Reasoning With and About Norms in Logical Argumentation

Kees VAN BERKEL [a,1] and Christian STRASSER [b]
[a] *Institute of Logic and Computation, TU Wien, Austria*
[b] *Institute for Philosophy II, Ruhr University Bochum, Germany*
This is a preprint of the article published in the proceedings of the 9th International Conference on Computational Models of Argument (COMMA 2022).

Abstract. Normative reasoning is inherently defeasible. Formal argumentation has proven to be a unifying framework for representing nonmonotonic logics. In this work, we provide an argumentative characterization of a large class of Input/Output logics, a prominent defeasible formalism for normative reasoning. In many normative reasoning contexts, one is not merely interested in knowing whether a specific obligation holds, but also in why it holds despite other norms to the contrary. We propose sequent-style argumentation systems called Deontic Argument Calculi (DAC), which serve transparency and bring meta-reasoning about the inapplicability of norms to the object language level. We prove soundness and completeness between DAC-instantiated argumentation frameworks and constrained Input/Output logics. We illustrate our approach in view of two deontic paradoxes.

Keywords. Nonmonotonic logic, Argumentation, Deontic logic, Normative reasoning

1. Introduction

Obligations and norms fulfil a crucial role in a variety of fields, including law, ethics, AI, and everyday life [1]. The logical study of normative reasoning investigates reasoning with such concepts in formal systems of logic, e.g., deontic logics. Its importance increases with the development of intelligent autonomous systems. Complex normative systems often require reasoning with normative conflicts, exceptions, preferences and priorities [1]. A central challenge is to provide transparent formal models of the underlying reasoning processes, e.g., by means of nonmonotonic logics.

Over the past decades, abstract argumentation has proven to be a unifying framework for the representation of large classes of nonmonotonic logics [2]. Formal argumentation provides both a natural and a transparent model of conflicts and their resolution in terms of conflicting arguments. In this way, it provides a promising basis for tackling the challenging requirements of normative reasoning. The logical analysis of normative reasoning is well-established [1] with the Input/Output framework (I/O) being one of the central approaches [3]. Nonmonotonicity is captured in constrained I/O logics through

[1]Corresponding Authors: Kees van Berkel (e-mail: kees@logic.at) and Christian Straßer (e-mail: christian.strasser@rub.de). Acknowledgments: This work is supported by WWTF and FWF projects MA16-028 and W1255-N23. We thank Leon van der Torre for discussions that significantly shaped this work.

considering maximal consistent families of norms. In recent years, also argumentative representations of deontic logics have attracted increasing interest [4,5,6,7,8,9].

This paper is the first to provide argumentative characterizations for a significant class of I/O logics, including all original logics from [3]. In this way, we are able to combine the advantages of I/O logic with those of formal argumentation. On the one hand, I/O is a highly expressive and robust framework with two decades of developments, including many applications (e.g., priorities, constitutive norms, cognitive modeling, causal reasoning [1,10]). On the other hand, it does not provide the level of transparency that comes with the explicit representation of conflicts in formal argumentation.

In particular, I/O leaves some central challenges of normative reasoning unaddressed. When answering the question as to *why* an obligation holds, one must state *reasons*. Moreover, often it does not suffice to know why a specific obligation holds, one must know why other obligations to the contrary *do not hold*. E.g., in order to understand why "I am permitted to overtake on the left, *despite* having to drive on the right" one must know how the first norm relates to the second. In this case, the first is an exception that renders the latter *inapplicable* in the context of "overtaking another vehicle". Common approaches to I/O logics—as well as deontic logics—do not provide means for making explicit the reasons why certain obligations are not derivable. Despite their central role in ethics and explanation [11], a general lack of explicit modeling of reasons in formal systems has been recently identified [12] (with some exceptions, e.g., [13]). Support and defeat relations are central in the context of reasons as well as in formal argumentation, which makes the latter an ideal framework to reason with and about reasons.

We address these problems by introducing a class of rule-based proof systems called *Deontic Argument Calculi* (DAC) for normative reasoning by means of argumentation.

Our conceptual contribution is twofold: First, we use labels on formulae to make the presentation transparent on the object level, i.e., we can syntactically distinguish between facts, obligations, and constraints without "burdening" the logics with modalities [10]. Second, we internalize some of the meta-reasoning in the I/O formalism by referring to the inapplicability of norms on the object language level. Consequently, our calculi generate both arguments that provide *explicit reasons* for obligations and arguments that defeat other arguments by giving explicit reasons for why certain norms are inapplicable. The second type of arguments concerns the nonmonotonicity of normative reasoning. The possibility to reason about the inapplicability of norms on the object language level distinguishes our work from other systems such as [14,15]. We illustrate the utility of our approach using the notion of *related admissibility* [16] to explain why some obligation holds *despite* certain norms to the contrary.

The technical contribution of this work contains two types of completeness results: First, we show adequacy between DAC and a significant class of monotonic I/O consequence relations. Second, we prove that formal argumentation frameworks instantiated with DAC arguments characterize a large class of nonmonotonic I/O logics. This makes our work the first to argumentatively characterize I/O logic. Moreover, DAC enjoys a modularity particularly suitable for expansions and our calculi are modular with respect to a large class of base logics. Last, our work contributes to previous representation results in formal argumentation concerning systems based on maximal consistent sets [2].

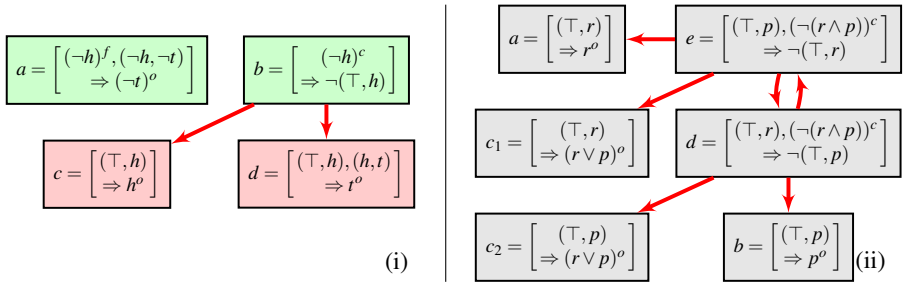

Figure 1. Defeasible normative reasoning examples: (i) The Chisholm scenario (Example 1). Arrows denote defeat relations between arguments, relative to $\mathscr{C}' = \{\neg h^c\}$ (Example 2). (ii) A deontic conflict (Example 3). Argument e defends $\{b, c_2, e\}$, whereas argument d defends $\{a, c_1, d\}$.

2. Basic Terminology and Benchmark Examples

We introduce basic terminology by considering two examples. Developments in deontic logic are driven by challenging examples [17]. Here, we focus on *contrary-to-duty* reasoning and *deontic conflicts*. Both can be effectively addressed using nonmonotonic reasoning [10] (for alternative approaches see [1]). The language $\mathscr{L}$ is defined as follows:

$$\varphi ::= p \mid \top \mid \bot \mid \neg\varphi \mid \varphi \vee \varphi \mid \varphi \wedge \varphi \mid \varphi \rightarrow \varphi$$

with $p \in \mathsf{Atoms}$. All connectives are primitive in order to be modular with respect to the base logic (Section 3). We use $p, q, r, \ldots$ for atoms, and reserve $\varphi, \psi, \theta, \ldots$ for arbitrary formulae of $\mathscr{L}$. In order to increase transparency we *label* formulae of $\mathscr{L}$, i.e., $\mathscr{L}^i = \{\varphi^i \mid \varphi \in \mathscr{L}\}$ for $i \in \{f, o, c\}$. We have formulae expressing *facts* $\mathscr{L}^f$, *obligations* $\mathscr{L}^o$, and *constraints* $\mathscr{L}^c$. Moreover, we employ pairs of formulae $\mathscr{L}^n = \{(\varphi, \psi) \mid \varphi, \psi \in \mathscr{L}\}$. A pair (φ, ψ) represents a *norm*: i.e., "given fact φ, it is obligatory that ψ" [3].

