Defeasible Normative Reasoning: A Proof-Theoretic Integration of Logical Argumentation

Ofer Arieli[1], **Kees van Berkel**[2], **Christian Straßer**[2]
[1]School of Computer Science, Tel-Aviv Academic College, Israel
[2]Institute for Philosophy II, Ruhr University Bochum, Germany
oarieli@mta.ac.il, {kees.vanberkel, christian.strasser}@rub.de

Abstract

We present a novel computational approach to resolving conflicts among norms by nonmonotonic normative reasoning (in constrained I/O logics). Our approach extends standard sequent-based proof systems and makes them more adequate to nonmonotonic reasoning by adding to the sequents annotations that keep track of what is known about the defeasible status of the derived sequents. This makes transparent the reasons according to which norms should be applicable or inapplicable, and accordingly the sequents that make use of such norms are accepted or retracted. We also show that this proof theoretic method has tight links to the semantics of formal argumentation frameworks. The outcome of this paper is thus a threefold characterization result that relates, in the context of nonmonotonic normative reasoning, three traditional ingredients of AI-based reasoning methods: maximally consistent sets of premises (in constrained I/O logics), derived sequents (which are accepted in corresponding annotated sequent calculi), and logical arguments (that belong to the grounded extensions of the induced logical argumentation frameworks).

1 Introduction

The central role of defeasible, or nonmonotonic reasoning (NMR) in symbolic AI is due to the need to manage uncertainty and resolve conflicts in complex reasoning tasks. Over the past decades, formal argumentation (Dung 1995; Baroni et al. 2018) has proven to be an effective framework for the unified representation and comparison of nonmonotonic logics in AI. Its auspiciousness is due to its pivotal notions of argumentative attack and defense, which are strongly akin to natural reasoning practices (Mercier and Sperber 2011). Due to the inevitability of norm conflicts and defeasibility in normative reasoning, recent years have also seen an increasing interest in argumentative characterizations of logics of normative reasoning (Straßer and Arieli 2015; da Costa Peirera et al. 2017; Beirlaen, Straßer, and Heyninck 2018; Governatori, Rotolo, and Riveret 2018; Liao et al. 2018; Pigozzi and van der Torre 2018; Pardo and Straßer 2022).

A natural perspective on (logical) reasoning is proof-theoretic, where inference rules are iteratively applied. In particular, sequent-style proof systems (Gentzen 1934; Troelstra and Schwichtenberg 2000) own their renown three features: their modularity, enabling simultaneous investigation of large classes of logics; their proven suitability for the analysis of meta-properties; and their rule-based characteristic, accommodating constructive reasoning via proof-search. Such proof systems are predominantly monotonic. This perspective faces serious challenges in the context of defeasible reasoning, where inferences deemed unproblematic at some stage of the reasoning process may become problematic at a later stage (of a derivation). We find nonmonotonic proof theory in the literature, from Hilbert style proofs in adaptive logics (Batens 2007; Straßer 2014) to sequent based systems (Bonatti and Olivetti 2002; Giordano et al. 2009; Arieli and Straßer 2019). However, many challenges remain unaddressed, such as providing a transparent bridge to argumentation theory and affording an explicit account of different commitment statuses (such as acceptance and rejection).

NMR comes with a variety of inference relations: from credulous ones mapping out different, possibly inconsistent, but ultimately defendable stances, to various skeptical approaches. The latter differ, e.g., in their treatment of floating conclusions, i.e., conclusions that can be obtained by several otherwise conflicting arguments. In this paper we focus on an approach that blocks floating conclusions. In the context of normative reasoning it has the advantage of providing conclusions that are non-controversial and therefore give a firm basis for actionable decisions. The approach is known as *grounded semantics* in formal argumentation (Dung 1995) or as *free consequences* when reasoning with maximal consistent sets (Rescher and Manor 1970).

The general contribution of this article is a class of nonmonotonic calculi that proof-theoretically characterize skeptical reasoning by internalizing formal argumentation's notion of argumentative attack. In particular, we develop a proof system for defeasible skeptical reasoning in the context of formal normative reasoning (generally referred to as deontic logic), and show its relations to two central traditions in nonmonotonic reasoning: formal argumentation (Dung 1995) and constrained Input/Output (I/O) logics (Makinson and van der Torre 2001). That is, we demonstrate a threefold correspondence between the following approaches to NMR:

1. The class of sequent-style calculi proposed in this article: Annotated Deontic Argumentation Calculi) (ADAC). These are nonmonotonic proof systems deriving sequents augmented with annotations for describing their status in

a derivation (Arieli, van Berkel, and Straßer 2022). The ADAC formalism extends the monotonic calculi DAC from (van Berkel and Straßer 2022) with annotations and special rules for acceptance and rejection of annotated sequents, according to which conclusions are made. A first version of ADAC was proposed in (van Berkel 2023).

2. Grounded semantics of DAC-induced argumentation frameworks (Dung 1995). These frameworks are directed graphs whose nodes are DAC-derivable sequents based on a given knowledge base, and whose edges are obtained by applications of attacking sequents generated by the calculi (Arieli and Straßer 2015).
3. Input/Output (I/O) logic (Makinson and van der Torre 2001), a renowned formalism defining a class of logics for defeasible normative reasoning.[1] In particular, we focus on the formulas entailed by norms that are in every maximally consistent subset of the knowledge base (Straßer, Beirlaen, and Van De Putte 2016). These *free formulas* are the obligations that must be complied with irrespective of the credulous defendable stance taken.

Our primary contribution is to show that the derived formulas by ADAC-based annotated proof systems (Item 1 above), coincide with the grounded extension of the ADAC-induced argumentation framework that is obtained from an (inconsistent) normative knowledge base $\mathbb{K}$ (Item 2), which in turn are the free formulas of $\mathbb{K}$ within the I/O formalism (Item 3). In this way, we provide new links between logical argumentation and nonmonotonic normative reasoning, via NMR-tuned enhancements of traditional sequent calculi, which enable promising computational approaches for the latter.

The rest of this paper is organized as follows: In Section 2, the preliminaries for (annotated) deontic argumentation calculi are provided. In Section 3, these calculi are extended with annotations and annotation revision rules. In Section 4, the correspondence with argumentation frameworks is demonstrated (Thm. 1). Relations to I/O logics are proven in Section 5 (Thm. 2 and 3). In Section 6, we conclude with some references to related work.

2 Labelled Deontic Logics

We start with a description of labeled logics (serving as the base logics) and their sequent calculi. Following (van Berkel and Straßer 2022), we use labeled propositional languages[2], where the labels f, o, and c, express *facts, obligations* and *constraints*, respectively. Thus, a formula φ^f reads "it is a fact that φ," φ^o states that "it is obligatory that φ," and φ^c denotes that "φ is a constraint with which obligations must be consistent." The language also contains expressions of the form (φ,ψ) that denote *norms*, i.e., "given the fact φ, it is obligatory that ψ." We also define expressions such as $\neg\{(\varphi,\psi),(\theta,\sigma)\}$ stating that "the norms (φ,ψ) and (θ,σ) are jointly inapplicable." The latter is an extension of the language in (van Berkel and Straßer 2022), where only expressions of the form $\neg(\varphi,\psi)$ were allowed. In what follows, we demonstrate the benefits of generalizing the language to referring to joint inapplicability of sets of norms.

[1] See (Olszewski, Parent, and Van der Torre 2023) for the most recent overview of systems in its two decades of developments.

[2] We avoid overburdening the language with modalities by using labels. This suffices to represent the roles propositional formulas play in normative reasoning (van Berkel and Straßer 2022). Modal representations of deontic (I/O) logics are available, e.g., in (Makinson and van der Torre 2000; Lellmann 2021).

