac{1 + g(a_C^+)}{1 + g(a_C^-)} * w_C$$

5 Results and Related Work

We take it that Principles 1–7 state conditions that should be satisfied by all functions, while Principle 8 (Distinct Roles) is up for debate. Our first result runs thus—its proof is straightforward and omitted for reasons of space.

Theorem 1. *Functions g_+ and $f_\times$ satisfy all principles stated in Definition 4. Functions g_+ and f_+ satisfy all principles but* Distinct Roles.

While ($f_\times$ and g_+) allow for attenuators to be as strong as undercutters and (f_+ and g_+) doesn't, a plausible result about the relationship between attenuators and undercutters can be established for both. Intuitively, it states that undercutters can be seen as a limit case of attenuators:

Theorem 2. *Given the functions ($f_\times$ and g_+) or (f_+ and g_+) and any consideration C_0, there is a series of attenuators C_i such that each C_i attenuates C_0 that can be added to the normative graph N, such that* $\lim_{i \to \infty} f^{N + \cup_{j=1}^{j=i} C_j}(C_0) = 0$.

Notice that Definition 4 is only the first step towards mapping out the space of normative principles that weighing functions can (should) satisfy. We leave fully-fledged principle-based analysis for future work. In the remainder of this section, we briefly compare our model to the work that comes the closest to it: Gordon and Walton's model of "balancing arguments" [4], the work on weighted argumentation by Amgoud et al. [1,2], and Horty's model of reasons [5].

We repurposed the model from [4] to a particular domain. As a result, we obtained a model that is simpler in a number of respects. For instance, the graphs that we work with are acyclic, and we have no need for labeling. The way we interpret weights is also different: where [4] assign values from $[0, 1]$ to arguments, interpreting them as the "level of acceptance", we allow all positive reals. As a result, we have also dropped the notion of "proof standard", which gets at the idea of a threshold of evidence for accepting a conclusion. No such threshold seems to exist in the context of practical normative reasons.

Amgoud et al. [1,2] work with the interval [0, 1], representing argument acceptance in their principle-based analysis of weighted argumentation. Their model assigns both default and final weights to arguments: attacks on arguments lower the final weight, while support increases it. The bulk of their principles can be restated in our framework, and many of them apply to reasons for action. In fact, our Principles 1–6 can be seen as translations from [2]. There are two main differences between the approach of Amgoud et al. and ours. First, the target of Amgoud et al. is the acceptability of arguments and not the interaction between normative reasons. Second, due to the interval scale used in their mode, an argument can be fully accepted simply due to the arguments "feeding into it". In our model, an issue is essentially contrastive: an option's normative status always depends on all reasons for all options. It also means that some of our principles (e.g., Principle 8) make little sense in the framework of Amgoud et al., while other principles aren't applicable in our model. Nevertheless, our principles both serve to show the (dis)similarities between normative reasons and arguments, we believe that there is a whole host of principles in the style of [2] unique to normative reasons that are still to be explored in future work.

Lastly, there is the model of Horty [5], which is meant to play a similar role to ours. The main advantage of our model over Horty's is its closer alignment to the idea of weight scales: our model associates magnitudes with reasons and lets us model combination or aggregation of weights in a straightforward way.

Acknowledgement. Vincent de Wit and Aleks Knoks acknowledge financial support from the Luxembourg National Research Fund (FNR). Knoks was supported through the project OPEN O20/14776480, de Wit through the project PRIDE19/14268506. We also thank our three anonymous referees for their generous comments.

References

1. Amgoud, L., Ben-Naim, J.: Evaluation of arguments in weighted bipolar graphs. Int. J. Approximate Reasoning **99**, 39–55 (2018)
2. Amgoud, L., Doder, D., Vesic, S.: Evaluation of argument strength in attack graphs: foundations and semantics. Artif. Intell. **302**, 1–61 (2022)
3. Bader, R.: Conditions, modifiers and holism. In: Lord, E., Maguire, B. (eds.) Weighing Reasons, pp. 27–55. Oxford University Press (2016)
4. Gordon, T., Walton, D.: Formalizing balancing arguments. In: Baroni, P., Gordon, T., Scheffler, T., Stede, M. (eds.) Proceedings of the 2016 Conference on Computational Models of Argument (COMMA 2016), pp. 327–38. IOS Press (2016)
5. Horty, J.: Reasons as Defaults. Oxford University Press, Oxford (2012)
6. Lord, E., Maguire, B.: An opinionated guide to the weight of reasons. In: Lord, E., Maguire, B. (eds.) Weighing Reasons, pp. 3–24. Oxford University Press (2016)

7. Maguire, B., Snedegar, J.: Normative metaphysics for accountants. Philos. Stud. **178**, 363–84 (2018)
8. Prakken, H., Sartor, G.: Modelling reasoning with precedents in a formal dialogue game. Artif. Intell. Law **6**, 231–87 (1998)
9. Tucker, C.: A holist balance scale. J. Am. Philos. Assoc. First View **9**, 1–21 (2022)