Sequent-Based Argumentation for Normative Reasoning

Christian Straßer   Ofer Arieli

March 13, 2014

Abstract

In this paper we present an argumentative approach to normative reasoning. Special attention is paid to normative conflicts, contrary-to-duty and specificity cases. These are modeled by means of argumentative attacks. For this, we adopt a recently proposed framework for logical argumentation in which arguments are generated by a sequent calculus of a given base logic (Arieli, CLIMA'2013, pp.69–85), and use an intuitionistic variant of standard deontic logic as our base logic. Argumentative attacks are realized by elimination rules that allow to discharge specific sequents. We demonstrate our system by means of various well-known benchmark examples.

1 Introduction

Normative reasoning concerns reasoning with and about norms such as obligations, imperatives, permissions, etc. A paradigmatic instance is so-called factual detachment which says: if φ holds, and there is a commitment to ψ conditional on φ, then there is a commitment to ψ. Another instance is aggregation: if there is a norm to bring about φ and another norm to bring about ψ then there should be a norm to bring about $\varphi \wedge \psi$. Allowing for unrestricted factual detachment or unrestricted aggregation is problematic in cases in which norms conflict [1]. For instance, aggregating two conflicting norms leads to a norm that commits us to do the impossible. Other problematic cases concern specificity: sometimes more specific norms override more general norms. In such cases we want to block factual detachment from the overridden norms. Logical accounts of normative reasoning that is tolerant with respect to normative conflicts and/or specificity cases have been shown to be challenging. This has given rise to a variety of approaches (e.g., [2, 3, 4, 5, 6, 7]).

In this paper we model normative reasoning by means of logical argumentation. Given a set of facts and a set of possibly conflicting and interdependent conditional norms we will demonstrate how this model helps us to identify sets of non-conflicting norms that are apt to guide the actions of a user. Furthermore, we will show how it offers an elegant tool to deal with specificity cases. It follows that the entailment relations that are obtained offer conflict-handling mechanisms for various types of conflicts, and as such they are adaptive to different application contexts.

Our starting point in modeling normative reasoning is concerned with Dung's well-known abstract argumentation frameworks [8]. These frameworks consist of a set of abstract objects (the 'arguments') and an attack relation between them. Their role is to serve as a tool to analyze and reason with arguments. Various procedures for selecting accepted arguments have been proposed, based on the dialectical relationships between the arguments. Usually, these methods avoid selecting arguments that conflict with each other and allows to respond to every possible attack on the argumentative stance with a counter-argument.

For formalizing normative reasoning we need to enhance abstract argumentation in order to model the structure of arguments. There are various ways of doing so (e.g., [9, 10]) In this paper, we settle for the representation in terms of *sequents* [11]. One advantage of this approach is that it immediately equips us with dynamic proof procedures in the style of adaptive logics [12, 13] that allow for automated reasoning [14]. Another advantage is that we can plug in any Tarskian logic that comes with an adequate sequent calculus as a base logic that produces our arguments.

In this paper we have found it useful to use ISDL (intuitionstic fragment of the standard deontic logic SDL) as our base logic (see Sec. 2). In this context, the modality O is used to model obligations and permissions are modeled by P, defined by $\neg\mathsf{O}\neg$. Accordingly, arguments are (proofs of) derivable sequents $\Gamma \Rightarrow \phi$ (for some finite set of formulas Γ and a formula ψ) in a sequent calculus for ISDL, based on Gentzen's LJ proof system [15]. Attacks between arguments are represented by attack rules that allow to derive elimination sequents of the form $\Gamma \not\Rightarrow \phi$, whose effect is the canceling or uncharging of $\Gamma \Rightarrow \phi$ (see [11]).

The following example illustrates (still on the intuitive level) how the sequent-based argumentation framework described above is useful for modeling normative reasoning.

Example 1. *Consider the following example by Horty [16]:*

- *When served a meal you ought to not eat with fingers.*
- *However, if the meal is asparagus you ought to eat with fingers.*

The statements above may be represented, respectively, by the formulas $m \supset \mathsf{O}\neg f$ *and* $(m \wedge a) \supset \mathsf{O}f$*. Now, in case we are indeed served asparagus* $(m \wedge a)$ *we expect to derive the (unconditional) obligation to eat with fingers* ($\mathsf{O}f$) *rather than to not eat with fingers* ($\mathsf{O}\neg f$)*. This is a paradigmatic case of* specificity*: a more specific obligations cancels (or overrides) a less specific one. In our setting this will be handled by an attack rule advocating specificity (*SPEC*, see Example 5 below), according to which the argument* $\{m \wedge a,\ (m \wedge a) \supset \mathsf{O}f\} \Rightarrow \mathsf{O}f$ *attacks the argument* $\{m,\ m \supset \mathsf{O}\neg f\} \Rightarrow \mathsf{O}\neg f$*, and as a consequence* $\mathsf{O}f$ *will be inferable in this case while* $\mathsf{O}\neg f$ *will not.*

2 Intuitionistic SDL

The base logic that we shall use in this paper is an intuitionistic variation of SDL (standard deontic logic, i.e., the normal modal logic KD), called ISDL. The underlying language $\mathcal{L}_{\mathsf{ISDL}}$ consists of a propositional constant $\bot$ (representing falsity), the standard operators for conjunction $\wedge$, disjunction $\vee$, and implication $\supset$, and the modal operator O representing obligations. Thus, for instance, the conditional obligation $\phi \supset \mathsf{O}\psi$ may be intuitively understood as "ϕ commits to bring about ψ".

We shall denote formulas in $\mathcal{L}_{\mathsf{ISDL}}$ by the lower Greek letter ψ, ϕ, and set of formulas by the upper Greek letters Γ, Δ, Σ. Following the usual conventions, we abbreviate $\psi \supset \bot$ by $\neg\psi$ and incorporate the modality P for representing permissions, where $\mathsf{P}\psi$ is defined by $\neg\mathsf{O}\neg\psi$. Other abbreviations that we shall use in the sequel are $\top$ for the formula $\bot \supset \bot$, $\mathsf{O}\Gamma$ for the set $\{\mathsf{O}\psi \mid \psi \in \Gamma\}$, and $\bigwedge\Gamma$ for the conjunction of the formulas in a finite set Γ.

The reason for choosing intuitionistic logic is to avoid undesirable phenomena caused by using a contrapositive implication [17].[1] Reasoning with ISDL is done by $\mathcal{L}_{\mathsf{ISDL}}$-*sequents* (or just sequents, for short), that is: expressions of the form $\Gamma \Rightarrow \psi$, where Γ is a finite set of $\mathcal{L}$-formulas and $\Rightarrow$ is a symbol that does not appear in $\mathcal{L}_{\mathsf{ISDL}}$. We shall denote $\mathsf{Prem}(\Gamma \Rightarrow \psi) = \Gamma$.

[1] Yet, it should be noted that this choice is not obligatory, and our setting is adjusted to other deontic logics such as SDL.

Given a set Σ of formulas in $\mathcal{L}_{\mathsf{ISDL}}$, we say that a formula ψ *follows* from Σ (in ISDL), and denote this by $\Sigma \vdash_{\mathsf{ISDL}} \psi$, if there is a subset $\Gamma \subseteq \Sigma$, such that the $\mathcal{L}_{\mathsf{ISDL}}$-sequent $\Gamma \Rightarrow \psi$ is provable in the sequent calculus $\mathcal{C}_{\mathsf{ISDL}}$ shown in Figure 1. It is easy to verify that $\vdash_{\mathsf{ISDL}}$ is a Tarskian consequence relation (that is, reflexive, monotonic and transitive).