Definition 1 *Let* Atoms *be a denumerable set of propositional atoms. The languages $\mathcal{L}^i$ with $i \in \{f,o,c\}$ are defined through the following BNF grammar (where $p \in$* Atoms *and $\top$ [$\bot$] is the propositional constant for truth [falsity]):*

$$\varphi^i ::= p^i \mid \top^i \mid \bot^i \mid (\neg\varphi)^i \mid (\varphi \wedge \varphi)^i \mid (\varphi \vee \varphi)^i \mid (\varphi \to \varphi)^i$$

Let $\mathcal{L}^{\downarrow}$ be the language $\mathcal{L}^i$ (for any i) stripped from its labels, and $\mathcal{L}^n = \{(\varphi,\psi) \mid \varphi,\psi \in \mathcal{L}^{\downarrow}\}$. The language of norms *is defined as $\mathcal{L}^n \cup \overline{\mathcal{L}^n}$, where $\overline{\mathcal{L}^n} = \{\neg\Delta \mid \emptyset \subset \Delta \subseteq \mathcal{L}^n\}$ is the language expressing the inapplicability of norms. $\mathcal{L}^{deon} = \mathcal{L}^f \cup \mathcal{L}^o \cup \mathcal{L}^c \cup \mathcal{L}^n \cup \overline{\mathcal{L}^n}$ is called a* labelled deontic language.

We use p,q,r for atoms, and φ,ψ for arbitrary formulas of $\mathcal{L}^{deon}$. Sets of formulas from $\mathcal{L}^{deon}$ are denoted by $\mathcal{S},\mathcal{T}$ and Γ,Δ,Σ refer to *finite* sets of formulas. We write $\bigwedge\Delta$ [$\bigvee\Delta$] to denote the conjunction [disjunction] of the elements in Δ.

In defining proof calculi for normative reasoning, the starting point is always a *normative knowledge base*

$$\mathbb{K} = \langle \mathcal{F}, \mathcal{C}, \mathcal{N} \rangle$$

where $\mathcal{F} \subseteq \mathcal{L}^f$ is a set of *facts*, $\mathcal{C} \subseteq \mathcal{L}^c$ a set of *constraints*, and $\mathcal{N} \subseteq \mathcal{L}^n$ a set of *norms*. In the remainder, we assume that $\mathbb{K}$ is finite. We write $\mathbb{K}^{\downarrow}$ for the non-labelled version of $\mathbb{K}$, that is, for the triple $\langle \{\varphi \mid \varphi^f \in \mathcal{F}\}, \{\varphi \mid \varphi^c \in \mathcal{C}\}, \mathcal{N}\rangle$.

For each language $\mathcal{L}^i$ ($i \in \{f,o,c\}$), we assume a base logic $\mathsf{L}^i = \langle \mathcal{L}^i, \vdash_{\mathsf{L}^i}\rangle$, where $\vdash_{\mathsf{L}^i}$ is a consequence relation (Tarski 1941) for the language $\mathcal{L}^i$, and we assume a sound and complete *sequent calculus* LC^i (Gentzen 1934; Troelstra and Schwichtenberg 2000). The latter are proof systems consisting of expressions of the form $\Gamma \Rightarrow \Delta$ (called *sequents*), such that $\Gamma \vdash_{\mathsf{L}^i} \bigvee\Delta$ iff the sequent $\Gamma \Rightarrow \Delta$ is LC^i-derivable.

For simplicity, we assume in the remainder that for each language $\mathcal{L}^i$ with $i \in \{f,o,c\}$ the base logic L^i is classical. However, the generality of our approach is in accordance with (Arieli, van Berkel, and Straßer 2022) and (van Berkel and Straßer 2022), allowing for a large class of underlying base logics, possibly varying among each labelled language.

Deontic Argumentation Calculi (DAC) are defined next:

Definition 2 *For each $i \in \{f,o,c\}$, let LC^i be the base calculus for the language $\mathcal{L}^i$ of Definition 1. The minimal system, referred to as $\mathsf{DAC}_{\emptyset}$ is based on* Ax, Taut, Detach, R-C, R-NS, *and* Cut *from Figure 1. The calculus $\mathsf{DAC}_{\mathcal{S}}$ extends $\mathsf{DAC}_{\emptyset}$ with the rules $\mathcal{S} \subseteq$ {TP, L-OR, L-CT}. This leads to a total of 8 $\mathsf{DAC}_{\mathcal{S}}$-systems.*

A $\mathsf{DAC}_{\mathcal{S}}$-derivation of $\Gamma \Rightarrow \Delta$ is a tree whose leaves are initial sequents of $\mathsf{DAC}_{\mathcal{S}}$, whose root is $\Gamma \Rightarrow \Delta$, and whose rule-applications are instances of the rules of $\mathsf{DAC}_{\mathcal{S}}$. We write $\vdash_{\mathsf{DAC}_{\mathcal{S}}} \Gamma \Rightarrow \Delta$ if $\Gamma \Rightarrow \Delta$ is $\mathsf{DAC}_{\mathcal{S}}$-derivable.

$\mathsf{DAC}_{\mathcal{S}}$ extends the calculi LC^f, LC^o and LC^c in view of the Ax rule. The $\mathsf{DAC}_{\mathcal{S}}$ calculi in Definition 2 are nearly identical to the ones introduced and discussed in (van Berkel and Straßer 2022), except that the rule R-N has been generalized to R-NS by allowing for negated *sets* of norms on the

$$\frac{\vdash_{\mathsf{LC}^i} \Gamma \Rightarrow \Delta}{\Gamma^i \Rightarrow \Delta^i}\ \mathsf{Ax},\ i \in \{f, o, c\} \qquad \frac{}{\Rightarrow (\top, \top)}\ \mathsf{Taut}$$

$$\frac{}{\varphi^f, (\varphi, \psi) \Rightarrow \psi^o}\ \mathsf{Detach} \qquad \frac{}{\varphi^f \Rightarrow \varphi^o}\ \mathsf{TP}$$

$$\frac{\Gamma \Rightarrow \varphi^o}{\Gamma, (\neg\varphi)^c \Rightarrow}\ \mathsf{R\text{-}C} \qquad \frac{\Gamma, \Delta \Rightarrow}{\Gamma \Rightarrow \neg\Delta}\ \mathsf{R\text{-}NS}^1$$

$$\frac{\varphi^f, \Gamma \Rightarrow \Delta}{\varphi^o, \Gamma \Rightarrow \Delta}\ \mathsf{L\text{-}CT}^2 \qquad \frac{\Gamma, \varphi^f \Rightarrow \Delta \qquad \Gamma', \psi^f \Rightarrow \Delta}{\Gamma, \Gamma', (\varphi \vee \psi)^f \Rightarrow \Delta}\ \mathsf{L\text{-}OR}^3$$

$$\frac{\Gamma \Rightarrow \varphi \qquad \varphi, \Gamma' \Rightarrow \Delta}{\Gamma, \Gamma' \Rightarrow \Delta}\ \mathsf{Cut}^4$$

Figure 1: Rules for $\mathsf{DAC}_{\mathcal{S}}$ (Def. 2). Ax, Detach, Taut, and TP introduce initial sequents. Side-conditions: (1) on R-NS: $\emptyset \subset \Delta \subseteq \mathcal{L}^n$, (2) on L-CT: $\Gamma \cap \mathcal{L}^n \neq \emptyset$; (3) on L-OR: $\Gamma \cap \mathcal{L}^n \neq \emptyset$ and $\Gamma' \cap \mathcal{L}^n \neq \emptyset$, and is only imposed when $\mathsf{TP} \notin \mathcal{S}$; (4) on Cut: $\varphi \in \mathcal{L}^{deon}$.

right hand side of sequents. The underlying idea is as follows: Given that the argument $\varphi^f, (\varphi, \psi) \Rightarrow \psi^o$ (obtained by Detach) expresses that the fact φ^f and norm (φ, ψ) provide reasons for concluding the obligation ψ^o, the argument $\varphi^f, (\varphi, \psi), \neg\psi^c \Rightarrow$ (obtained by R-C) states that φ^f and (φ, ψ) are inconsistent with the constraint requiring obligations to be consistent with $\neg\psi^c$. From the latter (by application of R-N) we obtain $\varphi^f, \neg\psi^c \Rightarrow \neg(\varphi, \psi)$ expressing that φ^f and $\neg\psi^c$ are reasons for the *inapplicability* of (φ, ψ). Similarly, from $\varphi^f, (\varphi, \psi), (\varphi, \psi'), \neg(\psi \wedge \psi')^c \Rightarrow$ we obtain $\varphi^f, \neg(\psi \wedge \psi')^c \Rightarrow \neg\{(\varphi, \psi), (\varphi, \psi')\}$ (by R-NS), which expresses that φ^f and $\neg(\psi \wedge \psi')^c$ are reasons against the *joint application* of the norms (φ, ψ) and (φ, ψ'). The latter cannot be expressed in the calculi introduced in (van Berkel and Straßer 2022), to which we refer as DAC^-. We reserve DAC for the calculi defined in this section.

