# A Kelsenian Deontic Logic

Agata CIABATTONI [a] Xavier PARENT [a] and Giovanni SARTOR [b]
[a] *Vienna University of Technology*
[b] *Cirsfid-Alma AI, University of Bologna; EUI, Florence*

**Abstract.** Inspired by Kelsen's view that norms establish causal-like connections between facts and sanctions, we develop a deontic logic in which a proposition is obligatory iff its complement causes a violation. We provide a logic for normative causality, define non-contextual and contextual notions of illicit and duty, and show that the logic of such duties is well-behaved and solves the main deontic paradoxes.

**Keywords.** Deontic Logic, Kelsen theory, Causality, Violations

## 1. Introduction

We develop a framework for deontic logic that combines violation and causality. Roughly speaking, an action is obligatory if refraining from performing it causes a violation. Our approach is inspired by the theory of norms developed by H. Kelsen, one of the most important legal scholars of the 20th century, in his "Pure Theory of Law," introduced in [1] and expanded in [2] (English translations in [3] and [4], respectively).

According to Kelsen, obligations and prohibitions are mere reflexes of sanction-norms: "the legal order [. . . ] prohibits a certain behavior by attaching to it a sanction or [. . . ] it commands a behavior by attaching a sanction to the opposite behavior" [4, p. 55]. This idea may be connected to the reduction of deontic logic to alethic modal logic as proposed by Anderson [5], though the latter does not refer to the work of Kelsen.

Here we use a reduction *à la* Anderson, but depart from it in two respects. First, we model the connection between a sanction (actually, a violation) and its triggering condition as a causal relationship, rather than as a necessity relationship, as proposed, e.g., in [6]. The necessity connection is usually understood as a strict implication which is known to generate counter-intuitive inferences, such as Ross's paradox [7]. It also satisfies the reflexivity postulate ("If $A$ then $A$"), which is often regarded as inappropriate for causal reasoning. Furthermore, Anderson's reduction is usually worked out within a possible worlds semantics, which is not compatible with Kelsen's view that norms do not bear a truth-value (see [8]). Second, in Anderson's perspective the same sanction is the consequent of each norm. In our approach instead different unlawful facts may lead to different violations. In this regard our approach corresponds to the working of legal and moral systems, where distinct unlawful or immoral acts lead to distinct sanctions or disvalues. This feature enables us to address contrary-to-duty CTD obligations, i.e., obligations which are applicable only if other obligations are violated. We can represent the original obligation by a norm linking the (prohibited) fact $f_1$ to a violation $v_1$, and the CTD obligation by a second norm linking the accomplishment of $f_1$ in combination with a further fact $f_2$ to an additional violation $v_2$. The first norm expresses the obligation of $\neg f_1$ (e.g., the obligation not to kill, in Forrester's famous paradox [9]), while the second norm expresses the prohibition of $f_2$, when $f_1$ is the case (the prohibition to be cruel, when killing). Hence we model CTD obligations by making the complement of their content into an aggravating circumstance, as in legal codes.

Before presenting our formal framework, we clarify some ideas about sanction and violation. Neither Kelsen nor Anderson claimed that every action for which a sanction is foreseen will necessarily be followed by an action of coercive enforcement by the state (forced execution, fine, detention, etc.), nor even by the pronouncement of a sanction by a competent authority. Once the condition for the sanction is realised, what necessarily happens is that (for Kelsen) the sanction is authorised and thus can legitimately be applied through the appropriate legal process, or (for Anderson) that something unwanted has happened. Hence we will model legal norms as connecting unlawful facts to violations, rather than to sanctions.

The paper is organised as follows. Sect. 2 presents the norms we deal with, which causally link (unlawful) states of affairs to violations, and the logic to reason about them. The latter is a simplified version of input-output logic [10], where part of the normative system consists of conditionals (regulative norms) having a violation constant on the right hand side. In Sect. 3 we discuss the notion of illicit, which is used to define contextual and non-contextual duties. In accordance with Kelsen's view that a normative system can be conflicting, Sect. 4 introduces the notion of (un-)obeyable system. The resulting logic of duties is analysed in Sect. 5, using as benchmarks well known properties and paradoxes from the deontic logic literature. Sect. 6 pinpoints a selection of topics for future research.