Axioms: $\psi \Rightarrow \psi, \quad \bot \Rightarrow \psi,$

Structural Rules:

Weakening: $$\frac{\Gamma \Rightarrow \psi}{\Gamma, \Gamma' \Rightarrow \psi}$$

Cut: $$\frac{\Gamma \Rightarrow \psi \quad \Gamma', \psi \Rightarrow \phi}{\Gamma, \Gamma' \Rightarrow \phi}$$

Logical Rules:

$[\wedge\Rightarrow]$ $$\frac{\Gamma, \psi, \varphi \Rightarrow \phi}{\Gamma, \psi \wedge \varphi \Rightarrow \phi}$$

$[\Rightarrow\wedge]$ $$\frac{\Gamma \Rightarrow \psi \quad \Gamma \Rightarrow \varphi}{\Gamma \Rightarrow \psi \wedge \varphi}$$

$[\vee\Rightarrow]$ $$\frac{\Gamma, \psi \Rightarrow \phi \quad \Gamma, \varphi \Rightarrow \phi}{\Gamma, \psi \vee \varphi \Rightarrow \phi}$$

$[\Rightarrow\vee]$ $$\frac{\Gamma \Rightarrow \psi}{\Gamma \Rightarrow \psi \vee \varphi} \qquad \frac{\Gamma \Rightarrow \varphi}{\Gamma \Rightarrow \psi \vee \varphi}$$

MP: $$\frac{}{\Gamma, \phi, \phi \supset \psi \Rightarrow \psi}$$

$[\Rightarrow\supset]$ $$\frac{\Gamma, \psi \Rightarrow \varphi}{\Gamma \Rightarrow \psi \supset \varphi}$$

KR: $$\frac{\Gamma \Rightarrow \phi}{\mathsf{O}\Gamma \Rightarrow \mathsf{O}\phi}$$

DR: $$\frac{\Gamma \Rightarrow \phi}{\mathsf{O}\Gamma \Rightarrow \neg\mathsf{O}\neg\phi}$$

NEC: $$\frac{\Rightarrow \phi}{\Rightarrow \mathsf{O}\phi}$$

Figure 1: The proof system $\mathcal{C}_{\mathsf{ISDL}}$

Note 2. The proof system $\mathcal{C}_{\mathsf{ISDL}}$ is equivalent to Gentzen's well-known sequent calculus *LJ* for intuitionistic propositional logic, extended with the rules for the modal operator O [18]. In particular, in $\mathcal{C}_{\mathsf{ISDL}}$ the rule [MP] is primitive and the rule

$$[\supset\Rightarrow] \quad \frac{\Gamma \Rightarrow \psi \quad \Gamma, \varphi \Rightarrow \phi}{\Gamma, \psi \supset \varphi \Rightarrow \phi}$$

is admissible (i.e., it is derivable from the rules of $\mathcal{C}_{\mathsf{ISDL}}$), while in *LJ* it is the other way around.

3 Logical Argumentation for Normative Reasoning

In what has become the orthodox approach based on Dung's representation [8], formal argumentation is studied on the basis of so-called argumentation frameworks. An argumentation framework in its most abstract form is a directed graph, where the nodes present (abstract) arguments and the arrows present argumentative attacks.

Definition 3. *An* (abstract) argumentation framework *is a pair* $\langle Args, Attack \rangle$*, where Args is an enumerable set of elements, called (abstract)* arguments*, and Attack is a relation between arguments whose instances are called* attacks*.*

When it comes to specific applications of formal argumentation it is often useful to provide an *instantiation* of (abstract) argumentation frameworks. Instantiations provide a specific account of the structure of arguments, and the concrete nature of argumentative attacks. There are various formal accounts available that provide frameworks for instantiating abstract argumentation such as assumption-based argumentation [9], ASPIC [10], etc. Here we settle for a recently proposed account based on sequent-based calculi [11].

The basic idea behind our instantiation is that arguments are $\mathcal{C}_{\mathsf{ISDL}}$-proofs.

Definition 4. $\mathsf{Arg}(\Sigma)$ *is the set of* $\mathcal{C}_{\mathsf{ISDL}}$*-proofs of sequents of the form* $\Gamma \Rightarrow \psi$ *for some* $\Gamma \subseteq \Sigma$*.*

For specifying the attack relation we complement $\mathcal{C}_{\mathsf{ISDL}}$ with *sequent elimination rules*. Unlike the inference (or, sequent introduction) rules of $\mathcal{C}_{\mathsf{ISDL}}$, the conclusions of sequent elimination rules are of the form $\Gamma \not\Rightarrow \psi$, and their intuitive meaning is the discharging sequent $\Gamma \Rightarrow \psi$.

Example 5. Consider the following sequent elimination rule:

$$\mathsf{SPEC}\ \frac{\Gamma,\, \phi \supset \psi \Rightarrow \psi \quad \Gamma \Rightarrow \phi \quad \Gamma' \Rightarrow \phi' \quad \phi \Rightarrow \phi' \quad \psi \Rightarrow \neg\psi' \quad \Gamma',\, \phi' \supset \psi' \Rightarrow \psi'}{\Gamma',\, \phi' \supset \psi' \not\Rightarrow \psi'}$$

This rule aims at formalizing the principle of specificity. It states that when two sequents $\Gamma' \Rightarrow \psi'$ and $\Gamma \Rightarrow \psi$ are conflicting, the one which is more specific gets higher precedence, and so the other one is discarded. Thus, in Example 1 for instance, SPEC allows to discharge the sequent $m,\, m \supset \mathsf{O}\neg f \Rightarrow \mathsf{O}\neg f$ in light of the more specific sequent $m \wedge a,\, (m \wedge a) \supset \mathsf{O}f \Rightarrow \mathsf{O}f$.

Some variations of SPEC are given below (where $\mathsf{NN'} \in \{\mathsf{OO}, \mathsf{OP}, \mathsf{PO}\}$):[2]

$$\mathsf{NN'SPEC}\ \frac{\begin{matrix}\Gamma,\, \phi \supset \mathsf{N}\psi \\ \Rightarrow \mathsf{N}\psi\end{matrix} \quad \Gamma \Rightarrow \phi \quad \Gamma' \Rightarrow \phi' \quad \phi \Rightarrow \phi' \quad \psi \Rightarrow \neg\psi' \quad \begin{matrix}\Gamma',\, \phi' \supset \mathsf{N}'\psi' \\ \Rightarrow \mathsf{N}'\psi'\end{matrix}}{\Gamma',\, \phi' \supset \mathsf{N}'\psi' \not\Rightarrow \mathsf{N}'\psi'}$$

For instance, POSPEC models permission as derogation [19]: a permission may suspend a more general obligation. Some further sequent elimination rules for handling conflicting sequents are listed in Figure 2. We will not further discuss them here but we will come back to them in Section 4.

Attacks between arguments are defined by the following notation and notion, referring to some $A \in \mathsf{Arg}(\Sigma)$:

- $\hat{A}$ denotes the top sequent in the proof A.
- We say that a sequent $\Gamma \Rightarrow \psi$ is a *subsequent of* A if it is contained in A, and $\mathsf{Prem}(\hat{A}) \vdash_{\mathsf{ISDL}} \bigwedge \Gamma$ (or, equivalently, if $\mathsf{Prem}(\hat{A}) \Rightarrow \bigwedge \Gamma \in \mathsf{Arg}(\Sigma)$).[3]

According to the next definition, an argument is attacked in some of its subsequents (including its top-sequent).

Definition 6. *Let* $R = \frac{\Gamma_1 \Rightarrow \phi_1 \ \dots \ \Gamma_n \Rightarrow \phi_n}{\Gamma_n \not\Rightarrow \phi_n}$ *be a sequent elimination rule in Figure 2, and let* $\mathcal{R}$ *be a set of such elimination rules.*

---

[2]Note that a 'PPSPEC'-variant would not be sensible since permissions with incompatible content do not conflict in any intuitive sense.

[3]Intuitively speaking, the second condition warrants that the subsequents of a proof A of $\mathfrak{s} = \Gamma \Rightarrow \psi$ are only those sequents whose premises are charged in the proof of $\mathfrak{s}$. Take for instance the proof of $\Rightarrow \phi \supset \phi$ from $\phi \Rightarrow \phi$ by $[\Rightarrow\supset]$. This prevents for instance attacks on A by $\neg\phi \Rightarrow \neg\phi$.