The next example illustrates the utility of DAC for normative reasoning and exemplifies various rule applications.

Example 1 *Suppose conferences* 1 *and* 2 *are two prestigious meetings Wilma is registered for. The two registrations (p_1 and p_2) induce normative reasons to attend the two conferences, expressed by (p_1, s_1) and (p_2, s_2) (e.g., the first norm reads "if Wilma registers for conference* 1*, she ought to attend it"). Suppose that Wilma promised her supervisor (p_3) to attend at least one of them, giving her a reason to participate in conference* 1 *or* 2 *$(p_3, s_1 \vee s_2)$. Later it is announced that the two conferences take place on the same day, far from each other, and attending both (s_1 and s_2) is impossible. Last, university policy dictates that by registering for a conference, Wilma ought to apply for funding (x).*

This scenario is captured by the normative knowledge base $\mathbb{K}_1 = \langle \mathcal{F}, \mathcal{C}, \mathcal{N} \rangle$, where $\mathcal{F} = \{p_1^f, p_2^f, p_3^f\}$, $\mathcal{N} = \{(p_1, s_1), (p_2, s_2), (s_1, x), (s_2, x), (p_3, s_1 \vee s_2)\}$, and $\mathcal{C} = \{\neg(s_1 \wedge s_2)^c\}$. Let $\mathsf{DAC}_{\mathcal{S}}$ be some calculus from Definition 2.

Now, consider (Part 1) of the proof in Figure 2, deriving the sequent A_1. From this sequent we can derive by R-NS the following sequents ($\mathsf{DAC}_{\mathcal{S}}$-derivable for any $\mathcal{S}$):

$\mathsf{A}_2 : p_1^f, p_2^f, \neg(s_1 \wedge s_2)^c, (p_1, s_1) \Rightarrow \neg\{(p_2, s_2)\}$

$\mathsf{A}_3 : p_1^f, p_2^f, \neg(s_1 \wedge s_2)^c \Rightarrow \neg\{(p_1, s_1), (p_2, s_2)\}$

$\mathsf{A}_4 : p_1^f, p_2^f, \neg(s_1 \wedge s_2)^c, (p_2, s_2) \Rightarrow \neg\{(p_1, s_1)\}$

A_2 expresses that $p_1^f, p_2^f, \neg(s_1 \wedge s_2)^c, (p_1, s_1)$ provide reasons not to apply (p_2, s_2), such as done in B_2. Similarly for A_4 and the application of (p_1, s_1) by B_1. According to A_3, (p_1, s_1) and (p_2, s_2) should not be applied jointly given p_1^f, p_2^f and the constraint $\neg(s_1 \wedge s_2)^c$.

We contrast two ways of obtaining the obligation x^o (of applying for funding). One, supposing L-CT, L-OR $\in \mathcal{S}$, is presented by Option 1 of Part 2 of the proof in Figure 2. The arguments B_3 and B_4 are unchallenged by A_2, A_3 and A_4, as they do not make use of the norms (p_1, s_1) and (p_2, s_2).

Alternatively, only supposing L-CT $\in \mathcal{S}$, one obtains x^o by chaining the norms (p_2, s_2) and (s_2, x) (resp. (p_1, s_1) and (s_2, x)), as shown in Option 2 of Part 2 of Figure 2.

In view of the conflict between the norms (p_1, s_1) and (p_2, s_2) (expressed in A_3), x^o is a floating conclusion *between B_5 and B_6, as it is obtained independently by two otherwise conflicting arguments. Unlike the reasons underlying B_4, the reasons given in arguments B_5 resp. B_6 are challenged by A_2 resp. A_4. In our example, this means that although Wilma cannot attend both conferences, if she is going to attend one of them, she ought to apply for funding.*

3 Annotated Deontic Argumentation Calculi

Let $\mathsf{DAC}_{\mathcal{S}}$ be a deontic argumentation calculus as in Definition 2. Our purpose in this section is to extend this monotonic calculus with *defeasible capabilities*. In the extended calculi it is not enough to derive a sequent for inferring its conclusion, but further evidence is needed, indicating that no conclusion of a counter-sequent can be inferred. In that case, we say that the sequent is *accepted*, and so not only that the sequent is derived, but also can be 'safely used' for making inferences.[3] Dually, sequents that are *rejected* are those for which a sequent with a counter-conclusion is accepted.

To express within the object level of the calculi the considerations above, we add annotations to sequents. An *annotated sequent* is an expression of the form $\Gamma \Rightarrow^{[\mathsf{s}]} \Delta$, where $\Gamma \Rightarrow \Delta$ is a standard sequent, and the annotation s is one of the following states, intuitively representing the status of the sequent in a derivation:

- [i] means that the sequent is introduced (derived);
- [a] indicates that the sequent is accepted;
- [r] denotes that the sequent is rejected.[4]

We use $[*]$ to indicate that the sequent's status is arbitrary.

Annotated versions for the rules of $\mathsf{DAC}_{\mathcal{S}}$ can now be defined as follows:

[3]The concept of (final) acceptance of sequents or arguments is also common in other calculi for nonmonotonic reasoning such as adaptive logic (Batens 2007; Straßer 2014) and argumentation-based proof systems (Arieli, van Berkel, and Straßer 2022).

[4]Our terminology is different from (Arieli, van Berkel, and Straßer 2022). What is there "finally accepted" is here "accepted." The status "rejected" is new and we do not need "eliminated" used in (Arieli, van Berkel, and Straßer 2022) for credulous reasoning.

Derivation Part 1:

$$\mathsf{R\text{-}C}\,\dfrac{\mathsf{Cut}\,\dfrac{\mathsf{Detach}\,\dfrac{}{\mathsf{B}_2:\ p_2^f,(p_2,s_2)\Rightarrow s_2^o}\qquad \mathsf{Cut}\,\dfrac{\mathsf{Detach}\,\dfrac{}{\mathsf{B}_1:\ p_1^f,(p_1,s_1)\Rightarrow s_1^o}\qquad \mathsf{Ax}\,\dfrac{}{s_1^o,s_2^o\Rightarrow(s_1\wedge s_2)^o}}{s_2^o,p_1^f,(p_1,s_1)\Rightarrow(s_1\wedge s_2)^o}}{p_1^f,p_2^f,(p_1,s_1),(p_2,s_2)\Rightarrow(s_1\wedge s_2)^o}}{\mathsf{A}_1:\ p_1^f,p_2^f,\neg(s_1\wedge s_2)^c,(p_1,s_1),(p_2,s_2)\Rightarrow}$$

Derivation Part 2 (Option 1):

$$\mathsf{Cut}\,\dfrac{\mathsf{Detach}\,\dfrac{}{\mathsf{B}_3:\ p_3^f,(p_3,s_1\vee s_2)\Rightarrow(s_1\vee s_2)^o}\qquad \mathsf{L\text{-}CT}\,\dfrac{\mathsf{L\text{-}OR}\,\dfrac{\mathsf{Detach}\,\dfrac{}{\mathsf{C}_1:\ s_1^f,(s_1,x)\Rightarrow x^o}\qquad \mathsf{Detach}\,\dfrac{}{\mathsf{C}_2:\ s_2^f,(s_2,x)\Rightarrow x^o}}{(s_1\vee s_2)^f,(s_1,x),(s_2,x)\Rightarrow x^o}}{(s_1\vee s_2)^o,(s_1,x),(s_2,x)\Rightarrow x^o}}{\mathsf{B}_4:\ p_3^f,(p_3,s_1\vee s_2),(s_1,x),(s_2,x)\Rightarrow x^o}$$