We work with knowledge bases of the type $\langle \mathscr{F}, \mathscr{N}, \mathscr{C} \rangle$, where $\mathscr{F} \subseteq \mathscr{L}^f$ constitutes the factual context, $\mathscr{N} \subseteq \mathscr{L}^n$ denotes a system of norms, and $\mathscr{C} \subseteq \mathscr{L}^c$ represents the constraints with which output must be consistent. The basic idea is that facts (input) trigger norms, from which obligations are detached (output). Moreover, constraints control the output to ensure consistency. The above is in the spirit of constrained I/O logic [3].

Suppose we have a single fact $\mathscr{F} = \{p^f\}$, a norm system $\mathscr{N} = \{(p, q)\}$, and no constraints, then an argument concluding that q is obligatory is of the following form,

$$p^f, (p, q) \Rightarrow q^o$$

The left-hand side (lhs) gives reasons for the conclusion on the right-hand side (rhs).

Example 1 (Chisholm scenario [1], Figure 1-i)**.** *Jones is under the obligation to go and help her neighbors* $(\top, h)$.[2] *Furthermore, Jones knows if she goes to help, she must tell them she goes* (h, t)*. Now, if Jones does not go, she ought not to tell them she goes* $(\neg h, \neg t)$*. It turns out that Jones does not go to help* $\neg h^f$*. Clearly, Jones has violated her primary obligation to go and help. Let the knowledge base be* $\mathscr{F} = \{\neg h^f\}$ *and* $\mathscr{N} = \{(\top, h), (h, t), (\neg h, \neg t)\}$*. Figure 1-i presents arguments* a, c*, and* d *that can be constructed from the knowledge base (we explain the meaning of* b *and the arrows in Exam-*

[2] A norm $(\top, \varphi)$ with a precondition $\top$ is triggered by default, that is, even by an empty factual context.

ple 2); e.g., in argument a, the reasons for not telling $\neg t^o$ are the fact $\neg h^f$ and the norm $(\neg h, \neg t)$. What must Jones do in this contrary-to-duty *scenario? The desired answer is that she ought not to tell the neighbors she goes $\neg t^o$. Formalizations of this scenario cause problems for (monadic) deontic logics, e.g., both t and $\neg t$ become obligatory.*

Arguments do not only provide reasons in support of an obligation, but also defend them from potential *defeaters*. A rebuttal defeat opposes the conclusion of an argument without pinpointing the reason as to why. In contrast, attacks on reasons—i.e., *undercuts*—are arguments that express *which reasons* are *inapplicable* given some context. We adopt undercuts since they are more transparent about attacks. Recall that constraints are consistency requirements and suppose $\mathscr{C} = \{\neg q^c\}$.[3] Then, a defeating argument

$$p^f, \neg q^c \Rightarrow \neg(p,q)$$

expresses that, if output is to be consistent with the constraint $\neg q^c$, in context p^f the norm (p,q) cannot be consistently asserted as a reason (since it would detach q^o). Hence, $\neg(p,q)$ expresses that this norm is *inapplicable* given $\mathscr{F}$ and $\mathscr{C}$. An argumentation framework is then simply a set of arguments with defeat relations holding between them.

Example 2 (Example 1 cont.)**.** *We want to know what Jones must do in the light of her violation $\neg h^f$. Thus, we impose the constraint that the output must be consistent with the fact that Jones does not help $\mathscr{C} = \{\neg h^c\}$ (i.e., $\mathscr{C} = \mathscr{F}$ modulo relabelling). This constraint gives us the argument $b : \neg h^c \Rightarrow \neg(\top, h)$ expressing that given consistency requirement $\neg h$, the norm $(\top, h)$ may not be asserted as a reason (it would output the inconsistent h^o). This argument serves as a defeater of any argument which appeals to $(\top, h)$ in its reasons, in this case both c and d; see the defeat arrows in Figure 1-i. So, that Jones ought not to tell $\neg t^o$ is explained by argument a together with the fact that arguments c and d concluding helping h^o, respectively telling t^o, cannot be defended in view of b. Namely, c and d both employ reasons that are inapplicable given $\mathscr{C}$.*

What makes this approach more transparent is the use of labels in arguments to indicate different types of information (factual, obligations, constraints), the internalized meta-reasoning about inapplicability of norms, and the argumentation framework revealing the contrastive dimension of defeasible reasoning. In Figure 1-i, the question "why shouldn't Jones help, *despite* argument c?" is answered by "since argument b attacks c and b is not attacked." These notions will be made precise in subsequent sections.

Example 3 (Deontic Conflict [1], Figure 1-ii)**.** *Suppose Smith has an obligation to return a borrowed weapon to a colleague $(\top, r)$. Smith knows the colleague is planning to commit a crime with this weapon and Smith is under the obligation to prevent crime $(\top, p)$. Furthermore, the constraint is that Smith cannot secure both r and p. What should Smith do? This is a* deontic conflict*. The knowledge base is $\mathscr{F} = \emptyset$, $\mathscr{N} = \{(\top, r), (\top, p)\}$, and $\mathscr{C} = \{\neg(r \wedge p)^c\}$. Suppose we reason classically, e.g., p entails $p \vee r$. The arguments that can be constructed are presented in Figure 1-ii. The two defeating arguments, d and e, express that given the constraints one of either two norms cannot be asserted.*

Intuitively, the defensible set $\{a, c_1, d\}$ justifies the obligation that Smith ought to return the weapon in Example 3, whereas $\{b, c_2, e\}$ does this for the prevention of crime.

[3] Since we do not allow for formulae with mixed labels, we can safely omit brackets w.r.t. using labels.

$$\frac{}{(\top,\top)}\ \mathrm{T} \qquad \frac{}{(\varphi,\varphi)}\ \mathrm{ID} \qquad \frac{(\varphi,\psi) \qquad \psi\vdash\gamma}{(\varphi,\gamma)}\ \mathrm{WO} \qquad \frac{(\varphi,\psi) \qquad (\varphi,\gamma)}{(\varphi,\psi\wedge\gamma)}\ \mathrm{AND}$$

$$\frac{(\varphi,\psi) \qquad \gamma\vdash\varphi}{(\gamma,\psi)}\ \mathrm{SI} \qquad \frac{(\varphi,\psi) \qquad (\varphi\wedge\psi,\gamma)}{(\varphi,\gamma)}\ \mathrm{CT} \qquad \frac{(\varphi,\psi) \qquad (\gamma,\psi)}{(\varphi\vee\gamma,\psi)}\ \mathrm{OR}$$

Figure 2. Rules for constructing $\mathsf{deriv}_{\mathscr{R},\mathsf{L}}$ proof-systems. The minimal set of deriv-rules is $\{\mathrm{WO},\mathrm{AND},\mathrm{SI}\}$.

Likewise, one can justify the floating conclusion $(r\vee p)^o$ in Figure 1-ii, by arguing that in every defensible stance *either* c_1 *or* c_2 is selected (cf. disjunctive response [13]). However, following a more skeptical reasoning style one can argue why $r\vee p$ is not obligatory since there is no single argument concluding $(r\vee p)^o$ that is selected in *every* defensible stance. Defeasible reasoning by means of argumentation gives rise to various reasoning styles, including the aforementioned. We will discuss these in Section 5.

3. Constrained Input/Output Logic

We briefly recall the basics of *Constrained Input/Output logic*, the systems for which we provide argumentative characterizations. The formalism was developed by Makinson and van der Torre [3] and is particularly suitable for normative reasoning [10]. Its central feature is the employment of syntactic objects of the form (φ,ψ), called *norms*.

I/O logics are construed over the *non*-labelled propositional language $\mathscr{L}$ (Section 2) and a base logic L. We use capital Greek letters $\Delta,\Gamma,\ldots$ for finite sets of $\mathscr{L}$-formulae and write $\bigwedge\Delta$ to denote the conjunction of elements of Δ. Let $\vdash$ denote the consequence relation of the base logic L. We assume that $\vdash$ is reflexive ($\Gamma,\varphi\vdash\varphi$), transitive ($\Gamma\vdash\psi$ and $\Gamma',\psi\vdash\varphi$ implies $\Gamma,\Gamma'\vdash\varphi$) and monotonic ($\Gamma\vdash\varphi$ implies $\Gamma,\Gamma'\vdash\varphi$). We also assume the presence of a conjunction $\wedge$, for which $\Gamma\vdash\psi\wedge\varphi$ iff $\Gamma\vdash\psi$ and $\Gamma\vdash\varphi$, a negation $\neg$, for which $\Gamma\vdash\varphi$ iff $\Gamma,\neg\varphi\vdash$, a disjunction $\vee$, for which $\Gamma,\varphi_1\vdash\psi$ and $\Gamma,\varphi_2\vdash\psi$ iff $\Gamma,\varphi_1\vee\varphi_2\vdash\psi$, and a falsum constant $\bot$ for which $\bot\vdash\varphi$ and $\varphi,\neg\varphi\vdash\bot$. We assume L has an adequate *sequent calculus* LC, i.e., $\Delta\vdash\varphi$ iff the sequent $\Delta\Rightarrow\varphi$ is LC-derivable.