$$\text{CON}\quad \frac{\Rightarrow \neg\bigwedge\Gamma \quad \Gamma,\Gamma' \Rightarrow \psi}{\Gamma' \not\Rightarrow \psi} \qquad \text{NIC}\quad \frac{\Gamma \Rightarrow \neg\phi \quad \Gamma' \Rightarrow \mathsf{N}\phi}{\Gamma' \not\Rightarrow \mathsf{N}\phi}$$

$$\text{CONFU}\quad \frac{\Gamma \Rightarrow \neg\bigwedge\Gamma' \quad \Gamma',\Gamma'' \Rightarrow \psi}{\Gamma',\Gamma'' \not\Rightarrow \psi} \qquad \text{CONF}\quad \frac{\Gamma' \Rightarrow \psi' \quad \psi' \Rightarrow \neg\psi \quad \Gamma \Rightarrow \psi}{\Gamma \not\Rightarrow \psi}$$

$$\mathsf{NN'CONF}\quad \frac{\Gamma \Rightarrow \mathsf{N}\psi \quad \psi \Rightarrow \neg\psi' \quad \Gamma' \Rightarrow \mathsf{N}'\psi'}{\Gamma' \not\Rightarrow \mathsf{N}'\psi'}$$

$$\mathsf{NN'CONFU}\quad \frac{\Gamma,\ \phi \supset \mathsf{N}\psi \Rightarrow \mathsf{N}\psi \quad \Gamma \Rightarrow \phi \quad \psi \Rightarrow \neg\psi' \quad \Gamma',\ \phi' \supset \mathsf{N}'\psi' \Rightarrow \psi''}{\Gamma',\ \phi \supset \mathsf{N}'\psi' \not\Rightarrow \psi''}$$

$$\mathsf{NCONFU'}\quad \frac{\Gamma \Rightarrow \neg(\phi \supset \mathsf{N}\psi) \quad \Gamma',\ \phi \supset \mathsf{N}\psi \Rightarrow \psi'}{\Gamma,\ \phi \supset \mathsf{N}\psi \not\Rightarrow \psi'}$$

$$\mathsf{NCTD}\quad \frac{\begin{array}{c}\Gamma,\ \phi \supset \mathsf{N}\psi\\ \Rightarrow \mathsf{N}\psi\end{array} \quad \Gamma \Rightarrow \phi \quad \Gamma' \Rightarrow \phi' \quad \phi \Rightarrow \phi' \quad \psi \Rightarrow \neg\psi' \quad \begin{array}{c}\Gamma',\ \phi' \supset \mathsf{O}\psi'\\ \Rightarrow \mathsf{O}\psi'\end{array}}{\Gamma',\ \phi' \supset \mathsf{O}\psi' \not\Rightarrow \mathsf{O}\psi'}$$

$$\mathsf{NN'SPECU}\quad \frac{\begin{array}{c}\Gamma,\ \phi \supset \mathsf{N}\psi\\ \Rightarrow \neg(\phi' \supset \mathsf{N}'\psi')\end{array} \quad \Gamma \Rightarrow \phi \quad \phi \Rightarrow \phi' \quad \psi \Rightarrow \neg\psi' \quad \begin{array}{c}\Gamma',\ \phi' \supset \mathsf{N}'\psi'\\ \Rightarrow \psi''\end{array}}{\Gamma',\ \phi' \supset \mathsf{N}'\psi' \not\Rightarrow \psi''}$$

Figure 2: Some sequent elimination rules for normative reasoning (where $\mathsf{NN'} \in \{\mathsf{OO}, \mathsf{OP}, \mathsf{PO}\}$ and $\mathsf{N} \in \{\mathsf{O}, \mathsf{P}\}$)

- *A sequent* $\mathfrak{s}$ *R*-attacks *a sequent* $\mathfrak{s}'$, *if there is an* $\mathcal{L}_{\mathsf{ISDL}}$*-substitution* θ *such that* $\mathfrak{s} = \theta(\Gamma_1) \Rightarrow \theta(\phi_1)$ *and* $\mathfrak{s}' = \theta(\Gamma_n) \Rightarrow \theta(\phi_n)$. *We say that* $\mathfrak{s}$ $\mathcal{R}$-attacks $\mathfrak{s}'$ *if* $\mathfrak{s}$ *R-attacks* $\mathfrak{s}'$ *for some* $R \in \mathcal{R}$.

- *An argument* $A \in \mathsf{Arg}(\Sigma)$ *R*-attacks *an argument* $B \in \mathsf{Arg}(\Sigma)$ *if* $\hat{A}$ *R-attacks some subsequent of B. Similarly A* $\mathcal{R}$-attacks *B if A R-attacks B for some* $R \in \mathcal{R}$.

Definition 7. *A* normative argumentation framework *induced by a set of elimination rules* $\mathcal{R}$ *is the logical argumentation framework* $\mathcal{AF}_{\mathcal{R}}(\Sigma) = \langle \mathsf{Arg}(\Sigma), \mathit{Attack} \rangle$ *in which* $(A, B) \in \mathit{Attack}$ *iff A* $\mathcal{R}$*-attacks B.*

Normative Entailments Induced by Argumentation Frameworks

We are ready now to use (normative) argumentation frameworks for normative reasoning. As usual in the context of abstract argumentation, we do so by incorporating Dung's notion of extension [8], defined next.

Definition 8. *Let* $\mathcal{AF} = \langle \mathit{Args}, \mathit{Attack} \rangle$ *be an argumentation framework, and let* $\mathcal{E} \subseteq \mathit{Args}$. *We say that* $\mathcal{E}$ attacks *an argument A if there is an argument* $B \in \mathcal{E}$ *that attacks A (i.e.,* $(B, A) \in \mathit{Attack}$*). The set of arguments that are attacked by* $\mathcal{E}$ *is denoted* $\mathcal{E}^+$. *We say that* $\mathcal{E}$ defends *A if* $\mathcal{E}$ *attacks every argument B that attacks A. The set* $\mathcal{E}$ *is called* conflict-free *if it does not attack any of its elements (i.e.,* $\mathcal{E}^+ \cap \mathcal{E} = \emptyset$*),* $\mathcal{E}$ *is called* admissible *if it is conflict-free and defends all of its elements, and* $\mathcal{E}$ *is* complete *if it is admissible and contains all the arguments that it defends. The minimal complete subset of Args is called the* grounded extension *of* $\mathcal{AF}$, *and a maximal complete subset of Args is called a* preferred extension *of* $\mathcal{AF}$.

Let $\mathcal{AF}_{\mathcal{R}}(\Sigma) = \langle \mathsf{Arg}(\Sigma), \mathit{Attack} \rangle$ be a normative argumentation framework.

- $\Sigma \mathrel{|\!\sim}_{\mathsf{gr}} \psi$ if there is $A \in \mathsf{Arg}(\Sigma)$ in the grounded extension of $\mathcal{AF}_{\mathcal{R}}(\Sigma)$ such that $\hat{A} = \Gamma \Rightarrow \psi$.[4]

- $\Sigma \mathrel{|\!\sim}^{\cap}_{\mathsf{pr}} \psi$ [$\Sigma \mathrel{|\!\sim}^{\cup}_{\mathsf{pr}} \psi$] if in every [some] preferred extension of $\mathcal{AF}_{\mathcal{R}}(\Sigma)$ there is $A \in \mathsf{Arg}(\Sigma)$ with $\hat{A} = \Gamma \Rightarrow \psi$.[5]

We will use the notation $\mathrel{|\!\sim}$ whenever a statement applies to each of the defined consequence relations.

4 Some Examples

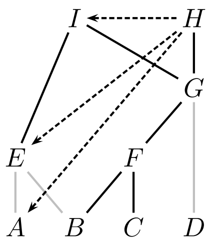
In this section we will demonstrate our argumentative model for normative reasoning by means of various examples.

Example 9. *Let us recall Example 1, where* $\Sigma = \{m, a, m \supset \mathsf{O}\neg f, (m \wedge a) \supset \mathsf{O}f\}$. *Some arguments in* $\mathsf{Arg}(\Sigma)$ *are listed in Figure 3 (right). We do not spell out the very simple proofs given by each argument but only list the top sequents and subsequent relationships. For instance, arguments A, B, C, D and E are one-liner proofs, argument F is obtained from B and C by weakening, etc. Figure 3 (left) shows an attack diagram where the only attack rule is* OOSPECU.

We observe that H OOSPECU*-attacks A and E, and since* $\hat{E}$ *is a subsequent of I, the latter is also attacked by H. It follows that, as expected, we have the following deductions:*

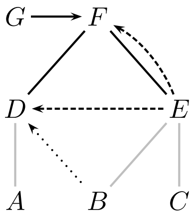
[4] Recall that by the definition of $\mathsf{Arg}(\Sigma)$, this implies that $\Gamma \subseteq \Sigma$.