Derivation Part 2 (Option 2):

$$\mathsf{Cut}\,\dfrac{\mathsf{B}_2\qquad \mathsf{L\text{-}CT}\,\dfrac{\mathsf{C}_2}{s_2^o,(s_2,x)\Rightarrow x^o}}{\mathsf{B}_5:\ p_2^f,(p_2,s_2),(s_2,x)\Rightarrow x^o}\qquad\qquad \mathsf{Cut}\,\dfrac{\mathsf{B}_1\qquad \mathsf{L\text{-}CT}\,\dfrac{\mathsf{C}_1}{s_1^o,(s_1,x)\Rightarrow x^o}}{\mathsf{B}_6:\ p_1^f,(p_1,s_1),(s_1,x)\Rightarrow x^o}$$

Figure 2: DAC-derivations for Example 1

- The conclusion of an initial $\mathsf{DAC}_\mathcal{S}$-rule (i.e., without conditions) is annotated by [i]. For instance, the *annotated* $\mathsf{DAC}_\mathcal{S}$-rule Detach from Definition 2 is:

$$\dfrac{}{\varphi^f,(\varphi,\psi)\Rightarrow^{[\mathsf{i}]}\psi^o}$$

- The conclusion of any other $\mathsf{DAC}_\mathcal{S}$-rule, except for $\mathsf{R\text{-}NS}$ (see below), is annotated by the minimum over the annotations of the sequents in the rule's condition, where minimization is taken w.r.t. the ordering $\mathsf{r}<\mathsf{i}<\mathsf{a}$. Thus:

$$\dfrac{\Delta_1\Rightarrow^{[\mathsf{x}]}\Gamma_1\qquad \Delta_2\Rightarrow^{[\mathsf{y}]}\Gamma_2}{\Delta_3\Rightarrow^{[\min(\mathsf{x},\mathsf{y})]}\Gamma_3}$$

- For the rule $\mathsf{R\text{-}NS}$, the annotation of the sequent in the rule's conclusion is reinitialized to [i], unless the sequent in the rule's condition is accepted. That is,

$$\dfrac{\Gamma,\Delta\Rightarrow^{[\mathsf{x}]}}{\Gamma\Rightarrow^{[\mathsf{reset}(\mathsf{x})]}\neg\Delta}$$

  where $\mathsf{reset}(\mathsf{r})=\mathsf{i}$, $\mathsf{reset}(\mathsf{i})=\mathsf{i}$, and $\mathsf{reset}(\mathsf{a})=\mathsf{a}$. We come back to this in Example 3 below.

The annotated versions of the rules in $\mathsf{DAC}_\mathcal{S}$ are extended with rules for accepted and rejected sequents. These rules incorporate the notion of attack between arguments where the attacks are defined in terms of sequents concluding the inapplicability of norms in another sequent. Thus,

$$\Gamma\Rightarrow^{[*]}\neg\Delta'\ \textit{attacks}\ \Delta\Rightarrow^{[*]}\Sigma \text{ whenever } \Delta'\subseteq\Delta.$$

We denote by $\mathsf{Att}(\Gamma)=\{\Delta\Rightarrow\neg\Gamma' \text{ is } \mathsf{DAC}_\mathcal{S}\text{-derivable}\mid\Delta\subseteq\mathbb{K},\emptyset\subset\Gamma'\subseteq\Gamma\cap\mathcal{L}^n\}$ the set of sequents attacking sequents whose premise is Γ, relative to a normative knowledge base $\mathbb{K}$. The acceptance and rejection rules are then the following:

$\mathsf{Acpt\text{-}1}$: If $\mathsf{Att}(\Gamma)=\emptyset$, then:

$$\dfrac{\Gamma\Rightarrow^{[\mathsf{i}]}\Delta}{\Gamma\Rightarrow^{[\mathsf{a}]}\Delta}$$

$\mathsf{Acpt\text{-}2}$: If $\mathsf{Att}(\Gamma_1)\neq\emptyset$, then:

$$\dfrac{\Gamma_1\Rightarrow^{[\mathsf{i}]}\Delta_1\qquad (\forall\Gamma_2\Rightarrow\neg\Delta_2\in\mathsf{Att}(\Gamma_1))\ \Gamma_2\Rightarrow^{[\mathsf{r}]}\neg\Delta_2}{\Gamma_1\Rightarrow^{[\mathsf{a}]}\Delta_1}$$

Rjct: If $\Gamma_2\Rightarrow\neg\Delta_2\in\mathsf{Att}(\Gamma_1)$, then:

$$\dfrac{\Gamma_1\Rightarrow^{[\mathsf{i}]}\Delta_1\qquad \Gamma_2\Rightarrow^{[\mathsf{a}]}\neg\Delta_2}{\Gamma_1\Rightarrow^{[\mathsf{r}]}\Delta_1}$$

Intuitively, $\mathsf{Acpt\text{-}1}$ allows to accept derived sequents that have no attackers, Rjct allows to reject sequents with at least one accepted attacker, and $\mathsf{Acpt\text{-}2}$ allows to accept a sequent of which all attackers are rejected. Note that $\mathsf{Acpt\text{-}1}$ is a particular case of $\mathsf{Acpt\text{-}2}$. Interestingly, $\mathsf{Acpt\text{-}2}$ and Rjct correspond to Caminada's rules for 3-valued grounded labeling of argumentation frameworks (Caminada 2017; Baroni, Caminada, and Giacomin 2018), where accepted (resp. rejected, introduced) sequents correspond to arguments that are labelled In (resp. Out, Undecided).

Definition 3 *Let* $\mathsf{DAC}_\mathcal{S}$ *be a calculus from Definition 2 and let* $\mathbb{K}$ *be a given knowledge base. The Annotated Deontic Argumentation Calculus (*$\mathsf{ADAC}_\mathcal{S}$*, for short) extends the annotated version of the rules in* $\mathsf{DAC}_\mathcal{S}$ *with the rules* $\mathsf{Acpt\text{-}1}$*,* $\mathsf{Acpt\text{-}2}$ *and* Rjct *defined above.*

$\mathsf{ADAC}_\mathcal{S}$-derivation and $\mathsf{ADAC}_\mathcal{S}$-derivability (denoted $\vdash_{\mathsf{ADAC}_\mathcal{S}}$) are defined as usual, i.e., just as in the case of $\mathsf{DAC}_\mathcal{S}$ (see Section 2), but with respect to the extended annotated calculi. Entailment relations induced by $\mathsf{ADAC}_\mathcal{S}$ are then defined as follows:

Definition 4 *For a calculus* $\mathsf{ADAC}_\mathcal{S}$ *and a normative knowledge base* $\mathbb{K}=\langle\mathcal{F},\mathcal{C},\mathcal{N}\rangle$*, we denote by* $\mathbb{K}\mid\!\sim^{[\mathsf{a}]}_{\mathsf{ADAC}_\mathcal{S}}\Delta$ *that the annotated sequent* $\Gamma\Rightarrow^{[\mathsf{a}]}\Delta$ *is* $\mathsf{ADAC}_\mathcal{S}$*-derivable for some* $\Gamma\subseteq\mathcal{F}\cup\mathcal{C}\cup\mathcal{N}$*.*

Example 2 *Henceforth,* $\mathsf{S}[\mathsf{s}]$ *denotes the annotated sequent that is obtained from the plain sequent* S *whose status is* s.