Constrained I/O logics work with knowledge bases of the type $\langle\mathscr{F},\mathscr{N},\mathscr{C}\rangle$, where $\mathscr{F}\subseteq\mathscr{L}$ is the *factual* input, $\mathscr{N}\subseteq\mathscr{L}\times\mathscr{L}$ a *normative system*, and $\mathscr{C}\subseteq\mathscr{L}$ a set of *constraints* containing the formulae with which output must be consistent. We assume $\mathscr{F}$ and $\mathscr{C}$ to be consistent, i.e., $\mathscr{F}\nvdash\bot$ and $\mathscr{C}\nvdash\bot$. The traditional I/O proof systems are only available for a class of *monotonic* I/O logics [10]. The system is referred to as "deriv" and contains inference rules that derive I/O pairs from other I/O pairs (Figure 2).

Definition 1. *Let* $\mathsf{deriv}_{\mathscr{R},\mathsf{L}}$ *be a proof-system, with* $\mathscr{R}$ *a set of rules from Table 2. Let* L *be the base logic, and let* $\mathscr{N}\subseteq\mathscr{L}^n$. *A* derivation *of* $(\varphi,\psi)\in\mathsf{deriv}_{\mathscr{R},\mathsf{L}}(\mathscr{N})$ *is a* tree *of rule-applications of* $\mathscr{R}$ *where the leaves are either members of* $\mathscr{N}$ *or instances of T and ID (if* $T,ID\in\mathscr{R}$*), all members of* $\mathscr{N}$ *are among the leaves, and the root is* (φ,ψ).

We say ψ *is* obligatory *(detached) under* $\mathscr{N}$ *and* $\mathscr{F}$ *if* $(\varphi,\psi)\in\mathsf{deriv}_{\mathscr{R},\mathsf{L}}(\mathscr{N}')$ *with* $\mathscr{F}\vdash\varphi$ *and* $\mathscr{N}'\subseteq\mathscr{N}$. *We write* $\psi\in\mathsf{deriv}_{\mathscr{R},\mathsf{L}}(\Delta,\mathscr{N})$ *if* $(\bigwedge\Delta,\varphi)\in\mathsf{deriv}_{\mathscr{R},\mathsf{L}}(\mathscr{N})$.

Paradigmatic I/O logics are characterized by the sets of rules $\mathscr{R}_1=\{\mathrm{T},\mathrm{WO},\mathrm{SI},\mathrm{AND}\}$, $\mathscr{R}_2=\{\mathrm{OR}\}\cup\mathscr{R}_1$, $\mathscr{R}_3=\mathscr{R}_1\cup\{\mathrm{CT}\}$, and $\mathscr{R}_4=\mathscr{R}_2\cup\mathscr{R}_3$. The system $\mathscr{R}_1$ represents a *single deontic detachment* procedure which allows for weakening of the output (WO), combining output (AND), and strengthening of the input (SI). All propositional tautolo-

gies are among the output (T). System $\mathscr{R}_2$ extends $\mathscr{R}_1$ with *reasoning by cases* (OR), i.e., if both φ and γ generate output ψ, then $\varphi \vee \gamma$ generates ψ too. System $\mathscr{R}_3$ extends $\mathscr{R}_1$ with reusability (CT) allowing for iterations of *successive deontic detachment* (cf. chaining reasons in Example 5). Last, $\mathscr{R}_4$ combines $\mathscr{R}_2$ and $\mathscr{R}_3$. The above systems may be closed under *throughput* (ID), i.e., input is 'put through' as output. We write $\mathscr{R}_i^+ = \mathscr{R}_i \cup \{\text{ID}\}$ for $i \in \{1,2,3,4\}$. The resulting eight systems are sound and complete with respect to their semantic characterizations [10]. We omit the semantics here.

The above systems are still monotonic. As Example 1 and 3 demonstrate, we require defeasible detachment. Constrained I/O logics enable this [3]. Constrained I/O logics work with maximal families of norms $\mathscr{N}' \subseteq \mathscr{N}$ under which the output remains consistent with the constraints $\mathscr{C}$. If the output is required to be consistent *per se*, we let $\mathscr{C} = \emptyset$. If the output is to be consistent with the input, we take $\mathscr{F} \subseteq \mathscr{C}$ (e.g., Example 2).

Definition 2. *Let* $\mathsf{deriv}_{\mathscr{R},\mathsf{L}}$ *be a system from Figure 2 and let* $\mathscr{K} = \langle \mathscr{F}, \mathscr{N}, \mathscr{C} \rangle$ *be a knowledge base. The set of* maximal consistent families *of* $\mathscr{N}$ *(*maxfam*) is defined as:*

- $\mathsf{maxfam}_{\mathscr{R},\mathsf{L}}(\mathscr{K})$ *is the set of* max-elements *of* $\{\mathscr{N}' \subseteq \mathscr{N} \mid$ *for all* $(\varphi,\psi) \in \mathsf{deriv}_{\mathscr{R},\mathsf{L}}(\mathscr{N}')$*, if* $\mathscr{F} \vdash \varphi$*, then* $\mathscr{C}, \psi \nvdash \bot\}$.

We define sceptic nonmonotonic inference $\mid\!\sim^s$ *for* constrained I/O logic *as follows:*

- $\mathscr{K} \mid\!\sim^s_{\mathscr{R},\mathsf{L}} \varphi$ *iff* $\forall\, \mathscr{N}' \in \mathsf{maxfam}_{\mathscr{R},\mathsf{L}}(\mathscr{K})$, $\exists(\psi,\varphi) \in \mathsf{deriv}_{\mathscr{R},\mathsf{L}}(\mathscr{N}')$ *with* $\mathscr{F} \vdash \psi$.

Example 4 (Example 1 cont.)**.** *Consider* $\mathscr{R}_3$ *with* L *a classical logic,* $\mathscr{F} = \{\neg h\}$*, and* $\mathscr{N} = \{(\top,h),(h,t),(\neg h,\neg t)\}$*. For* $\mathscr{C} = \emptyset$*, we have* $\mathsf{maxfam}_{\mathscr{R}_3,\mathsf{L}}(\mathscr{F},\mathscr{N},\mathscr{C}) = \{\{(\top,h),(h,t)\},\{(\neg h,\neg t),(\top,h)\},\{(\neg h,\neg t),(h,t)\}\}$*. We derive* $(\top, h \wedge t) \in \mathsf{deriv}_{\mathscr{R}_3,\mathsf{L}}(\{(\top,h),(h,t)\})$ *as follows:*

$$\dfrac{\dfrac{(\top,h) \qquad \dfrac{(h,t) \qquad \top \wedge h \vdash h}{(\top \wedge h, t)}\,\text{SI}}{(\top,t)}\,\text{CT} \qquad (\top,h)}{(\top, h \wedge t)}\,\text{AND}$$

with $\mathscr{F} \vdash \top$ *and* $\mathscr{C}, h \wedge t \nvdash \bot$*. However, once we set the constraints to Jones' violation, i.e.,* $\mathscr{C}' = \mathscr{F}$*, we obtain a single* maxfam *member* $\mathscr{N}' = \{(\neg h,\neg t),(h,t)\}$ *since now* $\mathscr{C}', h \vdash \bot$ *whereas* $\mathscr{C}', \neg t \nvdash \bot$ *(note that* $(h,t) \in \mathscr{N}'$ *cannot be triggered by* $\mathscr{F}$*). Given* $\mathscr{C}'$*, Jones is obliged to not tell, i.e.,* $\mathscr{K} \mid\!\sim^s_{\mathscr{R},\mathsf{L}} \neg t$*, and is not obliged to help, i.e.,* $\mathscr{K} \mid\!\not\sim^s_{\mathscr{R},\mathsf{L}} h$.