[5] A more cautious approach is to define: $\Sigma \mathrel{|\!\sim}^{\cap}_{\mathsf{pr}} \psi$ [$\Sigma \mathrel{|\!\sim}^{\cup}_{\mathsf{pr}} \psi$] if there is an $A \in \mathsf{Arg}(\Sigma)$ with $\hat{A} = \Gamma \Rightarrow \psi$ that is in every [some] preferred extension of $\mathcal{AF}_{\mathcal{R}}(\Sigma)$. Similar entailment relations may of-course be defined for other semantics of abstract argumentation such as [semi-]stable semantics, ideal semantics, etc.



$$\begin{aligned}
\hat{A} &= m \supset \mathsf{O}\neg f \Rightarrow m \supset \mathsf{O}\neg f\\
\hat{B} &= m \Rightarrow m\\
\hat{C} &= a \Rightarrow a\\
\hat{D} &= (m \wedge a) \supset \mathsf{O}f \Rightarrow (m \wedge a) \supset \mathsf{O}f\\
\hat{E} &= m, m \supset \mathsf{O}\neg f \Rightarrow \mathsf{O}\neg f\\
\hat{F} &= m, a \Rightarrow m \wedge a\\
\hat{G} &= m, a, (m \wedge a) \supset \mathsf{O}f \Rightarrow \mathsf{O}f\\
\hat{H} &= m, a, (m \wedge a) \supset \mathsf{O}f \Rightarrow \neg(m \supset \mathsf{O}\neg f)\\
\hat{I} &= m, a, m \supset \mathsf{O}\neg f, (m \wedge a) \supset \mathsf{O}f \Rightarrow \mathsf{O}\bot
\end{aligned}$$

Figure 3: (Part of) the normative argumentation framework of Example 9: dashed arrows are OOSPECU-attacks, solid black lines indicate subsequents (the top sequents of lower arguments are subsequents of higher ones) and the gray line merely helps the reader to see which sequents share premises.



$$\begin{aligned}
\hat{A} &= \top \supset \mathsf{O}\neg k \Rightarrow \top \supset \mathsf{O}\neg k\\
\hat{B} &= k \Rightarrow k\\
\hat{C} &= k \supset \mathsf{O}(k \wedge g) \Rightarrow k \supset \mathsf{O}(k \wedge g)\\
\hat{D} &= \top \supset \mathsf{O}\neg k \Rightarrow \mathsf{O}\neg k\\
\hat{E} &= k, k \supset \mathsf{O}(k \wedge g) \Rightarrow \mathsf{O}(k \wedge g)\\
\hat{F} &= k, \top \supset \mathsf{O}\neg k, k \supset \mathsf{O}(k \wedge g) \Rightarrow \bot\\
\hat{G} &= \Rightarrow \neg(k \wedge (\top \supset \mathsf{O}\neg k) \wedge (k \supset \mathsf{O}(k \wedge g)))
\end{aligned}$$

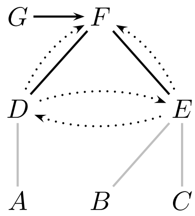


Figure 4: Forrester's Gentle Murderer

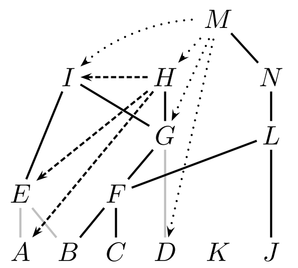
- $\Sigma \not\mathrel{|\!\sim} \mathsf{O}\neg f$. *Indeed, one cannot derive* $\mathsf{O}\neg f$ *since the application of MP to* $m \supset \mathsf{O}\neg f$ *(depicted by argument* E*) gets attacked by* H.[6]
- $\Sigma \mathrel{|\!\sim} \mathsf{O}f$. *Indeed,* G *is not* OOSPECU*-attackable by an argument in* $\mathsf{Arg}(\Sigma)$*, thus it is part of every grounded and preferred extension of the underlying normative argumentation framework, and so its descendent follows from* Σ.[7]

Example 10. *In the next example we take a look at contrary-to-duty (in short, CTD) obligations. A paradigmatic example is Forrester's Gentle Murderer scenario [20]: generally, one ought not to kill (*$\top \supset \mathsf{O}\neg f$*). However, upon killing, this should be done gently (*$k \supset \mathsf{O}(k \wedge g)$*). Let* $\Sigma_2 = \{k, \top \supset \mathsf{O}\neg k, k \supset \mathsf{O}(k \wedge g)\}$.

Van der Torre and Tan [21] distinguish CTD-obligations from cases of specificity. In the former the general obligations are not canceled or overridden but have still normative force (despite the fact that they are violated), while in cases of specificity the more general conditional obligations are canceled and thus deprived of normative force. There are various ways in which in our framework this distinction can be taken into account. One way of doing so is as follows. Instead of using strong rules such as OOSPECU *in Example 9 that 'destroy' overridden conditional obligations in the sense that they do not appear in the consequence set, we can make use of rules such as* OCTD *(Figure 2) that preserve 'overshadowed' conditional CTD obligations de-*

[6] Note that $m \supset \mathsf{O}\neg f$ cannot be derived either, due to the attack of H on A.

[7] It is important to note that G *is* OOSPECU-attackable by ISDL-derivable arguments, but none of them is in $\mathsf{Arg}(\Sigma)$. For instance, since intuitionistic implication allows for strengthening of antecedents ($\phi \supset \psi \Rightarrow (\phi \wedge \phi') \supset \psi$), we have that $m \supset \mathsf{O}\neg f \Rightarrow (m \wedge a) \supset \mathsf{O}\neg f$ is ISDL-derivable, and so G is attackable by an argument with, say, the ISDL-derivable top sequent $m, m \supset \mathsf{O}\neg f, m, a, (m \wedge a) \supset \mathsf{O}\neg f \Rightarrow \neg((m \wedge a) \supset \mathsf{O}\neg f)$. Yet, since $m \wedge a \supset \mathsf{O}\neg f \notin \Sigma$, this argument is not in $\mathsf{Arg}(\Sigma)$. We note, further, that the sequent $a, m, m \supset \mathsf{O}\neg f \Rightarrow \neg((m \wedge a) \supset \mathsf{O}f)$ is derivable, but it does not OOSPECU-attack $\hat{G}$ and $\hat{H}$ though it is attacked by $\hat{H}$.



$\hat{D} = (m \wedge a) \supset \mathsf{P}f \Rightarrow (m \wedge a) \supset \mathsf{P}f$
$\hat{G} = m, a, (m \wedge a) \supset \mathsf{P}f \Rightarrow \mathsf{P}f$
$\hat{H} = m, a, (m \wedge a) \supset \mathsf{P}f \Rightarrow \neg(m \supset \mathsf{O}\neg f)$
$\hat{I} = m, a, m \supset \mathsf{O}\neg f, (m \wedge a) \supset \mathsf{P}f \Rightarrow \mathsf{O}\bot$
$\hat{J} = c \Rightarrow c$
$\hat{K} = (m \wedge a \wedge c) \supset \mathsf{O}\neg f \Rightarrow (m \wedge a \wedge c) \supset \mathsf{O}\neg f$
$\hat{L} = m, a, c \Rightarrow m \wedge a \wedge c$
$\hat{M} = m, a, c, (m \wedge a \wedge c) \supset \mathsf{O}\neg f \Rightarrow \neg((m \wedge a) \supset \mathsf{P}f))$
$\hat{N} = m, a, c, (m \wedge a \wedge c) \supset \mathsf{O}\neg f \Rightarrow \mathsf{O}\neg f$

Figure 5: A normative argumentation framework for Example 11 (arguments A, B, C, E, F are as in Figure 3)

spite the fact that detachment is blocked, or incorporate OIC *that blocks detachment from violated norms. This is illustrated in Figure 4 (left) with the attack rules* OCTD *(dashed arrow),* OIC *(dotted arrow) and* CON *(solid arrow). Alternatively, we could model overshadowing by means of* OOCONF *instead of* OCTD*. This is illustrated in Figure 4 (right) with attack rules* OOCONF *(dotted arrows) and* CON *(solid arrow). Where $\Xi = \{A, B, C, G\}$, we have two preferred extensions: $\Xi \cup \{D\}$ and $\Xi \cup \{E\}$. Hence, $\Sigma_2 \mid\!\sim^{\cup}_{\mathsf{pr}} \mathsf{O}\neg k$ and $\Sigma_2 \mid\!\sim^{\cup}_{\mathsf{pr}} \mathsf{O}(k \wedge g)$. In the skeptical approach we get $\Sigma_2 \mid\!\sim^{\cap}_{\mathsf{pr}} \mathsf{O}(\neg k \vee (k \wedge g))$ and $\Sigma_2 \mid\!\sim^{\cap}_{\mathsf{pr}} \mathsf{O}\neg k \vee \mathsf{O}(k \wedge g)$. Yet another option is to use a very liberal approach with* CON *only. This will block arguments with inconsistent premises such as F but otherwise allows e.g., to derive both $\mathsf{O}\neg k$ and $\mathsf{O}(g \wedge k)$ even via the grounded approach: $\Sigma_2 \mid\!\sim_{\mathsf{gr}} \mathsf{O}\neg k$ and $\Sigma_2 \mid\!\sim_{\mathsf{gr}} \mathsf{O}(k \wedge g)$.*

Example 11. *Let us consider a variant of Example 9. Suppose that beside the obligation not to eat with your fingers we have the permission to do so in case asparagus is served, but it is considered impolite to eat asparagus with fingers if there is guest who considers this rude. The enriched set of premises may look as follows: $\Sigma_3 = \{a, m, c, m \supset \mathsf{O}\neg f, (m \wedge a) \supset \mathsf{P}f, (m \wedge a \wedge c) \supset \mathsf{O}\neg f\}$. The situation is depicted in Figure 5, where the attack rules* OPSPECU *(dotted arrows) and* POSPECU *(dashed arrows).*