## 2. The Violation Logic

We introduce our base violation logic, starting with its language.

**Definition 1** (Language). *Let $L$ be an ordinary propositional language containing the classical connectives and constants $\{\wedge, \vee, \neg, \rightarrow, t, f\}$, and a set $V$ of violation atoms.*

Each violation atom denotes a particular violation or offence (or, following Kelsen, the authorisation to enact a specific sanction [4, 108ff]). Norms consist in causal-like connections, denoted by the $\Rightarrow$ symbol. They link factual circumstances to violations (regulative or violation norms) or Boolean antecedents to conclusions other than violations (constitutive or counts-as norms).

**Definition 2** (Norm code). *A* norm code *is a finite set of:*

- Violations norms*: $A \Rightarrow v$, where each $A$ is formula from $L \setminus V$ and $v \in V$.*
- Constitutive norms *(count-as norms): $A \Rightarrow B$, where $A, B$ are formulas from $L \setminus V$.*

**Example 1** (Auto code). *A norm code capturing simple road traffic rules is:*

| | |
|---|---|
| Speed $\Rightarrow v_1$ | Red $\wedge \neg$Stop $\Rightarrow v_2$ |
| Dark $\wedge \neg$LightsOn $\Rightarrow v_3$ | BrokenLights $\Rightarrow v_4$ |
| Phone $\wedge$ Drive $\Rightarrow v_5$ | BrokenFrontLights $\Rightarrow$ BrokenLights |
| BrokenBackLights $\Rightarrow$ BrokenLights | Fog $\wedge \neg$LightsOn $\Rightarrow v_6$ |

Note that in our example different norms establish different violations. This feature is meant to capture the fact that unlawful or immoral situations may trigger distinct responses by a legal and moral system. Such responses have to be added up to determine how the situation is assessed by the system (expanding the generated violations makes things worse). Our approach does not exclude that distinct facts may lead to the same violation. However, in this case the normative system would generate a single violation, whenever one, some, or all such norms are triggered. For instance, the violation $v_4$ may be triggered in two ways (via *BrokenFrontLights* or via *BrokenBackLights*). Should by contrast distinct fines be applied for broken front lights and broken back lights (to be added up when both are the case), different violations would be triggered by each of them.

We now define our violation logic, that specifies a causal-like entailment for norms, in the spirit of Kelsen (who calls this entailment "imputation", see [4, 76ff]).

**Definition 3** (Violation Logic)**.** *A violation inference relation is a binary relation* $\Rightarrow$ *between the set of propositions in L satisfying the following rules (*$\models$ *is the semantical consequence in classical propositional logic):*

*(Truth)* $t \Rightarrow t$
*(Strengthening) If* $A \models B$ *and* $B \Rightarrow C$*, then* $A \Rightarrow C$*;*
*(Weakening) If* $A \Rightarrow B$ *and* $B \models C$*, then* $A \Rightarrow C$*;*
*(And) If* $A \Rightarrow B$ *and* $A \Rightarrow C$ *then* $A \Rightarrow B \wedge C$*;*
*(Cut) If* $A \Rightarrow B$ *and* $A \wedge B \Rightarrow C$*, then* $A \Rightarrow C$*.*
*(Or) If* $A \Rightarrow C$ *and* $B \Rightarrow C$*, then* $A \vee B \Rightarrow C$*.*

Although causal relations satisfy most of the rules for classical entailment, their distinctive feature is that they are irreflexive, that is, they do not satisfy $A \Rightarrow A$. Actually, the above relation corresponds to the "basic reusable" input-output logic ($out_4$) from [10]. It is also closely related to Bochman's causal calculus [11], see Remark 1. The semantics for the violation logic is essentially the one for $out_4$. It is "operational", and takes the form of a set of procedures yielding outputs for inputs. Roughly speaking, to determine if formula $A$ is in the output set, one considers in turn each maximal consistent extension of the input set that is closed under the norms, and checks if it contains $A$. If the answer is "yes", then $A$ is in the output. This semantics fits Kelsen's idea that norms do not bear a truth-value and that logic cannot add new norms to a code $N$ [12, Ch. 50], but rather identifies the input-output connections established by $N$, which are the object of "rules (propositions) of law" (in German *Rechtssätze*)−see [4, p. 72] and [12, Ch. 49].