First, maxfam sets (of arbitrary size) do not provide formal ways of pinpointing the reasons why some norms are inapplicable, e.g., why $(\top,h)$ in Example 4 is inapplicable given $\mathscr{C} = \mathscr{F}$. Second, deriv is unsuitable for generating transparent arguments, e.g., as a certificate the derivation in Example 4 may justify *that* $(\top, h \wedge t)$ is derivable, its conclusion does not explain *why* $h \wedge t$ is obligatory. In fact, although in general a derivation is a justification, it is not necessarily an explanation. Our calculi address both challenges.

4. Deontic Argument Calculi (DAC)

In order to generate more transparent I/O arguments, we label propositional formulae as facts $\mathscr{L}^f$, obligations $\mathscr{L}^o$, and constraints $\mathscr{L}^c$ (Section 2). What is more, we allow for Boolean operations over the more complex meta-logical objects (φ,ψ) denoting norms. Operations over these higher-order syntactic objects enable undercuts that explain why

$$\textbf{Ax}\ \vdash_{\mathsf{LC}} \Delta^i \Rightarrow \Gamma^i, \text{ for } i \in \{f,o,c\} \quad \textbf{Taut}\ \Rightarrow (\top,\top) \quad \textbf{Detach}\ \varphi^f,(\varphi,\psi) \Rightarrow \psi^o \quad \textbf{TP}\ \varphi^f \Rightarrow \varphi^o$$

$$\textbf{R-C}\ \frac{\Delta \Rightarrow \varphi^o}{\Delta,(\neg\varphi)^c \Rightarrow} \qquad \textbf{R-N}\ \frac{\Delta,(\varphi,\psi) \Rightarrow}{\Delta \Rightarrow \neg(\varphi,\psi)} \qquad \textbf{L-CT}^a\ \frac{\varphi^f,\Delta \Rightarrow \Theta}{\varphi^o,\Delta \Rightarrow \Theta}$$

$$\textbf{L-OR}^b\ \frac{\Delta,\varphi^f \Rightarrow \Theta \qquad \Delta',\psi^f \Rightarrow \Theta}{\Delta,\Delta',(\varphi \vee \psi)^f \Rightarrow \Theta} \qquad \textbf{Cut}^c\ \frac{\Delta \Rightarrow \varphi \qquad \varphi,\Delta' \Rightarrow \Theta}{\Delta,\Delta' \Rightarrow \Theta}$$

Figure 3. Rules for building $\mathsf{DAC}_{\mathscr{S}}$ (Definition 3). The upper level contains initial sequents and the lower level logical and structural rules. Side-condition (a) denotes $\Delta \cap \mathscr{L}^n \neq \emptyset$; (b) denotes that if $\textbf{TP} \notin \mathscr{S}$, then $\Delta \cap \mathscr{L}^n \neq \emptyset$ and $\Delta' \cap \mathscr{L}^n \neq \emptyset$; and (c) that $\varphi \in \mathscr{L}^{io}$.

certain norms should (not) be applied. For the present work, it suffices to consider negation only. Let $\overline{\mathscr{L}^n} = \{\neg(\varphi,\psi) \mid (\varphi,\psi) \in \mathscr{L}^n\}$. The *language of norms* is defined as $\mathscr{L}^n \cup \overline{\mathscr{L}^n}$. Furthermore, let $\mathscr{L}^{io} = \mathscr{L}^f \cup \mathscr{L}^o \cup \mathscr{L}^c \cup \mathscr{L}^n \cup \overline{\mathscr{L}^n}$ be the full labelled I/O language. In $\mathscr{L}^{io}$, norms are integrated into the object-level language. We write φ for an arbitrary formula of $\mathscr{L}^{io}$ and write Δ^i to denote that $\Delta^i \subseteq \mathscr{L}^i$ for $i \in \{f,o,c,n\}$.

We introduce *Deontic Argument Calculi* (DAC) for I/O logic. These calculi are *sequent-style* calculi, which are rule-based proof systems employing syntactic objects of the form $\Delta \Rightarrow \Gamma$, with $\Delta,\Gamma \subseteq \mathscr{L}^{io}$ and '$\Rightarrow$' as a sequent arrow. We call $\Delta \Rightarrow \Gamma$ a sequent or an argument, where Δ denotes the reasons for Γ (Section 2). Furthermore, Δ is interpreted as a regular finite set and Γ is restricted to at most one formula. The use of regular sets instead of multi-sets is more modular w.r.t. the base logic L. Let LC be an adequate sequent calculus for the base logic L, then, intuitively, DAC takes *labelled* versions of any LC-derivable $\Delta \Rightarrow \Gamma$ as an initial sequent (i.e., $\Delta^i \Rightarrow \Gamma^i$ for each $i \in \{f,o,c\}$) and contains logical- and structural rules for transforming labelled formulae of $\mathscr{L}^{io}$ (see Figure 3).

Definition 3. *Let* DAC *be the base system with the underlying logic* L*, containing the rules* **Ax**, **Detach**, **R-C**, **R-N***, and* **Cut** *from Figure 3. The calculus* $\mathsf{DAC}_{\mathscr{S}}$ *extends* DAC *with the set of rules* $\mathscr{S} \subseteq \{$**Taut**,**TP**,**L-OR**,**L-CT**$\}$*, leading to 16* DAC*-axiomatizations.*

A $\mathsf{DAC}_{\mathscr{S}}$*-derivation of* $\Delta \Rightarrow \Gamma$ *is a tree whose leafs are initial sequents of* $\mathsf{DAC}_{\mathscr{S}}$*, whose root is* $\Delta \Rightarrow \Gamma$*, and whose rule-applications are instances of the rules of* $\mathsf{DAC}_{\mathscr{S}}$*. We write* $\vdash_{\mathscr{S}} \Delta \Rightarrow \Gamma$ *(resp.* $\vdash^n_{\mathscr{S}} \Delta \Rightarrow \Gamma$*) if* $\Delta \Rightarrow \Gamma$ *is* $\mathsf{DAC}_{\mathscr{S}}$*-derivable (in at most n steps).*

Since $\mathsf{DAC}_{\mathscr{S}}$ takes labelled LC-derivable sequents as initial sequents, the rules of LC are not part of $\mathsf{DAC}_{\mathscr{S}}$. Still, LC rules can be straightforwardly shown admissible in DAC due to the presence of **Cut**. The rule **Taut** ensures that all propositional tautologies are considered as output. The rule **Detach** is an initial explanatory argument stating that the fact φ and the norm (φ,ψ) are reasons for the obligation ψ. Instead of deriving pairs from other pairs (as in deriv), we keep norms as primitive reasons from a given normative code $\mathscr{N}$ and only modify facts, obligations, and constraints. This gives us some explanatory advantages (see **R-C** and **R-N** below). The rule **TP** corresponds to throughput. The rule **L-CT** corresponds to successive detachment, expressing that a norm may likewise be triggered by the output of some other norm (cf. Example 5). **L-OR** reflects reasoning by cases over input. The side-condition on **L-OR** is dropped for $\textbf{TP} \in \mathscr{S}$ due to reasoning by cases with throughput. **Cut** suffices as the only structural rule.

More interesting are the rules **R-C** and **R-N**. Concerning **R-C**, think of a sequent with an empty right-hand side as an argument expressing inconsistent reasons. For instance, an argument $\varphi^f,(\varphi,\psi),(\neg\psi)^c \Rightarrow$ explains that the fact φ and the norm (φ,ψ)

(which are reasons for ψ) are inconsistent whenever the output must be consistent with $\neg\psi$. What is more, whenever such an argument expresses inconsistent reasons, we know at least one of its norms is inapplicable. The rule **R-N** expresses this: from $\varphi^f, (\varphi,\psi), (\neg\psi)^c \Rightarrow$ we obtain the defeating argument $\varphi^f, (\neg\psi)^c \Rightarrow \neg(\varphi,\psi)$. Hence, φ^f and $(\neg\psi)^c$ are reasons for the *inapplicability* of the norm (φ,ψ). $\mathsf{DAC}_{\mathscr{S}}$ sequents will be the building blocks for the desired argumentative characterizations (Section 5).