It follows that $\Sigma_3 \mid\!\sim \mathsf{O}\neg f$ (as expected), since N is defended, while G cannot be defended. Note that arguments A and E are also defended, since their only attacker H is attacked by the defended M. In argumentation theory A and E are said to be reinstated.[8]

Example 12. *Next we take a look at a simple conflict that is neither a specificity nor a CTD-case. Let $\Sigma_4 = \{a, b, a \supset \mathsf{O}(c \wedge d), b \supset \mathsf{O}(\neg c \wedge d)\}$. Figure 6 shows the situation for the attack rule* OOCONFU *(dotted arrows).*

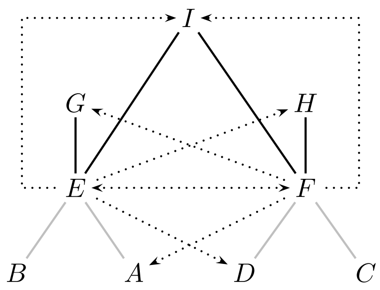
We have the following preferred extensions: $\{A, B, E, G\}$ and $\{C, D, F, H\}$. Note that we have the 'floating conclusion[9] *$\Sigma_4 \mid\!\sim^{\cap}_{\mathsf{pr}} \mathsf{O}d$ since one of G and H is in every preferred extension.*

Example 13. *The next example illustrates a conflict between three obligations. Let $\Sigma_5 = \{c, c \supset \mathsf{O}(a \vee b), c \supset \mathsf{O}(\neg a \vee b), c \supset \mathsf{O}\neg a\}$. It is interesting to note that modeling this scenario with* OOCONFU *is problematic. In this case no conflicts are triggered since the triple-conflict is not reducible to a binary conflict that fits the attack rule* OOCONFU*. This may be avoided by using* OCONFU' *instead of* OOCONFU*, as we get for instance $\Sigma_5 \mid\!\sim^{\cap}_{\mathsf{pr}} \mathsf{O}a \vee \mathsf{O}(\neg a \wedge b) \vee \mathsf{O}(\neg a \wedge \neg b)$. This example shows that elimination rules should be carefully chosen.*[10]

[8]In the full version of the paper we will discuss how Reinstatement may be avoided (if necessary) by altering the attack rules.

[9]In nonmonotonic reasoning *floating conclusions* are conclusions that are obtained from each of a set of otherwise conflicting arguments.

[10]In Section 5 we will prove that OCONFU' is rather well-behaved and can be used to give an argumentative



$\hat{A} = a \supset \mathsf{O}(c \wedge d) \Rightarrow a \supset \mathsf{O}(c \wedge d)$
$\hat{B} = a \Rightarrow a$
$\hat{C} = b \Rightarrow b$
$\hat{D} = b \supset \mathsf{O}(\neg c \wedge d) \Rightarrow b \supset \mathsf{O}(\neg c \wedge d)$
$\hat{E} = a, a \supset \mathsf{O}(c \wedge d) \Rightarrow \mathsf{O}(c \wedge d)$
$\hat{F} = b, b \supset \mathsf{O}(\neg c \wedge d) \Rightarrow \mathsf{O}(\neg c \wedge d)$
$\hat{G} = a, a \supset \mathsf{O}(c \wedge d) \Rightarrow \mathsf{O}d$
$\hat{H} = b, b \supset \mathsf{O}(\neg c \wedge d) \Rightarrow \mathsf{O}d$
$\hat{I} = a, b, a \supset \mathsf{O}(c \wedge d), b \supset \mathsf{O}(\neg c \wedge d) \Rightarrow \mathsf{O}\bot$

Figure 6: A simple conflict

5 Some Meta-Theory

We start with two basic observations which can easily be verified by the reader:

1. For any set of attack rules previously defined: whenever Σ is ISDL-consistent (i.e., $\Sigma \nvdash_{\mathsf{ISDL}} \bot$) then $\Sigma \vdash_{\mathsf{ISDL}} \psi$ iff $\Sigma \mid\!\sim \psi$. It is easy to verify that in this case all arguments in $\mathsf{Arg}(\Sigma)$ are selected since no argumentative attacks occur.

2. Where CON is part of the attack rules, (i) $\Sigma \mid\!\sim \phi$ implies that ϕ is ISDL-consistent (i.e., $\phi \nvdash_{\mathsf{ISDL}} \bot$) and, consequently, (ii) $\mid\!\sim$ is strongly paraconsistent (i.e., for all Σ, $\Sigma \not\mid\!\sim \bot$).

The main goal of this setion is to provide a link between of our approach to Input/Output logic [22] (see Theorem 20, Corollary 21, and Note 23 below). For this, we first recall the following semantic characterization of ISDL by means of a Kripkean possible worlds semantics [23].[11]

Definition 14. *An* IL*-model* M *is a tuple* $\langle W, \leq, v, @\rangle$*, where* W *is a nonempty set (of so-called worlds),* $\leq$ *is a partial order on* W*,* $@ \in W$ *is the so-called actual world, and* $v : W \to \wp(\mathcal{A})$ *(where* $\mathcal{A}$ *is the set of atomic formulas) is an assignment function that satisfies:*

(Fd) if $a \leq b$ *then* $v(a) \subseteq v(b)$*.*

For some $a \in W$*, we define:*

(M1) $M, a \models \rho$ *where* $\rho \in \mathcal{A}$ *iff* $\rho \in v(a)$
(M2) $M, a \models \psi \vee \phi$ *iff* $M, a \models \psi$ *or* $M, a \models \phi$
(M3) $M, a \models \psi \wedge \phi$ *iff* $M, a \models \psi$ *and* $M, a \models \phi$
(M4) $M, a \models \neg\psi$ *iff for all* $b \geq a$*,* $M, b \not\models \psi$
(M5) $M, a \models \psi \supset \phi$ *iff for all* $b \geq a$*,* $M, b \models \psi$ *implies* $M, b \models \phi$

We say that M *is an* IL-model of ψ *(*$M \models \psi$*) iff* $M, @ \models \psi$*.* M *is an* IL*-model of* Σ *if it is an* IL*-model of every* $\psi \in \Sigma$*. The set of all* IL*-models of* Σ *is denoted* $\mathcal{M}_{\mathsf{IL}}(\Sigma)$*. We also write* $\mathrm{Cn}_{\mathsf{IL}}(\Gamma) =_{\mathrm{df}} \{\psi \mid \Gamma \vdash_{\mathsf{IL}} \psi\}$*.*

Definition 15. *An* ISDL*-model* M *is a tuple* $\langle W, R, \leq, v, @\rangle\rangle$*, where* $\langle W, \leq, v, @\rangle$ *is an* IL*-model and* R *is a serial accessibility relation on* W*. In addition to (M1)–(M5) we have (where* $Rc =_{\mathsf{df}} \{d \in W \mid (c, d) \in R\}$*):*

(MO*)* $M, a \models \mathsf{O}\psi$ *iff for all* $c \geq a$ *and all* $b \in Rc$*,* $M, b \models \psi$*.*

account of a specific Input/Output logic.

[11] In [23] the reader can also find refinements of this semantics by frame-conditions and by letting P be primitive to get e.g. $\mathsf{P}(\phi \vee \psi) \vdash \mathsf{P}\phi \vee \mathsf{P}\psi$ or $\neg\mathsf{P}\phi \vdash \mathsf{O}\neg\phi$. With Definition 15 and $\mathsf{P}\phi =_{\mathsf{df}} \neg\mathsf{O}\neg f$ we get e.g. $\mathsf{O}\phi, \mathsf{P}\psi \vdash_{\mathsf{ISDL}} \mathsf{P}(\phi \wedge \psi)$, $\mathsf{P}\phi \vdash_{\mathsf{ISDL}} \mathsf{P}(\phi \vee \psi)$.

Again, we write $M \models \psi$ *iff* $M, @ \models \psi$ *and say that* M *is an* ISDL*-model of* ψ*.* $\mathcal{M}_{\mathsf{ISDL}}(\Sigma)$ *is the set of all* ISDL*-models of* Σ*.*

In the following we focus on premise sets Σ that consist of non-modal formulas (representing 'facts' or 'input') and formulas of the type $\phi \supset \mathsf{O}\psi$ (representing conditional obligations). For this let Σ_F be a set of non-modal propositional formulas, Σ_O a set of pairs of non-modal formulas (ψ, ϕ) ('I/O-pairs') and $\Sigma_O^* = \{\psi \supset \mathsf{O}\phi \mid (\psi, \phi) \in \Sigma_O\}$. The following definitions describe intuitionistic versions of the 'out' and the 'out$_2$'-function in [24]:

Definition 16. $\mathsf{out}(\Sigma_F, \Sigma_O) = \{\psi \mid (\phi, \psi) \in \Sigma_O,\ \Sigma_F \vdash_{\mathsf{IL}} \phi\}$.