## 3. From Illicits to Duties

In Kelsen's legal theory sanction norms (violation norms, in our framework) have a foundational status. Other normative notions are derivative. Every behaviour that may trigger a sanction against its author is a delict (*Unrecht*) and every delict is the content of the obligation that the delict does not take place [2, p. 39]. Here we prefer to speak of an illicit, rather than of a delict, to cover all elements (not only actions) that contribute to the triggering of a sanction. Elementary illicits represent the minimal conditions that lead to a violation, and elementary duties apply to their negations.

**Definition 4** (Elementary Illicit and Duty)**.** *A conjunction of literals* $\bigwedge\{l_1,\ldots,l_n\}$ *is an* elementary illicit *relatively to a norm code N if and only if (iff) there is a* $v \in V$ *such that:* (1) $\bigwedge\{l_1,\ldots,l_n\} \Rightarrow v$*, and* (2) *no proper subset of* $\{l_1,\ldots,l_n\}$ *satisfies condition 1.*

*A proposition A is (the content of) an* elementary duty*, relatively to a norm code N, iff* $A \equiv \neg B$ *and B is an elementary illicit relatively to N.*

**Example 2.** *In Ex. 1,* $\{\text{Speed}\}$, $\{\text{Red}, \neg\text{Stop}\}$, $\{\text{Dark}, \neg\text{LightsOn}\}$, $\{\text{BrokenFrontLights}\}$, $\{\text{BrokenBackLights}\}$ *and* $\{\text{Phone}, \neg\ \text{LightsOn}\}$ *are elementary illicits ; their negations (*$\neg$Speed*,* $\neg\text{Red} \vee \text{Stop}$*, etc.) are the content of elementary duties.*

To develop a deontic logic we introduce the idea of generalized illicit, which covers *all* possible conditions that may minimally lead to a disjunction of violations.

**Definition 5** (Generalised illicit)**.** *A boolean formula $A$ is a generalised illicit relatively to a norm code $N$ if there exists a set of violations $S \subseteq V$ such that:* (1) $A \Rightarrow \bigvee S$*, and* (2) *there is no $B$ such that $A \models B$, $B \not\models A$ and $B \Rightarrow \bigvee S$.*

The rationale behind this notion is to ensure that a generalised illicit is not only a sufficient, but also a necessary condition for the disjunction of the violations to take place. Condition 2. makes $A$ the weakest formula leading to the violations in question. This will be the key element to the solution of the deontic paradoxes in Sect. 5.2.

**Example 3.** *Consider again Ex. 1. The generalized illicits w.r.t.* $\{v_4\}$ *and* $\{v_1, v_5\}$ *are:* $\text{BrokenFrontLights} \vee \text{BrokenBackLights}$ *and* $\text{Speed} \vee (\text{Phone} \wedge \text{Drive})$*, respectively. Note that* BrokenFrontLights *alone is not a generalized illicits as it is not the only possible way of triggering the violation $v_4$ (i.e., condition 2. in Def. 5 fails).*

Generalised illicts have the following logical properties: they are reducible to a disjunction of elementary illicits, and their disjunction constitutes a new generalised illict.

**Proposition 1.** *If $A$ is a generalised illicit relatively to a norm code $N$, then there exists a disjunction of elementary illicits $\{L_1, \ldots, L_n\}$ such that $A \equiv L_1 \vee \ldots \vee L_n$.*

*Proof.* Let $L_1, \ldots L_n$ be *all* possible elementary illicits relative to $N$ w.r.t. the violations $v \in S \subseteq V$. $L_1 \vee \ldots \vee L_n$ is a generalized illicit w.r.t. $\bigvee S$. Condition 1 of Def. 5 is satisfied. Indeed, if $L_i \Rightarrow v \in S$, then by $(Weakening)$ $L_i \Rightarrow \bigvee S$, and by $(Or)$ $L_1 \vee \ldots \vee L_n \Rightarrow \bigvee S$.