Example 5 (Example 1 cont.)**.** *The* DAC*-argument d (Figure 1-i), concluding that Jones should tell her neighbors she is coming to help, is derived through chaining* $(\top,h)$ *and* (h,t)*. The following* $\mathsf{DAC}_{\mathscr{S}}$*-derivation (left) shows this, where* $\mathbf{L\text{-}CT} \in \mathscr{S}$*:*

$$\dfrac{\dfrac{}{\top^f,(\top,h)\Rightarrow h^o}\,\textbf{Detach} \qquad \dfrac{\dfrac{}{h^f,(h,t)\Rightarrow t^o}\,\textbf{Detach}}{h^o,(h,t)\Rightarrow t^o}\,\textbf{L-CT}}{\top^f,(\top,h),(h,t)\Rightarrow t^o}\,\textbf{Cut} \qquad\qquad \dfrac{\dfrac{}{\top^f,(\top,h)\Rightarrow h^o}\,\textbf{Detach}}{\dfrac{\top^f,(\neg h)^c,(\top,h)\Rightarrow}{\top^f,(\neg h)^c\Rightarrow\neg(\top,h)}\,\textbf{R-N}}\,\textbf{R-C}$$

Given $\mathscr{C}' = \{\neg h^c\}$*, "why should Jones not tell,* despite *argument d?" is answered by the (right) derivable argument b (Figure 1-i). The fact* $\top^f$ *is omitted by a* **Cut** *with* $\Rightarrow \top^f$*.*

Example 6 (Example 3 cont.)**.** *In the dilemma, Smith cannot both return the weapon and prevent the crime. So, we find* $(\top,r)$ *applicable if and only if* $(\top,p)$ *is inapplicable. This is expressed by arguments e and f. The* $\mathsf{DAC}_{\mathscr{S}}$*-derivations of e and f from Figure 1-ii are obtained similarly to argument b in Example 5, using* **Detach** *twice, the* DAC*-admissible rule from* LC *for right conjunction introduction,* **R-C**, and **R-N** *consecutively.*

5. Argumentation and DAC-Instantiations

DAC arguments are of two types: they either give reasons for obligations, or they give reasons for why certain norms are inapplicable, i.e., defeated. The latter arguments capture the defeasibility of normative reasoning and define the interaction among arguments. We define DAC-induced *argumentation frameworks* (*AF*s) to model this interaction.

Definition 4. *Let* $\mathsf{DAC}_{\mathscr{S}}$ *be a calculus and* $\mathscr{K} = \langle \mathscr{F}, \mathscr{N}, \mathscr{C} \rangle$ *a labelled knowledge base (i.e.,* $\mathscr{F} \subseteq \mathscr{L}^f, \mathscr{N} \subseteq \mathscr{L}^n$*, and* $\mathscr{C} \subseteq \mathscr{L}^c$*). We define* $\mathrm{AF}_{\mathscr{S}}(\mathscr{K}) = \langle \mathsf{Arg}, \mathsf{Att} \rangle$ *as follows:*

- $\Delta \Rightarrow \Gamma \in \mathsf{Arg}$ *iff* $\Delta \Rightarrow \Gamma$ *is* $\mathsf{DAC}_{\mathscr{S}}$*-derivable,* $\Delta \subseteq \mathscr{F} \cup \mathscr{N} \cup \mathscr{C}$*, and* $\Gamma \subseteq \mathscr{L}^{io}$*;*
- a defeats b*, i.e.,* $(a,b) \in \mathsf{Att}$ *iff* $a = \Delta \Rightarrow \neg(\varphi,\psi)$ *and* $b = \Gamma, (\varphi,\psi) \Rightarrow \Theta$*.*

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

For an $AF_{\mathscr{S}}(\mathscr{K})$ it suffices to only consider arguments relevant to $\mathscr{K}$, i.e., $\mathsf{Arg}(\mathscr{F} \cup \mathscr{N} \cup \mathscr{C})$. We are interested in what combinations of arguments (*extensions*) can be collectively accepted given an *AF*. For our purpose, stable extensions suffice.

Definition 5. *Let* $\langle \mathsf{Arg}, \mathsf{Att} \rangle$ *be an* AF *and let* $\mathscr{E} \subseteq \mathsf{Arg}$*:*

- $\mathscr{E}$ defeats *an argument* $a \in \mathsf{Arg}$ *if there is a* $b \in \mathscr{E}$ *that defeats a, i.e.,* $(b,a) \in \mathsf{Att}$*;*
- $\mathscr{E}$ *is* conflict-free *if it does not defeat any of its own elements;*
- $\mathscr{E}$ *is* stable *if it is conflict-free and defeats all* $b \in \mathsf{Arg} \setminus \mathscr{E}$*.*

Table 1. Lemmas for $\vdash_{\mathscr{S}}$. Let $\Delta^{\downarrow}$ and $\varphi^{\downarrow}$ be the set of formulae in Δ, resp. φ stripped from any labels.

Lemma :	if	then
1	$\vdash_{\mathscr{S}} \Delta, \Gamma_1^c \Rightarrow \Sigma$ and $\mathscr{C} \vdash \bigwedge \Gamma_1$	$\exists \Gamma_2 \subseteq \mathscr{C} : \vdash_{\mathscr{S}} \Delta, \Gamma_2^c \Rightarrow \Sigma$ and $\Gamma_2 \vdash \bigwedge \Gamma_1$
2	$\vdash^n_{\mathscr{S}} \Delta \Rightarrow \neg(\varphi, \psi)$	$\vdash^n_{\mathscr{S}} \Delta, (\varphi, \psi) \Rightarrow$
3		$\Delta^{\downarrow} \vdash \gamma^{\downarrow}$, where $\Delta \subseteq \mathscr{L}^f \cup \mathscr{L}^o \cup \{(\top, \top)\}$, $\gamma \in \mathscr{L}^f \cup \mathscr{L}^o$
4		$\vdash \gamma^{\downarrow}$, where $\mathbf{TP} \notin \mathscr{S}$, $\Delta \subseteq \mathscr{L}^f \cup \{(\top, \top)\}$, $\gamma \in \mathscr{L}^o$
5	$\vdash^n_{\mathscr{S}} \Delta \Rightarrow$	$\vdash^n_{\mathscr{S}} \Delta \setminus \mathscr{L}^c \Rightarrow \varphi^o$ s.t. $\varphi \vdash \neg \bigwedge (\Delta \cap \mathscr{L}^c)^{\downarrow}$, where $\neg \bigwedge \emptyset =_{\mathsf{df}} \bot$

Let `Stable` *be the set of stable extensions of* AF. *We define sceptic (s), sceptic* (s*), and credulous (c) nonmonotonic inference as follows:*

- $\mathrm{AF} \mid\!\sim^{s}_{\text{stable}} \varphi$ *iff for each* $\mathscr{E} \in$ `Stable`*, there is an* $a \in \mathscr{E}$ *concluding* φ*;*
- $\mathrm{AF} \mid\!\sim^{s^*}_{\text{stable}} \varphi$ *iff there is an* $a \in \bigcap$`Stable` *concluding* φ*;*
- $\mathrm{AF} \mid\!\sim^{c}_{\text{stable}} \varphi$ *iff there is a* $\mathscr{E} \in$ `Stable` *s.t. there is an* $a \in \mathscr{E}$ *concluding* φ*.*

The use of DAC-arguments introduces nuances in sceptic inference: e.g., the distinction between s and s^* corresponds to the discussion of floating conclusions in Section 2.