Let $M = \langle W, \leq, v, @\rangle$. We write $w\!\uparrow = \{w' \in W \mid w \leq w'\}$, $V_M = \{\psi \mid M \models \psi\}$ and where $w \in W$, $V_w = \{\phi \mid M, w \models \phi\}$. We say that M is *consistent with* Σ_O iff for all $w \in @\!\uparrow$, $\mathsf{out}(V_w, \Sigma_O)$ is IL-consistent.

Definition 17. $\phi \in \mathsf{out}_2(\Sigma_F, \Sigma_O)$ *iff* $\phi \in \mathrm{Cn}_{\mathsf{IL}}(\mathsf{out}(V_M, \Sigma_O))$ *for all* $M \in \mathcal{M}_{\mathsf{IL}}(\Sigma_F)$ *that are consistent with* Σ_O*. If there are no* $M \in \mathcal{M}_{\mathsf{IL}}(\Sigma_F)$ *that are consistent with* Σ_O *then define* $out_2(\Sigma_F, \Sigma_O)$ *to be* $\mathrm{Cn}_{\mathsf{IL}}(\{\psi \mid (\psi', \psi) \in \Sigma_O\})$*.*[12]

Theorem 18. $\Sigma_F \cup \Sigma_O^* \Vdash_{\mathsf{ISDL}} \mathsf{O}\phi$ *iff* $\phi \in \mathsf{out}_2(\Sigma_F, \Sigma_O)$.

Sketch. We will make use of the following simple fact (the proof of which is left to the reader):

(†) Let $M = \langle W, \leq, v, @\rangle$ be an IL-model and $M' = \langle @\!\uparrow, \leq_{|(@\uparrow \times @\uparrow)}, v_{|@\uparrow}, @\rangle$ its submodel, restricted to $@\!\uparrow$. Then $M \models \psi$ iff $M' \models \psi$.

($\Rightarrow$) Suppose $\phi \notin \mathsf{out}_2(\Sigma_F, \Sigma_O)$. We show that $\Sigma_F \cup \Sigma_O^* \nVdash_{\mathsf{ISDL}} \mathsf{O}\phi$ by constructing a model $M^* \in \mathcal{M}_{\mathsf{ISDL}}(\Sigma_F \cup \Sigma_O^*)$ for which $M^* \not\models \mathsf{O}\phi$. By the supposition there is an $M \in \mathcal{M}_{\mathsf{IL}}(\Sigma_F)$ that is consistent with Σ_O and for which $\phi \notin \mathrm{Cn}_{\mathsf{IL}}(\mathsf{out}(V_M, \Sigma_O))$. By (†), we can suppose that $M = \langle @\!\uparrow, \leq, v, @\rangle$ (where $\leq$ and v are restricted to $@\!\uparrow$). Since $\phi \notin \mathrm{Cn}_{\mathsf{IL}}(\mathsf{out}(V_M, \Sigma_O))$, there is an $M_@ \in \mathcal{M}_{\mathsf{IL}}(\mathsf{out}(V_M, \Sigma_O))$ such that $M_@ \not\models \phi$. With (†) we suppose that $M_@$ is of the form $\langle \underline{@}\!\uparrow, \leq_@, v_@, \underline{@}\rangle$. Where $a \in @\!\uparrow \setminus \{@\}$, let $M_a \in \mathcal{M}_{\mathsf{IL}}(\mathsf{out}(V_a, \Sigma_O))$ of the form $\langle \underline{a}\!\uparrow, \leq_a, v_a, \underline{a}\rangle$. We suppose that (‡) $@\!\uparrow$, $\underline{a}\!\uparrow$ and $\underline{b}\!\uparrow$ (where $a \neq b \in @\!\uparrow$) are distinct sets of worlds. We define an ISDL-model $M^* = \langle W^*, R^*, \leq^*, v^*, @\rangle$ as in Figure 7.[13] We now show that $M^* \models \Sigma_F \cup \Sigma_O^*$ and $M^* \not\models \mathsf{O}\phi$.

$$W^* = @\!\uparrow \cup \textstyle\bigcup_{a \in @\uparrow} \underline{a}\!\uparrow,$$
$$v^* = v \cup \textstyle\bigcup_{a \in @\uparrow} v_a,$$
$$R^* = \textstyle\bigcup_{a \in @\uparrow} \big(\{(a, \underline{a}), (\underline{a}, \underline{a})\} \cup \bigcup_{b \in \underline{a}\uparrow} \{(b, b)\}\big),$$
$$\leq^* = \leq \cup \textstyle\bigcup_{a \in @\uparrow} \leq_a.$$

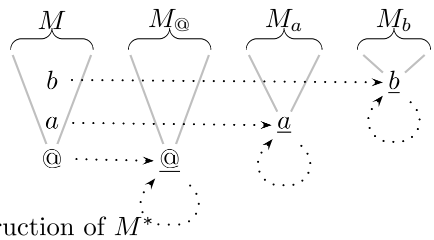


Figure 7: The construction of M^*

[12] According to the original definition of out_2 relative to classical logic, $\phi \in \mathsf{out}_2(\Sigma_F, \Sigma_O)$ iff for all classical models M of Σ_F for which $\mathsf{out}(V_M, \Sigma_O)$ is (classically) non-trivial, $\phi \in \mathsf{out}(V_M, \Sigma_O)$ (for the non-degenerated case). Our requirement that $M \in \mathcal{M}_{\mathsf{IL}}(\Sigma_F)$ is consistent with Σ_O is slightly more complicated. One (common) way to think about worlds in $@\!\uparrow \setminus \{@\}$ is in terms of possible increased information states accessible from the actual world. In this reading the requirement that M is consistent with Σ_O says that the output remains consistent for each possible increased information state. Take e.g., $\Sigma_O = \{(\rho, \bot), (\neg\rho, \bot)\}$ (where ρ is an atom) and $\Sigma_F = \emptyset$. There are no $M \in \mathcal{M}_{\mathsf{IL}}(\Sigma_F)$ that are consistent with Σ_O (for the same reason that $\mathcal{M}_{\mathsf{IL}}(\{\rho \supset \bot, \neg\rho \supset \bot\}) = \emptyset$). Clearly, there are $M \in \mathcal{M}_{\mathsf{IL}}(\Sigma_F)$ such that $\mathsf{out}(V_M, \Sigma_O)$ is IL-consistent. However, in each such M there is a w accessible from the actual world such that $\mathsf{out}(V_w, \Sigma_O)$ is IL-trivial.

[13] It is easy to see that M^* is an ISDL-model: First, in view of (‡) it is easy to check that the resulting relation v^* is right-unique and left-total and hence a function. (Fd) is fulfilled since $\leq$ and $\leq_a$ (where $a \in @\!\uparrow$) fulfill (Fd) and the latter relations partition $\leq^*$ by (‡). Finally, R^* is serial by the construction.

First notice that by the definition of $\leq^*$ and (‡), for every non-modal formula ψ: (i) $M^*, @ \models \psi$ iff $M, @ \models \psi$ and (ii) where $a \in @\uparrow$, $M_a, \underline{a} \models \psi$ iff $M^*, \underline{a} \models \psi$. An immediate consequence of (i) is that $M^* \models \Sigma_F$. Also, by (ii), since $R@ = \{\underline{@}\}$ and $M_@ \not\models \phi$, we have $M^* \not\models \mathsf{O}\phi$.

We now show that $M^*, \underline{a} \models \mathsf{O}\psi$ for all $\psi \in \mathsf{out}(V_a, \Sigma_O)$ and all $a \in @\uparrow$. Let for this $b \in W^*$ such that $a \leq^* b$. By the construction of M^* this means that $b \in @\uparrow$ and $Rb = \{\underline{b}\}$. By (Fd), $v(b) \supseteq v(a)$ and hence $\mathsf{out}(V_a, \Sigma_O) \subseteq \mathsf{out}(V_b, \Sigma_O)$. Thus, $\psi \in \mathsf{out}(V_b, \Sigma_O)$ and by (ii), $M^*, \underline{b} \models \psi$. Altogether, this shows that $M^*, a \models \mathsf{O}\psi$. This immediately shows that $M^* \models \Sigma_O^*$.