Reconsider Example 1. By annotated versions of the derivations in that example (Figure 2), one obtains ADAC*-proofs for* $\mathsf{A_n}[\mathsf{i}]$ $(1 \leq \mathsf{n} \leq 4)$, $\mathsf{B_m}[\mathsf{i}]$ $(1 \leq \mathsf{m} \leq 6)$, *and* $\mathsf{C_k}[\mathsf{i}]$ $(1 \leq \mathsf{k} \leq 2)$. *Note that* $\mathsf{A}_3[\mathsf{i}]$ *has no attackers, simply because the sequent does not have norms on its left hand side, therefore* A_3 *is accepted:*

$$\frac{\mathsf{A}_3[\mathsf{i}]:\ p_1^f, p_2^f, \neg(s_1 \wedge s_2)^c \Rightarrow^{[\mathsf{i}]} \neg\{(p_1, s_1), (p_2, s_2)\}}{\mathsf{A}_3[\mathsf{a}]:\ p_1^f, p_2^f, \neg(s_1 \wedge s_2)^c \Rightarrow^{[\mathsf{a}]} \neg\{(p_1, s_1), (p_2, s_2)\}}\ \textit{Acpt-1}$$

Let $\mathsf{A}_1^\star = p_1^f, p_2^f, p_3^f, \neg(s_1 \wedge s_2)^c, (p_1, s_1), (p_2, s_2) \Rightarrow^{[\mathsf{i}]} \neg\{(p_3, s_1 \vee s_2)\}$. *This argument can be derived in a similar way as* A_1*, using* B_3*. Since,* $\mathsf{A}_1^\star$ *is attacked by* A_3*, which is already accepted, by* Rjct, $\mathsf{A}_1^\star$ *becomes rejected:*

$$\frac{\mathsf{A}_1^\star[\mathsf{i}] \qquad \mathsf{A}_3[\mathsf{a}]}{\mathsf{A}_1^\star[\mathsf{r}]}\ \textit{Rjct}$$

Similar considerations show that all the attackers of B_3 *and* B_4 *can be rejected, and so by* Acpt-2, B_3 *and* B_4 *are accepted as well (namely,* $\mathsf{B}_3[\mathsf{a}]$ *and* $\mathsf{B}_4[\mathsf{a}]$ *are* ADAC*-derived). Thus, e.g., (assuming* $\{\textit{L-CT}, \textit{L-OR}\} \subseteq \mathcal{S}$*),* $\mathbb{K}_1 \mid\!\sim^{[\mathsf{a}]}_{\mathsf{ADAC}_\mathcal{S}} x^o$.

Example 3 *As the following extension of Example 2 shows,* R-NS *does not necessarily preserve the sequent's status:*

$$\frac{\dfrac{\mathsf{A}_1[\mathsf{i}] \qquad \mathsf{A}_3[\mathsf{a}]}{\mathsf{A}_1[\mathsf{r}]}\ \textit{Rjct}}{\mathsf{A}_3:\ p_1^f, p_2^f, \neg(s_1 \wedge s_2) \Rightarrow^{[\mathsf{reset(r)}]} \neg\{(p_1, s_1), (p_2, s_2)\}}\ \textit{R-NS}$$

The inconsistent A_1 *is rejected by the unattackable* A_3 *and applying* **R-NS** *to the resulting* $\mathsf{A}_1[\mathsf{r}]$ *yields again* A_3 *which cannot be rejected. To avoid such situations, a reset is initiated (e.g., we may not yet have derived* A_3*'s acceptability).*

Last, we point out that the entailment relations $\mid\!\sim^{[\mathsf{a}]}_{\mathsf{ADAC}_\mathcal{S}}$ have some nice properties. The following are easily verified.

Proposition 1 *The entailments* $\mid\!\sim^{[\mathsf{a}]}_{\mathsf{ADAC}_\mathcal{S}}$ *are* paraconsistent *(i.e.,* $\varphi^o, \neg\varphi^o \not\mid\!\sim^{[\mathsf{a}]}_{\mathsf{ADAC}_\mathcal{S}} \psi^o$*) and* nonmonotonic *(e.g.,* $\varphi^f, (\varphi, \psi) \mid\!\sim^{[\mathsf{a}]}_{\mathsf{ADAC}_\mathcal{S}} \psi^o$*, but* $\varphi^f, (\varphi, \psi), (\varphi, \neg\psi) \not\mid\!\sim^{[\mathsf{a}]}_{\mathsf{ADAC}_\mathcal{S}} \psi^o$*).*

4 Formal Argumentation

Formal argumentation enables reasoning with possibly inconsistent knowledge bases. The merit of this method lies with its transparent way to track conflicts in the knowledge base by means of argumentative attacks. Reasoning with inconsistent *normative* knowledge bases is characterized as a form of defeasible reasoning naturally defined in terms of argumentative attacks that explicitly model norm conflicts.

Argumentation frameworks (Dung 1995) are directed graphs whose nodes represent arguments and whose edges represent argumentative attacks. We adopt a specific type of argumentation framework, called a *sequent-based framework*, whose arguments are sequents, induced by a given knowledge base and a sequent calculus (Arieli and Straßer 2015). In the context of DAC, the arguments are generated by a calculus $\mathsf{DAC}_\mathcal{S}$ and a normative knowledge base $\mathbb{K}$. An important property of sequent-based frameworks is that, unlike other settings for logic-based argumentation (e.g., Besnard and Hunter (2001)), arguments $\Gamma \Rightarrow \Delta$ are determined purely by their validity w.r.t. the underlying logic, and their support sets (Γ) need not be minimal nor consistent.

Let $\mathsf{A} = \Gamma \Rightarrow \neg\Delta$ and $\mathsf{B} = \Gamma' \Rightarrow \Delta'$ be two arguments (with $\Delta \neq \emptyset$). We are interested in two types of attack:

A **norm-attacks** B, iff $\neg\Delta = \neg\{(\varphi, \psi)\}$ for a $(\varphi, \psi) \in \Gamma'$.

A **consistency-attacks** B, iff $\Gamma \cap \mathcal{L}^n = \emptyset$ and $\Delta \subseteq \Gamma' \cap \mathcal{L}^n$.

If A norm-attacks B, A states reasons as to why some norm that is used in B is not applicable. If A consistency-attacks B, A states that some of the norms employed in B are not jointly applicable, since they lead to an inconsistency with respect to the constraints $\mathcal{C}$. In case of a consistency-attack the only reasons stated in A are constraints and facts (and so A is unassailable). Notice that some consistency-attacks are also norm-attacks, e.g., $\varphi^f \Rightarrow \neg\{(\varphi, \bot)\}$.

Let us make the above formally precise:

Definition 5 *An* argumentation framework *induced by a normative knowledge base* $\mathbb{K} = \langle \mathcal{F}, \mathcal{C}, \mathcal{N} \rangle$ *and a calculus* $\mathsf{DAC}_\mathcal{S}$*, is a pair* $\mathcal{AF}_{\mathsf{DAC}_\mathcal{S}}(\mathbb{K}) = \langle \mathsf{Arg}, \mathsf{Att} \rangle$*, such that:*

- Arg *is the set of all the sequents* $\Gamma \Rightarrow \Delta$ *that are* $\mathsf{DAC}_\mathcal{S}$*-derivable from* $\mathbb{K}$*, i.e., for which* $\Gamma \subseteq \mathcal{F} \cup \mathcal{C} \cup \mathcal{N}$.
- $(\mathsf{A}, \mathsf{B}) \in \mathsf{Att}$ *for* $\mathsf{A}, \mathsf{B} \in \mathsf{Arg}$ *if* A *norm-attacks* B *or* A *consistency-attacks* B.

We shall sometimes write $\mathsf{Arg}_{\mathsf{DAC}_\mathcal{S}}(\Sigma)$ *to denote the set of the* $\mathsf{DAC}_\mathcal{S}$*-arguments* $\Gamma \Rightarrow \Delta$ *for which* $\Gamma \subseteq \Sigma$.

Example 4 *Consider the arguments from Example 1 and the argument* $\mathsf{A}_1^\star$ *from Example 2. The argumentation framework based on* $\mathsf{Arg}_{\mathsf{DAC}_\mathcal{S}}(\mathbb{K}_1)$ *is depicted in Figure 3. (Recall that the presence of* B_4 *supposes* $\textit{L-CT}, \textit{L-OR} \in \mathcal{S}$ *and the presence of* B_5 *and* B_6 *only supposes* $\textit{L-CT} \in \mathcal{S}$*. The other arguments are derivable for any* $\mathcal{S}$*.) According to* A_2*, the norm* (p_2, s_2) *is inapplicable. Since arguments* $\mathsf{A}_1^\star, \mathsf{A}_4, \mathsf{B}_2$ *and* B_5 *make use of* (p_2, s_2) *they are norm-attacked by* A_2*. Similarly,* A_4 *norm-attacks* $\mathsf{A}_2, \mathsf{B}_1$ *and* B_6*. The argument* $\mathsf{A}_1^\star$ *is based on an inconsistent norm set including* (p_1, s_1) *and* (p_2, s_2)*. It is inconsistent in view of the constraint* $\neg(s_1 \wedge s_2)^c$ *and the facts* p_1^f *and* p_2^f*. By concluding* $\neg\{(p_1, s_1), (p_2, s_2)\}$*, argument* A_3 *thus consistency-attacks* $\mathsf{A}_1^\star$*, stating the joint inapplicability of these norms.*

Dung (1995) devised several semantics which provide different rationales to selecting justifiable sets of arguments from an argumentation framework. We recall the definitions required for establishing the main results of this paper.