Example 7 (Example 3 cont.)**.** *Smith is in a dilemma of conflicting duties. The* AF *of Figure 1-ii represents this conflict, where* $\mathsf{Arg} = \{a, b, c_1, c_2, d, e\}$ *and* $\mathsf{Att} = \{(e,a), (e,c_1), (e,d), (d,e), (d,c_2), (d,b)\}$*. It has two* stable *extensions* $\{a, c_1, d\}$ *and* $\{b, c_2, e\}$*, defending the views that Smith ought to return the weapon, resp. prevent the crime. Hence,* $\mathrm{AF} \mid\!\sim^{c}_{\text{stable}} r^o, p^o$*, whereas* $\mathrm{AF} \not\mid\!\sim^{c}_{\text{stable}} (r \wedge p)^o$*,* $\mathrm{AF} \mid\!\sim^{s}_{\text{stable}} r^o$*, and* $\mathrm{AF} \mid\!\sim^{s}_{\text{stable}} p^o$*. For the floating conclusion* $(r \vee p)^o$ *we have* $\mathrm{AF} \mid\!\sim^{s}_{\text{stable}} (r \vee p)^o$ *but* $\mathrm{AF} \not\mid\!\sim^{s^*}_{\text{stable}} (r \vee p)^o$*. (The* AF *of Example 2 in Figure 1-i has one stable extension* $\{a, b\}$*, and so* $\mathrm{AF} \mid\!\sim^{s,s^*,c}_{\text{stable}} (\neg t)^o$*.)*

To illustrate the utility of our approach, we consider the notion of related admissibility [16]. An extension $\mathscr{E}$ is *admissible* if it is conflict-free and $\mathscr{E}$ defeats all arguments defeating some $a \in \mathscr{E}$. An argument a *defends* b iff $a = b$, or there is a c s.t. a defeats c and c defeats b, or there is a c s.t. a defends c and c defends b. A set $\mathscr{E}_a \subseteq \mathsf{Arg}$ is *related admissible with topic* a iff $a \in \mathscr{E}_a$, for all $b \in \mathscr{E}_a$, b defends a, and $\mathscr{E}_a$ is admissible. Thus, a related admissible set $\mathscr{E}_a$ identifies the relevant arguments that justify the acceptability of a. Let $\mathscr{E}^+ = \{a \in \mathsf{Arg} \mid \mathscr{E} \text{ defeats } a\}$ and $\mathscr{E}^- = \{a \in \mathsf{Arg} \mid a \text{ defeats some } b \in \mathscr{E}\}$. In Example 3, the answer to "why is Smith obliged to prevent crime (b)?" is given by the related admissible set $\mathscr{E}_b = \{b, e\}$ where $\mathscr{E}_b^- = \{d\}$ and $\{d\}^- \cap \mathscr{E}_b = \{e\}$ explain that the only counterargument to b is d which is defeated by e expressing that the norm $(\top, r)$ used in d is inapplicable given the reasons $(\top, p)$ and $\neg(r \wedge p)^c$ offered in e. Hence, using only undercuts enables a more refined analysis of the *relevant* norms explaining the (non-)acceptability of certain arguments and obligations. The DAC approach is therefore more precise compared to using maximal consistent families of norms in traditional I/O.

6. Metatheory: Soundness and Completeness

We demonstrate two soundness and completeness results: First, we prove adequacy between I/O proof systems and DAC (Theorem 1). Second, we prove adequacy between constrained I/O logics and DAC-based argumentation frameworks (Theorem 2). We pro-

Table 2. Correspondence between $\mathsf{deriv}_{\mathscr{R},\mathsf{L}}$ rules and $\mathsf{DAC}_{\mathscr{S}}$ rules with the underlying logic L. For instance, $\{\mathrm{ID},\mathrm{OR}\} \subseteq \mathscr{R}$ iff $\{\mathbf{TP},\mathbf{L\text{-}OR}\} \subseteq \mathscr{S}$. The first column represents the minimal sets the systems must contain.

Rules of $\mathsf{deriv}_{\mathscr{R},\mathsf{L}}$	{WO, AND, SI}	T	ID	CT	OR
Rules of $\mathsf{DAC}_{\mathscr{S}}$	$\{\mathbf{Ax},\mathbf{Detach},\mathbf{R\text{-}C},\mathbf{R\text{-}N},\mathbf{Cut}\}$	**Taut**	**TP**	**L-CT**	**L-OR**

vide explicit proofs of the main results. Table 1 lists several technical lemmas whose proofs are omitted: Lemma 1 follows by the compactness of L, while Lemmas 2 to 5 are proven by a straightforward induction on the length of the derivation.

We first show adequacy between deriv and DAC. Both systems are modular and correspondence between the rules of these systems is defined in Table 2. In referring to $\mathsf{deriv}_{\mathscr{R}\mathsf{L}}$ and $\mathsf{DAC}_{\mathscr{S}}$ we assume this correspondence. We state the two directions of Theorem 1 separately, we prove Lemma 7, and omit the similar proof of Lemma 6.

Lemma 6. *Let* $\Theta \subseteq \mathscr{L}^n$*, If* $\vdash_{\mathscr{S}} \Delta^f, \Theta \Rightarrow \varphi^o$*, then* $\varphi \in \mathsf{deriv}_{\mathscr{R},\mathsf{L}}(\Delta,\Theta)$.

Lemma 7. *If* $(\varphi,\psi) \in \mathsf{deriv}_{\mathscr{R},\mathsf{L}}(\Theta)$*, then* $\vdash_{\mathscr{S}} \varphi^f, \Theta \Rightarrow \psi^o$.

Proof. By induction on the length of the $\mathsf{deriv}_{\mathscr{R},\mathsf{L}}$-derivation of (φ,ψ). *Base case.* Case $\{(\varphi,\psi)\} = \Theta$. By **Detach**, $\vdash_{\mathscr{S}} \varphi^f, (\varphi,\psi) \Rightarrow \psi^o$. Case $(\top,\top)$ is derived by T with $\Theta = \emptyset$. By **Detach**, $\vdash_{\mathscr{S}} \top^f, (\top,\top) \Rightarrow \top^o$ and by **Taut**, $\vdash_{\mathscr{S}} \Rightarrow (\top,\top)$. By **Cut**, $\vdash_{\mathscr{S}} \top^f \Rightarrow \top^o$. Case (φ,φ) is derived by ID with $\Theta = \emptyset$. By **TP**, $\varphi^f \Rightarrow \varphi^o$.

Inductive step. To illustrate, we consider the case of CT. The other cases are similar or straightforward. Suppose that (φ,ψ) is derived from $(\varphi,\sigma) \in \mathsf{deriv}_{\mathscr{R},\mathsf{L}}(\Theta_1)$ and $(\varphi \wedge \sigma,\psi) \in \mathsf{deriv}_{\mathscr{R},\mathsf{L}}(\Theta_2)$ by CT, where $\Theta = \Theta_1 \cup \Theta_2$. By the IH, $\vdash_{\mathscr{S}} \varphi^f, \Theta_1 \Rightarrow \sigma^o$ and $\vdash_{\mathscr{S}} (\varphi \wedge \sigma)^f, \Theta_2 \Rightarrow \psi^o$. By R$\wedge$2, $\varphi,\sigma \vdash \varphi \wedge \sigma$. By **Ax**, $\vdash_{\mathscr{S}} \varphi^f, \sigma^f \Rightarrow (\varphi \wedge \sigma)^f$ and by **Cut**, $\varphi^f, \sigma^f, \Theta_2 \Rightarrow \psi^o$. Then, if $\emptyset \neq \Theta_2$, by **L-CT**, $\vdash_{\mathscr{S}} \varphi^f, \sigma^o, \Theta_2 \Rightarrow \psi^o$ and by **Cut**, $\vdash_{\mathscr{S}} \varphi^f, \Theta \Rightarrow \psi^o$. Else, $\Theta_2 = \emptyset$ (and hence $\Theta = \Theta_1$). We consider: (i) $\mathbf{TP} \in \mathscr{S}$ and (ii) $\mathbf{TP} \notin \mathscr{S}$. Ad (i). By Lemma 3.1, $\varphi,\sigma \vdash \psi$ and by **Ax**, $\vdash_{\mathscr{S}} \varphi^o, \sigma^o \Rightarrow \psi^o$. By **TP**, $\vdash_{\mathscr{S}} \varphi^f \Rightarrow \varphi^o$ and by twice **Cut**, $\vdash_{\mathscr{S}} \varphi^f, \Theta \Rightarrow \psi^o$. Ad (ii). By Lemma 3.2, $\vdash \psi$ and so $\sigma \vdash \psi$. By **Ax**, $\vdash_{\mathscr{S}} \sigma^o \Rightarrow \psi^o$. By **Cut**, $\vdash_{\mathscr{S}} \varphi^f, \Theta \Rightarrow \psi^o$. □

Theorem 1. *Let* $\Delta \subseteq \mathscr{L}$*,* $\psi \in \mathscr{L}$*, and* $\Theta \subseteq \mathscr{L}^n$*. Then,* $\vdash_{\mathscr{S}} \Delta^f, \Theta \Rightarrow \psi^o$ *iff* $\psi \in \mathsf{deriv}_{\mathscr{R},\mathsf{L}}(\Delta,\Theta)$.