($\Leftarrow$) Suppose $\Sigma_F \cup \Sigma_O^* \not\Vdash_{\mathsf{ISDL}} \mathsf{O}\phi$. Hence, there is an $M = \langle W, R, \leq, v, @\rangle \in \mathcal{M}_{\mathsf{ISDL}}(\Sigma_F \cup \Sigma_O^*)$, for which $M \not\models \mathsf{O}\phi$. Thus, there is an $a \in R@$, such that $M, a \not\models \phi$. Let $M_a = \langle W, \leq, v, a\rangle$ and $M_@ = \langle W, \leq, v, @\rangle$. Since $M \in \mathcal{M}_{\mathsf{ISDL}}(\Sigma_F \cup \Sigma_O^*)$, it is easy to see that $M_@ \in \mathcal{M}_{\mathsf{IL}}(\Sigma_F)$ is consistent with Σ_O. Then, $M_a \in \mathcal{M}_{\mathsf{IL}}(\mathsf{out}(V_{M_@}, \Sigma_O))$ and $M_a \not\models \phi$. This shows that $\phi \notin \mathsf{out}_2(\Sigma_F, \Sigma_O)$. □ □

In order to deal with situations in which $\mathsf{out}(\Sigma_F, \Sigma_O)$ is inconsistent, Makinson and Van Der Torre [22] 'contextualize' their output-functions to maximal sets of conditionals that are consistent with Σ_F, so-called maxfamilies:[14]

Definition 19.

- $\Gamma_O \in \mathsf{maxfamily}(\Sigma_F, \Sigma_O)$ *iff* $\mathsf{out}_2(\Sigma_F, \Gamma_O)$ *is* IL-*consistent and for all* $(\psi, \phi) \in \Sigma_O \setminus \Gamma_O$, $\mathsf{out}_2(\Sigma_F, \Gamma_O \cup \{(\psi, \phi)\})$ *is not* IL-*consistent.*
- $\psi \in \mathsf{out}_2^{\cup}(\Sigma_F, \Sigma_O)$ *iff* $\psi \in \bigcup_{\Gamma_O \in \mathsf{maxfamily}(\Sigma_F, \Sigma_O)} \mathsf{out}_2(\Sigma_F, \Gamma_O)$.
- $\psi \in \mathsf{out}_2^{\cap}(\Sigma_F, \Sigma_O)$ *iff* $\psi \in \bigcap_{\Gamma_O \in \mathsf{maxfamily}(\Sigma_F, \Sigma_O)} \mathsf{out}_2(\Sigma_F, \Gamma_O)$.

We now show that in our argumentative approach the Input/Output logics in Definition 19 can be characterized by means of the attack rule OCONFU'.

Theorem 20. $\{\mathsf{Arg}(\Sigma_F \cup \Gamma_O^*) \mid \Gamma_O \in \mathsf{maxfamily}(\Sigma_F, \Sigma_O)\}$ *is the set of all preferred extensions of* $\mathcal{AF}_{\mathsf{OCONFU'}}(\Sigma_F \cup \Sigma_O^*)$.

Sketch. Let $\Gamma_O \in \mathsf{maxfamily}(\Sigma_F, \Sigma_O)$. By Theorem 18, $\Sigma_F \cup \Gamma_O^*$ is ISDL-consistent and hence $\mathsf{Arg}(\Sigma_F \cup \Gamma_O^*)$ is conflict-free. Thus, each argument A attacking any argument in $\mathsf{Arg}(\Sigma_F \cup \Gamma_O^*)$ is such that $A \notin \mathsf{Arg}(\Sigma_F \cup \Gamma_O^*)$. Let $A \in \mathsf{Arg}(\Sigma_F \cup \Sigma_O^*) \setminus \mathsf{Arg}(\Sigma_F \cup \Gamma_O^*)$. This means that there is a $\psi \supset \mathsf{O}\phi \in \mathsf{Prem}(\hat{A}) \cap (\Sigma_O^* \setminus \Gamma_O^*)$. Since $\mathsf{out}_2(\Sigma_F, \Gamma_O \cup \{(\psi, \phi)\})$ is IL-inconsistent we have by Theorem 18 that $\Sigma_F \cup \Gamma_O^* \cup \{\psi \supset \mathsf{O}\phi\}$ is ISDL-inconsistent. Thus, there is a finite $\Theta \subseteq \Sigma_F \cup \Gamma_O^*$ such that $\Theta, \psi \supset \mathsf{O}\phi \Rightarrow \bot$ is $\mathcal{C}_{\mathsf{ISDL}}$-provable. By $[\Rightarrow\supset]$, we derive $\mathfrak{s} = \Theta \Rightarrow \neg(\psi \supset \mathsf{O}\phi)$. Let C be the corresponding proof with $\hat{C} = \mathfrak{s}$. Then $C \in \mathsf{Arg}(\Sigma_F \cup \Gamma_O^*)$ and C OCONFU'-attacks A. We have shown that $\mathsf{Arg}(\Sigma_F \cup \Gamma_O^*)$ is defended and that it is maximally so.

Now assume there is an admissible extension Ξ of $\mathcal{AF}_{\mathsf{OCONFU'}}(\Sigma_F \cup \Sigma_O^*)$ such that there is no $\Gamma_O \in \mathsf{maxfamily}(\Sigma_F, \Sigma_O)$ for which $\Xi \subseteq \mathsf{Arg}(\Sigma_F \cup \Gamma_O^*)$. Hence, there is no $\Gamma_O \in \mathsf{maxfamily}(\Sigma_F, \Sigma_O)$ for which $\Gamma_\Xi = \bigcup_{A \in \Xi}\{(\psi, \phi) \mid \psi \supset \mathsf{O}\phi \in \mathsf{Prem}(\hat{A})\} \subseteq \Gamma_O$. This means $\mathsf{out}_2(\Sigma_F, \Gamma_\Xi)$ is IL-inconsistent. By Theorem 18, $\Sigma_F \cup \Gamma_\Xi^*$ is ISDL-inconsistent. Hence, there are finite $\Theta_F \subseteq \Sigma_F$ and $\Theta_O^* \subseteq \Gamma_\Xi^*$ such that $\Theta_F, \Theta_O^* \Rightarrow \bot$ is $\mathcal{C}_{\mathsf{ISDL}}$-derivable. With Weakening and $[\Rightarrow\supset]$ we have an argument C, with $\hat{C} = \Theta_F, \Theta_O^* \setminus \{\psi \supset \mathsf{O}\phi\} \Rightarrow \neg(\psi \supset \mathsf{O}\phi)$. By the subformula property we can suppose that:[15]

(†) for all $\gamma \supset \mathsf{O}\gamma'$ that occur in subsequents of C, $(\gamma, \gamma') \in \Theta_O$.

Then C OCONFU'-attacks A. Also, by (†), the only way to attack C leads to an attack on Ξ as well. Thus, Ξ cannot be defended from C. □ □

[14] The approach in [22] is more general since it takes into account sets of additional constraints beside our requirement of consistency.

[15] Similar considerations to those in [18] (where cut elimination has been shown for the modal enrichment of LK) show that $\mathcal{C}_{\mathsf{ISDL}}$ has the subformula property.

Corollary 21. *Where the only attack rule is* $\mathsf{OCONFU'}$*, for every* $\lambda \in \{\cup, \cap\}$ *it holds that* $\psi \in \mathsf{out}_2^{\lambda}(\Sigma_F, \Sigma_O)$ *iff* $\Sigma_F \cup \Sigma_O^* \mathrel{|\!\sim}^{\lambda}_{\mathsf{pr}} \mathsf{O}\psi$*.*

Example 22. *Let us look once more at Example 10. Let* $\Sigma_O = \{(\top, \neg k), (k, k \wedge g)\}$ *and* $\Sigma_F = \{k\}$*. We have* $\mathsf{maxfamily}(\Sigma_F, \Sigma_O) = \{\{(\top, \neg k)\}, \{k, k \wedge g\}\}$*. Since* $\neg k \vee (k \wedge g) \in \mathsf{out}_2(\Sigma_F, \{(\top, \neg k)\}) \cap \mathsf{out}_2(\Sigma_F, \{(k, k \wedge g)\})$*, also* $\neg k \vee (k \wedge g) \in \mathsf{out}_2^{\cap}(\Sigma_F, \Gamma_O)$*. In the normative argumentation framework* $\mathcal{AF}_{\mathsf{OCONFU'}}(\Sigma_F, \Sigma_O^*)$ *we have two preferred extensions: one with e.g. arguments with top sequents* $k, k \supset \mathsf{O}(k \wedge g) \Rightarrow \neg(\top \supset \mathsf{O}\neg k)$*,* $k, k \supset \mathsf{O}(k \wedge g) \Rightarrow \mathsf{O}(k \wedge g)$*, and* $k, k \supset \mathsf{O}(k \wedge g) \Rightarrow \mathsf{O}(\neg k \vee (k \wedge g))$*; and another one with e.g. arguments with top sequents* $k, \top \supset \mathsf{O}\neg k \Rightarrow \neg(k \supset \mathsf{O}(k \wedge g))$*,* $\top \supset \mathsf{O}\neg k \Rightarrow \mathsf{O}\neg k$*, and* $\top \supset \mathsf{O}\neg k \Rightarrow \mathsf{O}(\neg k \vee (k \wedge g))$*. Thus,* $\Sigma_F \cup \Sigma_O^* \mathrel{|\!\sim}^{\cap}_{\mathsf{pr}} \mathsf{O}(\neg k \vee (k \wedge g))$*.*