Definition 6 *Let* $\mathcal{AF}_{\mathsf{DAC}_\mathcal{S}}(\mathbb{K}) = \langle \mathsf{Arg}, \mathsf{Att} \rangle$ *be an argumentation framework and* $\mathcal{A} \subseteq \mathsf{Arg}$ *a set of arguments. We say that* $\mathcal{A}$ defends *some argument* $\mathsf{A} \in \mathsf{Arg}$ *if for every* $\mathsf{B} \in \mathsf{Arg}$ *that attacks* A *there is a* $\mathsf{C} \in \mathcal{A}$ *that attacks* B.

- $\mathcal{A}$ *is* conflict-free *if there are no attacks between its elements:* $(\mathcal{A} \times \mathcal{A}) \cap \mathsf{Att} = \emptyset$*;*
- $\mathcal{A}$ *is* complete *if it is conflict-free, defends every argument in* $\mathcal{A}$*, and contains every argument that it defends;*
- $\mathcal{A}$ *is* grounded *if it is the unique* $\subseteq$*-minimally complete set of arguments;*
- $\mathcal{A}$ *is* stable *if it is a conflict-free set attacking every argument outside of* $\mathcal{A}$.

Figure 3: The argumentation framework of Example 4. Solid arrows represent norm-attacks, dashed arrows consistency-attacks. (Left) The gray arguments form the grounded set. (Right) The gray arguments form one of the two stable sets.

Example 5 *In Figure 3 we highlight in gray the grounded set (left) and one of the two stable sets (right) of the argumentation framework that is described in Example 4. The arguments in the grounded extension of the framework are exactly those accepted by the* ADAC*-derivation in Example 2. Theorem 1 shows that this is not a coincidence.*

Argumentation semantics can be used to define nonmonotonic entailment relations (see, e.g., Modgil and Prakken (2014)). Here, this is vindicated for the grounded entailment:

Definition 7 *Let* $\mathbb{K}$ *be a knowledge base and* $\mathsf{AF}_{\mathsf{DAC}_\mathcal{S}}(\mathbb{K}) = \langle \mathsf{Arg}, \mathsf{Att} \rangle$ *the argumentation framework induced by* $\mathbb{K}$ *and* $\mathsf{DAC}_\mathcal{S}$*. We denote by* $\mathbb{K} \mid\!\sim^{\mathsf{grd}}_{\mathsf{DAC}_\mathcal{S}} \varphi^o$ *that there is an argument of the form* $\Gamma \Rightarrow \varphi^o$ *in the grounded set of* $\mathsf{AF}_{\mathsf{DAC}_\mathcal{S}}(\mathbb{K})$.

Theorem 1 (Equivalence Characterization 1) *Let* $\mathbb{K}$ *be a knowledge base,* $\mathsf{DAC}_\mathcal{S}$ *a deontic calculus, and* $\mathsf{ADAC}_\mathcal{S}$ *be its annotated extension. Then:*

$$\mathbb{K} \mid\!\sim^{[\mathsf{a}]}_{\mathsf{ADAC}_\mathcal{S}} \varphi^o \text{ iff } \mathbb{K} \mid\!\sim^{\mathsf{grd}}_{\mathsf{DAC}_\mathcal{S}} \varphi^o.$$

5 Input/Output Logics

Next, we demonstrate that ADAC characterizes the free consequences of *nonmonotonic* Input/Output (I/O) logics (Makinson and van der Torre 2001). The I/O formalism is a prominent normative reasoning framework with various mechanisms to defeasibly detach obligations from norms in a given context and has been employed to defeasibly reason with norm conflicts, contrary-to-duty scenarios, and norm exceptions (Olszewski, Parent, and Van der Torre 2023). There is a wide range of applications of this formalism, covering other AI fields, e.g., dealing with causal and legal reasoning; see (Parent and van der Torre 2013; Bochman 2021).

Traditionally, the *monotonic* I/O formalism (Makinson and van der Torre 2000) has two equivalent characterizations: a proof-theoretic and a semantic one. We recall the former here, referred to as deriv. The basic I/O idea is to detach an obligation ψ from a norm $(\varphi, \psi) \in \mathcal{N}$ when φ expresses a fact (the I/O language is unlabelled). However, one wants to reason with the complex interaction of norms and the possible detachable obligations. One way to achieve this goal is to close a set of norms $\mathcal{N}$ under meta-rules before applying detachment. This idea is captured by the various deriv systems. For instance, one can adopt the meta-rule CT (below) that expresses a form of transitivity enabling successive detachment (cf. the application of $\mathsf{L\text{-}CT}$ in Figure 2), or the meta-rule OR that allows for reasoning by cases: if an obligation ψ follows from some norms under input φ as well as under input φ', it follows under input $\varphi \lor \varphi'$ (cf. the application of $\mathsf{L\text{-}OR}$ in Figure 2). Figure 4 (left) represents the eight I/O logics from (Makinson and van der Torre 2000) obtained by the following seven meta-rules, where $\vdash$ is the entailment relation of the base logic:[5]

$$\frac{(\varphi, \psi)}{(\varphi \land \varphi', \psi)}\ \mathsf{SI} \qquad \frac{(\varphi, \psi) \qquad \psi \vdash \psi'}{(\varphi, \psi')}\ \mathsf{WO}$$

$$\frac{(\varphi, \psi) \qquad (\varphi, \psi')}{(\varphi, \psi \land \psi')}\ \mathsf{AND} \qquad \frac{(\varphi, \psi) \qquad (\varphi', \psi)}{(\varphi \lor \varphi', \psi)}\ \mathsf{OR}$$

$$\frac{(\varphi, \psi) \qquad (\varphi \land \psi, \psi')}{(\varphi, \psi')}\ \mathsf{CT} \qquad \frac{}{(\top, \top)}\ \mathsf{T} \qquad \frac{}{(\varphi, \varphi)}\ \mathsf{ID}$$

Definition 8 *Given a set of norms* $\mathcal{N}$ *and a set of meta-rules* $\mathcal{R}$*, we let* $\mathsf{deriv}_\mathcal{R}(\mathcal{N})$ *be the closure of* $\mathcal{N}$ *under* $\mathcal{R}$*. Given a set of (unlabelled) facts* $\mathcal{F}$*, we define* $\varphi \in \mathsf{deriv}_\mathcal{R}(\mathcal{F}, \mathcal{N})$ *for* $\varphi \in \mathcal{L}^\downarrow$ *iff* $(\psi, \varphi) \in \mathsf{deriv}_\mathcal{R}(\mathcal{N})$ *for some* $\mathcal{F} \vdash \psi$.

The minimal system, denoted $\mathsf{deriv}_{\mathcal{R}_1}$*, is defined by* $\mathcal{R}_1 = \{\mathsf{SI}, \mathsf{WO}, \mathsf{AND}, \mathsf{T}\}$ *(Fig. 4). The other systems* $\mathsf{deriv}_{\mathcal{R}_i}$ *are extensions, where* $\mathcal{R}_1 \subset \mathcal{R}_i \subseteq \mathcal{R}_1 \cup \{\mathsf{OR}, \mathsf{CT}, \mathsf{ID}\}$ *(Fig. 4).*

As we have seen in the context of Example 1, the addition of constraints to a normative knowledge base may lead to inconsistency. This is demonstrated in the next example:

Example 6 *For the normative knowledge base in the running example (recall Example 1) we have that* $s_1 \land s_2 \in \mathsf{deriv}_{\mathcal{R}_i}(\mathcal{F}, \mathcal{N})$ *(for any* i*), while* $\neg(s_1 \land s_2) \in \mathcal{C}$.