Proof. ($\Rightarrow$) This is Lemma 6. ($\Leftarrow$) Suppose $\psi \in \mathsf{deriv}_{\mathscr{R},\mathsf{L}}(\Delta,\Theta)$. So, $(\bigwedge\Delta,\psi) \in \mathsf{deriv}_{\mathscr{R},\mathsf{L}}(\Theta)$. By Lemma 7, $\vdash_{\mathscr{S}} (\bigwedge\Delta)^f, \Theta \Rightarrow \psi^o$. Since $\Delta \vdash \bigwedge\Delta$, $\vdash_{\mathscr{S}} \Delta^f \Rightarrow (\bigwedge\Delta)^f$ by **Ax**. By **Cut**, $\vdash_{\mathscr{S}} \Delta^f, \Theta \Rightarrow \psi^o$. □

We now prove our second adequacy result concerning constrained I/O logics and DAC-instantiated argumentation frameworks.

Theorem 2. *Let* $\mathscr{K} = \langle \mathscr{F}, \mathscr{N}, \mathscr{C} \rangle$ *be a knowledge base. Let* $\mathscr{R}$ *be a set of* deriv*-rules and* $\mathscr{S}$ *the set of corresponding* DAC*-rules (Table 2). Let* $\mathrm{AF} = \mathrm{AF}_{\mathscr{S}}(\mathscr{K}) = \langle \mathsf{Arg},\mathsf{Att} \rangle$.

1. *If* $\mathscr{N}' \in \mathsf{maxfam}_{\mathscr{R},\mathsf{L}}(\mathscr{K})$ *then* $\mathsf{Arg}(\mathscr{F}^f \cup \mathscr{N}' \cup \mathscr{C}^c)$ *is stable in* AF.
2. *If* $\mathscr{A}$ *is stable in* AF *then there is a* $\mathscr{N}' \subseteq \mathscr{N}$ *such that* $\mathscr{N}' \in \mathsf{maxfam}_{\mathscr{R},\mathsf{L}}(\mathscr{K})$ *for which* $\mathscr{A} = \mathsf{Arg}(\mathscr{F}^f \cup \mathscr{N}' \cup \mathscr{C}^c)$.

Proof. (1) Let $\mathscr{N}' \in \mathsf{maxfam}_{\mathscr{R},\mathsf{L}}(\mathscr{K})$ and $\mathscr{A} = \mathsf{Arg}(\mathscr{F}^f \cup \mathscr{N}' \cup \mathscr{C}^c)$. For conflict-freeness assume towards a contradiction that there are $a = \Delta^f, \Theta, \Gamma^c \Rightarrow \neg(\varphi,\psi) \in \mathscr{A}$

(where $\Theta \subseteq \mathscr{N}'$) and $b = \Omega, (\varphi,\psi) \Rightarrow \Sigma \in \mathscr{A}$ such that a attacks b. By Lemma 2 and since $(\varphi,\psi) \in \mathscr{N}'$, we have, $\Delta^f, \Theta, \Gamma^c, (\varphi,\psi) \Rightarrow \in \mathscr{A}$. By Lemma 5, $\Delta^f, \Theta, (\varphi,\psi) \Rightarrow \sigma^o \in \mathscr{A}$ for some σ for which $\sigma \vdash \neg\bigwedge\Gamma$. By Theorem 1, $\sigma \in \mathsf{deriv}_{\mathscr{R},\mathsf{L}}(\Delta, \Theta \cup \{(\varphi,\psi)\})$, which contradicts the $\mathscr{C}$-consistency of $\mathscr{N}'$.

For $\mathscr{A}$ defeats all $b \in \mathsf{Arg} \setminus \mathscr{A}$ let $a = \Delta_1^f, \Theta_1, \Gamma_1^c \Rightarrow \Sigma \in \mathsf{Arg} \setminus \mathscr{A}$, where $\Theta_1 \subseteq \mathscr{L}^n$. So, there is a $(\varphi,\psi) \in \Theta_1 \setminus \mathscr{N}'$. By the maximal consistency of $\mathscr{N}'$, $\mathscr{N}' \cup \{(\varphi,\psi)\}$ is inconsistent with $\mathscr{C}$. So, there is a $\theta \in \mathsf{deriv}_{\mathscr{R},\mathsf{L}}(\Delta_2, \Theta_2)$ for some $\Delta_2 \subseteq \mathscr{F}$ and $\Theta_2 \subseteq \mathscr{N}' \cup \{(\varphi,\psi)\}$ such that $\mathscr{C} \vdash \neg\theta$. By Theorem 1, $\Delta_2, \Theta_2 \Rightarrow \theta^o \in \mathsf{Arg}$. Note that $(\varphi,\psi) \in \Theta_2$ since otherwise $\Theta_2 \subseteq \mathscr{N}'$ in contradiction to the consistency of $\mathscr{N}'$. By **R-C** and **R-N**, $\vdash_{\mathscr{S}} \Delta_2, \Theta_2 \setminus \{(\varphi,\psi)\}, (\neg\theta)^c \Rightarrow \neg(\varphi,\psi)$. By Lemma 1, $b = \Delta_2, \Theta_2 \setminus \{(\varphi,\psi)\}, \Gamma_2^c \Rightarrow \neg(\varphi,\psi) \in \mathsf{Arg}$ for some $\Gamma_2 \subseteq \mathscr{C}$ for which $\Gamma_2 \vdash \neg\theta$. Note, $b \in \mathscr{A}$ and b attacks a.

(2) Let $\mathscr{A}$ be a stable extension of $AF(\mathscr{K})$. Let $\mathscr{N}' = \{(\varphi,\psi) \in \mathscr{N} \mid \neg\exists a = \Delta \Rightarrow \neg(\varphi,\psi) \in \mathscr{A}\}$. We first show that $\mathscr{A} = \mathsf{Arg}(\mathscr{F}^f \cup \mathscr{N}' \cup \mathscr{C}^c)$:

"$(\supseteq)$" Let $a \in \mathsf{Arg}(\mathscr{F}^f \cup \mathscr{N}' \cup \mathscr{C}^c)$. By the definition of $\mathscr{N}'$ there is no $b \in \mathscr{A}$ that attacks a and since $a \in \mathsf{Arg}$ and by the stability of $\mathscr{A}$, $a \in \mathscr{A}$. "$(\subseteq)$" Let $a \in \mathsf{Arg} \setminus \mathsf{Arg}(\mathscr{F}^f \cup \mathscr{N}' \cup \mathscr{C}^c)$ with $a = \Delta \Rightarrow \Gamma$. So, there is a $(\varphi,\psi) \in \Delta$ for which there is a $b \in \mathscr{A}$ with $b = \Theta \Rightarrow \neg(\varphi,\psi)$. So b attacks a and by the stability of $\mathscr{A}$, $a \notin \mathscr{A}$.

We now show that $\mathscr{N}' \in \mathsf{maxfam}_{\mathscr{R},\mathsf{L}}(\mathscr{K})$. Assume for a contradiction that $\mathscr{N}'$ is inconsistent with $\mathscr{C}$. So, there is a $\theta \in \mathsf{deriv}_{\mathscr{R},\mathsf{L}}(\Delta, \Theta)$ for some $\Delta \subseteq \mathscr{F}$ and $\Theta \subseteq \mathscr{N}'$ for which $\mathscr{C} \vdash \neg\theta$. By Theorem 1, $a = \Delta^f, \Theta \Rightarrow \theta^o \in \mathscr{A}$. Assume first that $\Theta = \emptyset$.