Note 23. *It is well-known that, by Glivenko's transformation* $A \rightsquigarrow \neg\neg A$*, classical logic is embedded in* IL*. By Corollary 21, then, the translation of* Σ_F *to* $\{\neg\neg\psi \mid \psi \in \Sigma_F\}$ *and* Σ_O *to* $\{(\neg\neg\phi, \neg\neg\psi) \mid (\phi, \psi) \in \Sigma_O\}$ *gives a characterization of classical Input/Output logic within our account.*

Further investigations of entailment relations resulting from the application of attack rules other than OCONFU' will be considered in a future work.

6 Discussion and Outlook

The idea to use argumentation and abstract argumentation in particular to model normative reasoning is not new. Two examples are [25, 26]. The approach in [25] is based on bipolar abstract argumentation frameworks: beside an attack arrow a support arrow is used to express conditional obligations. Also in [26] Dung's framework is enhanced by a support relation this time signifying evidential support. Prolog-like predicates are used to encode argument schemes of normative reasoning and an algorithm is provided to translate them into an argumentation framework. One of the main differences in our approach based on logical argumentation is that we use a base logic (ISDL) that generates all the given arguments (on the basis of a premise set). As a consequence an additional support relation is not needed since argumentative support is intrinsically modeled by considering arguments as proofs in ISDL. A by-product of this is that our approach is closer linked to deontic logic.

Deontic logicians mainly agree that modeling conditional obligations on the basis of SDL and material implication is futile due to problems with CTD-norms and specificity [1]. Therefore more research interest has been directed towards bi-conditionals. Specificity cases for instance call for weakened principles of strengthening the antecedent which are still strong enough to support many intuitively valid inferences. E.g., the principle of Rational Monotonicity has been challenged in [27] and replaced by a weakened version which itself has been criticized in [28]. In contrast, our base logic uses the standard implication of IL to model conditional obligations and allows for full strengthening of the antecedent. Unwanted applications of the latter are avoided by means of argumentative attacks that are triggered e.g. in cases of specificity. As a consequence, our consequence relations are non-monotonic. There are other non-monotonic accounts of normative reasoning such as [6] based on default logic, Input/Output logic [22], or adaptive logics [2, 29, 4, 5]. Due to space restrictions we postpone a more elaborate comparison with these frameworks to future work.

In future work we also plan to investigate ways to combine and prioritize among attack rules, to distinguish preferences/priorities among norms, and to relate our work to different accounts of permission [30, 19]. Finally, we will investigate whether other nonmonotonic approaches and non truth-functional logics can be expressed in our framework.[16] Also, we shall examine base logics that are obtained from ISDL by removing some of the inference rules in $\mathcal{C}_{\mathsf{ISDL}}$, and so such logics may not have deterministic matrices. There is also forthcoming work on dynamic proofs for sequent-based argumentation [14], which may be useful to automatize normative reasoning as modeled in this paper.

References

[1] Aqvist, L.: Deontic logic. In Gabbay, D.M., Guenthner, F., eds.: Handbook of Philosophical Logic, Vol.8 (2nd edition). Kluwer (2002) 147–264

[2] Beirlaen, M., Straßer, C., Meheus, J.: An inconsistency-adaptive deontic logic for normative conflicts. Journal of Philosophical Logic **2**(42) (2013) 285–315

[3] Carmo, J., Jones, A.J.I.: Deontic logic and contrary-to-duties. In Gabbay, D.M., Guenthner, F., eds.: Handbook of Philosophical Logic, Vol.8 (2nd edition). Kluwer (2002) 265–343

[4] Straßer, C.: A deontic logic framework allowing for factual detachment. Journal of Applied Logic **9**(1) (2010) 61–80

[5] Goble, L.: Deontic logic (adapted) for normative conflicts. Logic Journal of the IGPL (2013)

[6] Horty, J.F.: Reasoning with moral conflicts. Noûs **37**(4) (2003) 557–605

[7] Straßer, C., Beirlaen, M.: An Andersonian deontic logic with contextualized sanctions. In: Deontic Logic in Computer Science. Springer (2012) 151–169

[8] Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artifical Intelligence **77** (1995) 321–358

[9] Dung, P.M., Kowalski, R.A., Toni, F.: Assumption-based argumentation. Argumentation in Artificial Intelligence (2009) 199–218

[10] Prakken, H.: An abstract framework for argumentation with structured arguments. Argument and Computation **1**(2) (2011) 93–124

[11] Arieli, O.: A sequent-based representation of logical argumentation. In: Proc. CLIMA'13. LNCS 8143, Springer (2013) 69–85

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

[13] Straßer, C.: Adaptive Logic and Defeasible Reasoning. Applications in Argumentation, Normative Reasoning and Default Reasoning. Volume 38 of Trends in Logic. Springer (2014)

[14] Arieli, O., Straßer, C.: Dynamic derivations for sequent-based logical argumentation (Submitted).

[15] Gentzen, G.: Investigations into logical deduction (1934) (In German. An English translation appears in 'The Collected Works of Gerhard Gentzen', edited by M. E. Szabo, North-Holland, 1969).

[16] For instance, we will check whether out_4 from [22] may be expressible by using $\Gamma^{\star}_{O} = \Gamma^{*}_{O} \cup \{\mathsf{O}\phi \supset \mathsf{O}\psi \mid \phi \supset \psi \in \Gamma_O\}$ instead of Γ^{*}_{O} as in Corollary 21.

[16] Horty, J.F.: Moral dilemmas and nonmonotonic logic. Journal of Philosophical Logic **23** (1994) 35–65

[17] Caminada, M.: On the issue of contraposition of defeasible rules. Frontiers in Artificial Intelligence and Applications **172** (2008) 109–115

[18] Valentini, S.: The sequent calculus for the modal logic D. Unione Matematica Italiana. Bollettino. A. Serie **7**(7) (1993) 3

[19] Stolpe, A.: A theory of permission based on the notion of derogation. Journal of Applied Logic **8**(1) (2010) 97–113

[20] Forrester, J.: Gentle murder, or the adverbial samaritan. Journal of Philosophy **81** (1984) 193–197

[21] Van Der Torre, L., Y.-H., T.: Cancelling and overshadowing: two types of defeasibility in defeasible deontic logic. In: Proc. IJCAI'95. (1995) 1525–1532

[22] Makinson, D., Van Der Torre, L.: Constraints for Input/Output logics. Journal of Philosophical Logic **30**(2) (2001) 155–185

[23] Simpson, A.K.: The proof theory and semantics of intuitionistic modal logic. PhD thesis, University of Edinburgh (1994)

[24] Makinson, D., van der Torre, L.: Input/Output logics. Journal of Philosophical Logic (29) (2000) 383–408

[25] Gabbay, D.: Bipolar argumentation frames and contrary to duty obligations, preliminary report. In: Proc. CLIMA'12. Springer (2012) 1–24

[26] Oren, N., Luck, M., Miles, S., Norman, T.: An argumentation inspired heuristic for resolving normative conflict. In: Proc. AAMAS-08. (2008) 41–56

[27] Goble, L.: A proposal for dealing with deontic dilemmas. In: Proc. DEON'04. LNCS 3065, Springer (2004) 74–113

[28] Straßer, C.: An adaptive logic framework for conditional obligations and deontic dilemmas. Logic and Logical Philosophy **19**(1–2) (2010) 95–128

[29] Meheus, J., Beirlaen, M., Van De Putte, F.: Avoiding deontic explosion by contextually restricting aggregation. In: Proc. DEON'10. LNAI 6181, Springer (2010) 148–165

[30] Makinson, D., van der Torre, L.: Permission from an input/output perspective. Journal of Philosophical Logic **32**(4) (2003) 391–416