As for condition 2, assume there is a boolean formula $B$ such that $L_1 \vee \ldots \vee L_n \models B$ and $B \Rightarrow \bigvee S$. Assume w.l.o.g. that $B$ is in disjunctive normal form, say $B := N_1 \vee \cdots \vee N_m$. By $(Strenghtening)$ and the assumption $N_1 \vee \cdots \vee N_m \Rightarrow \bigvee S$ it follows that $N_i \Rightarrow \bigvee S$, for all $i = 1, \ldots, m$. Being the $L_i$'s all the elementary illicits triggering the violations in $S$ (and hence representing minimal conditions to trigger them), for each $N_i$ there are some literals $\{L_1, \ldots L_n\}$ which are included in the literals of $N_i$. Hence $B \models L_1 \vee \ldots \vee L_n$. $\square$

**Corollary 1.** *If $A_1$ and $A_2$ are generalized illicit relatively to $N$, so is $A_1 \vee A_2$.*

The concept of generalised illicits leads us to noncontextual duties, i.e., states of affairs whose non-realisation lead to a violation. More formally:

**Definition 6** (Noncontextual Duty)**.** *A Boolean formula $A$ is (the content of) a noncontextual duty relatively to a norm code $N$, denoted as $\mathbf{O}_N A$, iff $A \equiv \neg B$ and $B$ is a generalised illicit relatively to $N$.*

**Example 4** (Ctd. from Ex. 3)**.** $\mathbf{O}_N(\neg\ \text{BrokenFrontLights} \wedge \neg\text{BrokenBackLights})$ *and* $\mathbf{O}_N \neg(\text{Speed} \vee (\text{Fog} \wedge \text{LightsOn})) \equiv \mathbf{O}_N(\neg\text{Speed} \wedge (\text{Fog} \rightarrow \text{LightsOn}))$ *are duties.*

**Corollary 2.** *If $\mathbf{O}_N A$ then $A \equiv \neg L_1 \wedge \ldots \wedge \neg L_n$, for $L_1, \ldots, L_n$ elementary illicits.*

**Remark 1.** *$Out_4$ [10], and hence our violation logic, differ from the causal calculus in [11] by the presence in the latter of axiom $f \Rightarrow f$. This axiom would create the following counter-intuitive situation when considering the corresponding obligations. From $A \Rightarrow B$ follows $A \wedge \neg B \Rightarrow f$.*[1] *Now, by (Weakening), for any violation $v$, $A \wedge \neg B \Rightarrow v$ (using the fact that $f \models v$), which leads to $\mathbf{O}_N(\neg A \vee B)$, for any constitutive norm $A \Rightarrow B$.*

### 3.1. Putting Illicits and Duties in Context

We consider how norms operate relatively to contexts, i.e., in circumstances considered to be settled. We will regard contexts as kind of restrictions of the set of possible world, which are limited to those satisfying the context. These restrictions may depend on different considerations such as natural necessity (laws of nature), temporal necessity (the immutability of the past), or even the settled choices of the agent.

**Definition 7** (Context). *A context for a norm code is a consistent set of literals.*

Here we focus on the illicits that are not settled by the context (entailed by it), so that their happening is contingent on the choice (deliberation) of the involved agent.

**Definition 8** (Contextual Illicit). *A Boolean formula $A$ is a contextual illicit (c-illicit) relative to a norm code $N$ and context $C$ iff there is a set of violations $S \subseteq V$ s.t.*

1. $\bigwedge(C \cup \{A\}) \Rightarrow \bigvee S$
2. *there is no $B$ such that $A \models B$, $B \not\models A$ and* $\bigwedge(C \cup \{B\}) \Rightarrow \bigvee S$
3. $\bigwedge C \not\Rightarrow \bigvee S$ *and* $C \not\models \neg A$

Establishing that no weaker formula generates the considered violations, condition 2, is useful to resolve several deontic paradoxes, in particular those following from the assumption of closure of the deontic operator under logical consequence. Condition 3 tells us that $A$ is “needed” to generate the violation in question, and also that the truth or falsity of $A$ is not settled by the context, as shown in Remark 2.