If $\mathbf{TP} \notin \mathscr{S}$, by Lemma 3.3, $\vdash \theta$ and thus $\mathscr{C}$ is inconsistent which is a contradiction. Thus, $\Theta \neq \emptyset$. If $\mathbf{TP} \in \mathscr{S}$ then, by Lemma 3.2, $\Delta \vdash \theta$. But then $\mathscr{F} \cup \mathscr{C}$ is inconsistent, a contradiction and so $\Theta \neq \emptyset$. So, in both cases $\Theta \neq \emptyset$.

Let $(\varphi,\psi) \in \Theta$. By **R-N** and **R-C**, $\vdash_{\mathscr{S}} \Delta, \Theta \setminus \{(\varphi,\psi)\}, (\neg\theta)^c \Rightarrow \neg(\varphi,\psi)$. By Lemma 1, there is a $\Gamma \subseteq \mathscr{C}$ for which $b = \Delta, \Theta \setminus \{(\varphi,\psi)\}, \Gamma^c \Rightarrow \neg(\varphi,\psi) \in \mathscr{A}$. Since b attacks a, this contradicts conflict-freeness of $\mathscr{A}$, which shows $\mathscr{N}'$ is consistent with $\mathscr{C}$.

Assume for a contradiction that there is a $(\varphi,\psi) \in \mathscr{N} \setminus \mathscr{N}'$ such that $\mathscr{N}' \cup \{(\varphi,\psi)\}$ is consistent with $\mathscr{C}$ (i.e., $\mathscr{N}'$ is not maximal). By the definition of $\mathscr{N}'$, there is a $b = \Delta^f, \Theta, \Gamma^c \Rightarrow \neg(\varphi,\psi) \in \mathscr{A}$. By Lemma 2, $\vdash_{\mathscr{S}} \Delta^f, \Theta, (\varphi,\psi), \Gamma^c \Rightarrow$. By Lemma 5, $\vdash_{\mathscr{S}} \Delta^f, \Theta, (\varphi,\psi) \Rightarrow \sigma^o$ such that $\sigma \vdash \neg\bigwedge\Gamma$. By Theorem 1, $\sigma \in \mathsf{deriv}_{\mathscr{R},\mathsf{L}}(\Delta, \Theta \cup \{(\varphi,\psi)\})$ which shows that $\mathscr{N}' \cup \{(\varphi,\psi)\}$ is inconsistent with $\mathscr{C}$ (note that $\Gamma \subseteq \mathscr{C}$). This completes our proof since it shows that $\mathscr{N}' \in \mathsf{maxfam}_{\mathscr{R},\mathsf{L}}(\mathscr{K})$. □

Corollary 1. *Let $\mathscr{K}$ be a knowledge base, $\mathscr{R}$ a set of* deriv*-rules, and $\mathscr{S}$ a set of corresponding* DXC*-rules (Table 2). For $i \in \{s,c\}$, $AF_{\mathscr{S}}(\mathscr{K}) \mid\sim^i_{stable} \varphi$ iff $\mathscr{K} \mid\sim^i_{\mathscr{R},\mathsf{L}} \varphi$.*

7. Related Work and Conclusion

In [14], a sequent-style system for monotonic I/O logics without constraints is presented. It utilizes a correspondence between I/O and conditional logics. In [15] proof systems for constrained I/O logic are developed, where modalities for 'input' and 'output' allow for meta-reasoning in the object language. DAC uses labels instead of modalities and additionally allows for meta-reasoning about the (in)applicability of norms. In [9], sequent argumentation is used for defeasible reasoning with deontic logic. Norms are modelled with material implications which allows for less fine-tuning of norms than in DAC.

In [5,6,18] argumentative characterizations of normative systems employing priority orderings are studied. Their language is restricted to literals only, whereas our approach adopts a full propositional language. In [6,18] arguments consist only of (sets of) norms. In future work, we aim to incorporate priority and preference reasoning in the more transparent context of DAC. Moreover, the I/O formalism has other applications including reasoning with consistency checks, permissions, and constitutive norms [8,17]. In particular, we aim to exploit the internalization of meta-reasoning in DAC to characterize various types of permission [17], for instance, negative permissions as defined in terms of the absence of applicable norms to the contrary.

An alternative approach to model reasoning with norms is to instantiate ASPIC$^+$ [2] with conditionals representing norms and a defeasible modus ponens rule. This approach leads to a "greedier" style of reasoning than our approach. Consider $\mathscr{F} = \emptyset$ and $\mathscr{N} = \{(\top, p), (p, q), (\top, \neg q)\}$. An ASPIC$^+$-based approach yields the obligation to p with stable semantics since the argument for p from $(\top, p)$ is unchallenged. In contrast, our approach generates the argument $(\top, \neg q), (p, q) \Rightarrow \neg(\top, p)$ concluding the inapplicability of $(\top, p)$. The latter is in line with the I/O approach to normative reasoning.

We illustrated our approach with the notion of related admissibility [16]. For future work, we will investigate other argumentative approaches to explanation and how these can be used in the context of DAC, e.g., explicit reasoning about the inapplicability of norms in DAC can be harnessed to explain the non-acceptability of arguments [19].

Last, explanations typically occur in dialogues, through an exchange of reasons, questions, and arguments [20]. Consequently, explanations are often tailored to the background of the explainee. We will adopt our approach to dialogue models in future work.

In conclusion, in normative reasoning contexts one is not just interested in whether a specific obligation holds, but also in why it holds despite other norms to the contrary. To address this challenge, we developed Deontic Argument Calculi (DAC) which are rule-based proof systems that use labels to facilitate transparency and incorporate meta-normative reasoning with norms into the object language.

References

[1] Gabbay D, Horty J, Parent X, van der Meyden R, van der Torre L. Handbook of Deontic Logic and Normative Systems. vol. 1. United Kingdom: College Publications; 2013.

[2] Gabbay D, Giacomin M, Simari GR, Thimm M. Handbook of Formal Argumentation. vol. 2. United Kingdom: College Publications; 2021.

[3] Makinson D, van der Torre L. Constraints for Input/Output Logics. Journal of Philosophical Logic. 2001;30(2):155-85.

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

[5] Governatori G, Rotolo A, Riveret R. A deontic argumentation framework based on deontic defeasible logic. In: Poceedings of PRIMA 2018. Springer; 2018. p. 484-92.

[6] Liao B, Oren N, van der Torre L, Villata S. Prioritized norms in formal argumentation. Journal of Logic and Computation. 2018;29(2):215-40.

[7] Peirera C, Tettamanzi AG, Villata S, Liao B, Malerba A, Rotolo A, et al. Handling norms in multi-agent system by means of formal argumentation. IfCoLog. 2017;4(9):1-35.

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

[9] Straßer C, Arieli O. Normative reasoning by sequent-based argumentation. Journal of Logic and Computation. 2015 07;29(3):387-415.

[10] Parent X, van der Torre L. Introduction to deontic logic and normative systems. College Publications; 2018.
[11] Brunero J. Reasons, Evidence, and Explanations. Oxford Handbooks Online. 2018 Jul.
[12] Nair S, Horty J. The Logic of Reasons. Oxford Handbooks Online. 2018 Jul.
[13] Horty JF. Reasons as defaults. Oxford University Press, USA; 2012.
[14] Lellmann B. From Input/Output Logics to Conditional Logics via Sequents – with Provers. In: Automated Reasoning with Analytic Tableaux and Related Methods. Cham: Springer; 2021. p. 147-64.
[15] Straßer C, Beirlaen M, Van De Putte F. Adaptive logic characterizations of input/output logic. Studia Logica. 2016;104(5):869-916.
[16] Fan X, Toni F. On computing explanations in argumentation. In: AAAI; 2015. p. 1496-502.
[17] Tosatto SC, Boella G, van der Torre L, Villata S. Abstract normative systems: Semantics and proof theory. In: Proceedings of KR12; 2012. .
[18] Straßer C, Pardo P. Prioritized Defaults and Formal Argumentation. In: Proceedings of 14th International Conference of Deontic Logic and Normative Systems. College Publications; 2021. p. 427-46.
[19] Borg A, Bex F. A Basic Framework for Explanations in Argumentation. IEEE Int Systems. 2021:25-35.
[20] Walton D. A dialogue system specification for explanation. Synthese. 2010 Apr;182(3):349–374.