In order to reason consistently in such circumstances, I/O-logics work on the basis of $\subseteq$-maximally consistent sets.

Definition 9 *Let* $\vdash$ *be the entailment relation of the underlying base logic and let* $\mathbb{K}$ *be a normative knowledge base. A subset* $\mathcal{N}'$ *of* $\mathcal{N}$ *is an* $\mathcal{R}$-maxicon set *for* $\mathbb{K}^\downarrow = \langle \mathcal{F}, \mathcal{C}, \mathcal{N} \rangle$ *in case* $\mathcal{C} \cup \mathsf{deriv}_\mathcal{R}(\mathcal{F}, \mathcal{N}') \nvdash \bot$ *and for all* $\mathcal{N}'' \subseteq \mathcal{N}$*, if* $\mathcal{N}' \subsetneq \mathcal{N}''$ *then* $\mathcal{C} \cup \mathsf{deriv}_\mathcal{R}(\mathcal{F}, \mathcal{N}'') \vdash \bot$*. We denote by* $\mathsf{maxicon}_\mathcal{R}(\mathbb{K}^\downarrow)$ *the set of all* $\mathcal{R}$*-maxicon sets for* $\mathbb{K}^\downarrow$*. The* free set $\mathcal{N}_{\mathsf{free}}$ *for* $\mathbb{K}^\downarrow$ *and* $\mathcal{R}$ *is defined as the intersection of all* $\mathcal{R}$*-maxicon sets for* $\mathbb{K}^\downarrow$*, i.e.,* $\mathcal{N}_{\mathsf{free}} = \bigcap \mathsf{maxicon}_\mathcal{R}(\mathbb{K}^\downarrow)$.

[5]The rule ID stipulates that the input (facts) are part of the output (obligations). In contrast to, e.g., Default Logic (Reiter 1980), I/O logics do not necessarily satisfy identity. See (Makinson and van der Torre 2000) for an extensive introduction and (Pardo and Straßer 2022) for a deontic default logic without identity.

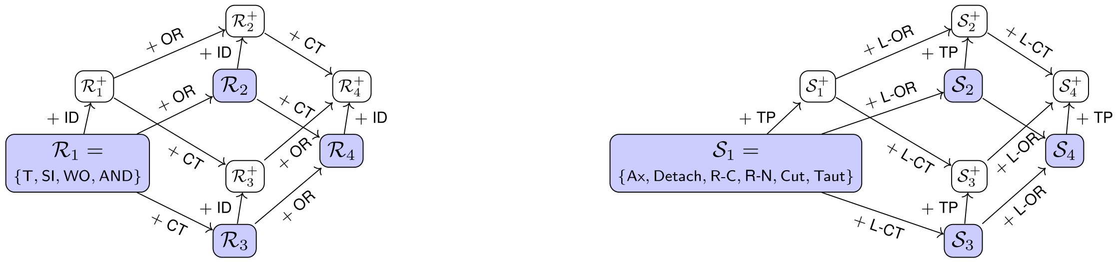

Figure 4: Left: The 8 sets of meta-rules for I/O-logics. Right: The 8 sets of inference rules in DAC. The set $\mathcal{S}_i^j$ corresponds to $\mathcal{R}_i^j$ for $i \in \{1, \ldots, 4\}$ and $j \in \{+, \emptyset\}$. Correspondence with the inference rules of ADAC is straightforwardly obtained using Definition 3. The superscript + conventionally indicates the presence of the ID, respectively TP rule.

Several consequence relations can be defined on the basis of maxicon sets. We are interested in the free consequences, which are well-known from reasoning with maximal consistent subsets in propositional logic (Rescher and Manor 1970) and have been studied in the context of I/O logics (Straßer, Beirlaen, and Van De Putte 2016). In the following, the I/O entailment is defined for unlabelled obligations.

Definition 10 *Let $\mathbb{K}$ be a normative knowledge base, $\mathbb{K}^{\downarrow} = \langle \mathcal{F}, \mathcal{C}, \mathcal{N} \rangle$ and $\mathcal{R}$ a set of meta-rules (Figure 4). We define $\mathbb{K}^{\downarrow} \mid\!\sim^{\mathsf{free}}_{\mathsf{I/O}_{\mathcal{R}}} \varphi$ iff $\varphi \in \mathsf{deriv}_{\mathcal{R}}(\mathcal{F}, \mathcal{N}_{\mathsf{free}})$.*

Example 7 *The maxicon sets for $\mathcal{R} \in \{\mathcal{R}_3, \mathcal{R}_4\}$ (Fig. 4) and $\mathbb{K}_1^{\downarrow}$ of the running example are: $\mathcal{N}_1 = \{(p_1, s_1), (s_1, x), (s_2, x), (p_3, s_1 \vee s_2)\}$ and $\mathcal{N}_2 = \{(p_2, s_2), (s_1, x), (s_2, x), (p_3, s_1 \vee s_2)\}$. Hence, the free set of norms is*

$$\mathcal{N}_{\mathsf{free}} = \{(s_1, x), (s_2, x), (p_3, s_1 \vee s_2)\}.$$

Note that $s_1 \vee s_2 \in \mathsf{deriv}_{\mathcal{R}}(\mathcal{F}, \mathcal{N}_1) \cap \mathsf{deriv}_{\mathcal{R}}(\mathcal{F}, \mathcal{N}_2)$. Moreover, $s_1 \vee s_2 \in \mathsf{deriv}_{\mathcal{R}_4}(\mathcal{F}, \mathcal{N}_{\mathsf{free}}) \setminus \mathsf{deriv}_{\mathcal{R}_3}(\mathcal{F}, \mathcal{N}_{\mathsf{free}})$. To see this, we observe that $(p_3, x) \in \mathsf{deriv}_{\mathcal{R}_4}(\mathcal{N}_{\mathsf{free}})$ by applications of OR and CT, but $(p_3, x) \notin \mathsf{deriv}_{\mathcal{R}_3}(\mathcal{N}_{\mathsf{free}})$. We therefore conclude that $\mathbb{K}_1^{\downarrow} \mid\!\sim^{\mathsf{free}}_{\mathsf{I/O}_{\mathcal{R}_4}} x$ and $\mathbb{K}_1^{\downarrow} \not\mid\!\sim^{\mathsf{free}}_{\mathsf{I/O}_{\mathcal{R}_3}} x$, while $\mathbb{K}_1^{\downarrow} \mid\!\sim^{\mathsf{free}}_{\mathsf{I/O}_{\mathcal{R}_i}} s_1 \vee s_2$ for $i = 3$ and $i = 4$. Thus, the obligation to apply for funding (x) is only inferred in the case of $\mathcal{R}_4$ and the obligation to attend either of the two conferences $(s_1 \vee s_2)$ is inferred in both $\mathcal{R}_3$ and $\mathcal{R}_4$ (cf. Example 5). That this is not a coincidence, is shown in Theorem 2 below.

Theorem 2 shows that nonmonotonic entailments in ADAC correspond to the free consequences of the I/O formalism. Theorem 3 is a corollary of Theorems 1 and 2. We assume that $\mathcal{R}$ and $\mathcal{S}$ correspond according to Figure 4.

Theorem 2 (Equivalence Characterization 2) *Let $\mathbb{K}$ be a normative knowledge base. Then:*

$$\mathbb{K} \mid\!\sim^{[\mathsf{a}]}_{\mathsf{ADAC}_{\mathcal{S}}} \varphi^{o} \text{ iff } \mathbb{K}^{\downarrow} \mid\!\sim^{\mathsf{free}}_{\mathsf{I/O}_{\mathcal{R}}} \varphi.$$

Theorem 3 (Equivalence Characterization 3) *Let $\mathbb{K}$ be a normative knowledge base. Then:*

$$\mathbb{K} \mid\!\sim^{\mathsf{grd}}_{\mathsf{DAC}_{\mathcal{S}}} \varphi^{o} \text{ iff } \mathbb{K}^{\downarrow} \mid\!\sim^{\mathsf{free}}_{\mathsf{I/O}_{\mathcal{R}}} \varphi.$$

6 Related Works and Concluding Remarks

The primary contribution of this paper comprises a formal model of normative reasoning that is intuitive (since it is rule-based and iterative), transparent and explanatory (since it is based on stating reasons in arguments and the dialectic interplay among arguments), and unifying (since it characterizes two central paradigms in nonmonotonic inference). The main technical contributions of the paper are the correspondences between acceptability in proof systems for defeasible logics, grounded inference in formal argumentation, and the nonmonotonic semantic entailment relation of the I/O formalism. This threefold characterization result (Figure 5) demonstrates the merits of adopting an argumentative proof-theoretic approach to defeasible normative reasoning.

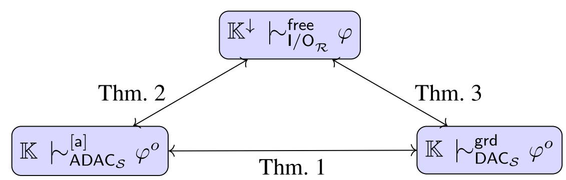

Figure 5: Equivalences for defeasible normative reasoning.

Proof theoretic characterizations of *monotonic* I/O-logics can be found in (Lellmann 2021; Ciabattoni and Rozplokhas 2023), while a proof theory of *nonmonotonic* I/O-logics has been presented in (Straßer, Beirlaen, and Van De Putte 2016) in terms of adaptive logics. We mention some main differences to the latter work. First, in our calculus derived sequents are labelled in terms of the status of their derivation (accepted, rejected, introduced), unlike adaptive proofs in which proof lines (of a Hilbert-style proof) are marked as *defeasibly* defeated. In contrast, our labels of accepted and rejected represent *final* statuses. Second, our sequents represent arguments that explicitly state reasons for their conclusions in terms of norms, facts, and constraints on their left hand side. As argued in (van Berkel and Straßer 2022), the stated reasons serve an explanatory function. For this to be transparent it is important that no new norms are generated, and that only the norms in the knowledge base serve as potential reasons. In contrast, in (Straßer, Beirlaen, and Van De Putte 2016) new norms are generated from old ones following the meta-rules in Fig. 4 (left). Finally, we establish a link between the set of accepted sequents and the grounded semantics of argumentation frameworks and thereby closely link our calculi to the tradition of abstract argumentation (see Baroni, Caminada, and Giacomin (2018)).

Acknowledgments

This work is partly supported by the Israel Science Foundation (Grant No. 550/19), and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project Number 511915728.

References

Arieli, O.; and Straßer, C. 2015. Sequent-based logical argumentation. *Argument and Computation*, 6(1): 73–99.

Arieli, O.; and Straßer, C. 2019. Logical argumentation by dynamic proof systems. *Theoretical Computer Science*, 781: 63–91.

Arieli, O.; van Berkel, K.; and Straßer, C. 2022. Annotated sequent calculi for paraconsistent reasoning and their relations to logical argumentation. In *Proceedings of IJCAI-22*, 2532–2538. ijcai.org.

Baroni, P.; Caminada, M.; and Giacomin, M. 2018. Abstract argumentation frameworks and their semantics. In *Handbook of Formal Argumentation, Volume 1*, 159–236. College Publications.

Baroni, P.; Gabbay, D.; Giacomin, M.; and van der Torre, L. 2018. *Handbook of Formal Argumentation, Volume 1*. College Publications.

Batens, D. 2007. A universal logic approach to adaptive logics. *Logica universalis*, 1(1): 221–242.

Beirlaen, M.; Straßer, C.; and Heyninck, J. 2018. Structured argumentation with prioritized conditional obligations and permissions. *Journal of Logic and Computation*, 29(2): 187–214.

van Berkel, K. 2023. *A Logical Analysis of Normative Reasoning: Agency, Action, and Argumentation*. Ph.D. thesis, TU Wien.

van Berkel, K.; and Straßer, C. 2022. Reasoning with and about norms in logical argumentation. In *Frontiers in Artificial Intelligence and Applications*, volume 353 of *Proceedings of COMMA'22*, 332 – 343. IOS press.

Besnard, P.; and Hunter, A. 2001. A logic-based theory of deductive arguments. *Artificial Intelligence*, 128(1-2): 203–235.

Bochman, A. 2021. *A Logical Theory of Causality*. MIT Press.

Bonatti, P. A.; and Olivetti, N. 2002. Sequent calculi for propositional nonmonotonic logics. *ACM Transactions on Computational Logic (TOCL)*, 3(2): 226–278.

Caminada, M. 2017. Argumentation semantics as formal discussion. In *Handbook of Formal Argumentation, Volume 1*, 487–518. College Publications.

Ciabattoni, A.; and Rozplokhas, D. 2023. Streamlining Input/Output logics with sequent calculi. In *Proceedings of KR'23*, 146–155.

da Costa Peirera, C.; Tettamanzi, A. G.; Villata, S.; Liao, B.; Malerba, A.; Rotolo, A.; and van der Torre, L. 2017. Handling norms in multi-agent system by means of formal argumentation. *IfCoLog Journal of Logics and Their Applications*, 4(9): 1–35.

Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2): 321–357.

Gentzen, G. 1934. Untersuchungen über das logische Schließen I, II. *Mathematische Zeitschrift*, 39: 176–210, 405–431.

Giordano, L.; Gliozzi, V.; Olivetti, N.; and Pozzato, G. L. 2009. Analytic tableaux calculi for KLM logics of nonmonotonic reasoning. *ACM Transactions on Computational Logic (TOCL)*, 10(3): 18.

Governatori, G.; Rotolo, A.; and Riveret, R. 2018. A deontic argumentation framework based on deontic defeasible logic. In *Proceedings of PRIMA'2018*, 484–492.

Lellmann, B. 2021. From Input/Output logics to conditional logics via sequents – with provers. In *Proceedings of Tableaux'21*, 147–164. Springer.

Liao, B.; Oren, N.; van der Torre, L.; and Villata, S. 2018. Prioritized norms in formal argumentation. *Journal of Logic and Computation*, 29(2): 215–240.

Makinson, D.; and van der Torre, L. 2000. Input/Output logics. *Journal of Philosophical Logic*, 383–408.

Makinson, D.; and van der Torre, L. 2001. Constraints for Input/Output logics. *Journal of Philosophical Logic*, 30(2): 155–185.

Mercier, H.; and Sperber, D. 2011. Why do humans reason? Arguments for an argumentative theory. *Behavioral and brain sciences*, 34(2): 57–74.

Modgil, S.; and Prakken, H. 2014. The ASPIC+ framework for structured argumentation: a tutorial. *Argument & Computation*, 5(1): 31–62.

Olszewski, M.; Parent, X.; and Van der Torre, L. 2023. Permissive and regulative norms in deontic logic. *Journal of Logic and Computation*, exad024.

Pardo, P.; and Straßer, C. 2022. Modular orders on defaults in formal argumentation. *Journal of Logic and Computation*.

Parent, X.; and van der Torre, L. 2013. Input/Output Logic. In *Handbook of Deontic Logic and Normative Systems, Volume 1*, 499–544. College Publications.

Pigozzi, G.; and van der Torre, L. 2018. Arguing about constitutive and regulative norms. *Journal of Applied Non-Classical Logics*, 28(2-3): 189–217.

Reiter, R. 1980. A logic for default reasoning. *Artificial intelligence*, 13(1-2): 81–132.

Rescher, N.; and Manor, R. 1970. On inference from inconsistent premises. *Theory and Decision*, 1: 179–217.

Straßer, C.; and Arieli, O. 2015. Normative reasoning by sequent-based argumentation. *Journal of Logic and Computation*, 29(3): 387–415.

Straßer, C. 2014. *Adaptive Logics for Defeasible Reasoning. Applications in Argumentation, Normative Reasoning and Default Reasoning*, volume 38 of *Trends in Logic*. Springer.

Straßer, C.; Beirlaen, M.; and Van De Putte, F. 2016. Adaptive logic characterizations of input/output logic. *Studia Logica*, 104(5): 869–916.

Tarski, A. 1941. *Introduction to Logic*. Oxford University Press.

Troelstra, A. S.; and Schwichtenberg, H. 2000. *Basic proof theory*. 43. Cambridge University Press.