**Remark 2.** *$\bigwedge C \not\Rightarrow \bigvee S$ in Def 8 (3) implies that $C \not\models A$. By condition 1 in Def 8, $\bigwedge(C \cup \{A\}) \Rightarrow \bigvee S$. Suppose $C \models A$. By propositional logic $\bigwedge C \models \bigwedge(C \cup \{A\})$. By (Strengthening), one gets $\bigwedge C \Rightarrow \bigvee S$. Contradiction.*

**Example 5.** *(Ctd. from Ex 1) In context $C = \{BrokenFrontLights\}$, $BrokenBackLights$ is not a c-illicit (due to the first half of condition 3 in Definition 8). The intuition is that when it is settled that one the two requirements leading to the same violation $v_4$ is met, meeting the other becomes irrelevant.*

On the basis of c-illicts we define contextual duties.

**Definition 9** (Contextual Duty). *A Boolean formula $A$ is a contextual duty relatively to a norm set $N$ and to a context $C$, denoted as $\mathbf{O}_{(N,C)}(A)$ iff $A \equiv \neg B$ and $B$ is a c-illicit relatively to $N$ and $C$ (we omit $N$ and $C$, when no ambiguity occurs).*

We apply below our approach to a pair of well-known deontic paradoxes pertaining to contrary-to-duty (CTD) scenarios (see Sect. 5.2 for more paradoxes).

---

[1] Proof: If $A \Rightarrow B$, then $A \wedge \neg B \Rightarrow B$ by (Strengthening). By $f \Rightarrow f$ and (Strengthening), $A \wedge \neg B \wedge B \Rightarrow f$. By (Cut), $A \wedge \neg B \Rightarrow f$.

**Example 6** (Forrester paradox [9])**.** *Consider the following premises: (1) You should not kill (2) If you kill, you should kill gently, (3) You kill. In Standard Deontic Logic SDL [13] (1)-(3) entail that you should both kill and not kill. By considering under what circumstance obligations (1) and (2) would be violated we have the violation norms*

$\text{Kill} \Rightarrow v_1 \qquad \text{Kill} \wedge \neg\text{KillGently} \Rightarrow v_2$

*where* KillGently $\Rightarrow$ Kill. *In context* $\{\}$, $\mathbf{O}\neg$Kill *holds, in context* $\{$Kill$\}$, $\mathbf{O}$KillGently.

**Example 7** (Chisholm paradox [14])**.** *It consists of: (1) You ought to go to the assistance of your neighbours; (2) If you go you ought to tell them that you are coming; but (3) If you do not go then you ought not to tell them that you are coming; and (4) You do not go. SDL [13] entails that both you ought and you ought not to tell your neighbours that you are coming. In our framework the norms involved in this scenario are formalised as:*

$\neg\text{Go} \Rightarrow v_1 \qquad \text{Go} \wedge \neg\text{TellGo} \Rightarrow v_2 \qquad \neg\text{Go} \wedge \text{TellGo} \Rightarrow v_3$

*In context* $\{\}$, $\mathbf{O}$Go *holds, in context* $\{\neg$Go$\}$, $\mathbf{O}\neg$TellGo *holds.*

**Remark 3.** *Some c-illicit $A$ relative to a norm code $N$ and context $C_1$, might not be a c-illicit relative to a superset $C_2$ of $C_1$ that is consistent with $A$ and such that $C_2 \not\models \vee S$, for any (sub)set $S$ of the violations in $N$. E.g., put $N = \{A \wedge B \Rightarrow v_1, D \wedge E \Rightarrow v_2\}$; $A \wedge B$ is a c-illicit in context $C_1 := \{E\}$, but not in $C_2 := \{A, E\}$ (condition 2 in Def. 8 fails).*

The following property, connecting violations entailed by contexts and duties, will be useful in Sect.5.2.

**Lemma 1.** *Let $N$ be a norm code. If $C \not\models \neg A_1 \vee \cdots \vee \neg A_n$, for all duties $\mathbf{O}_N A_1, \ldots, \mathbf{O}_N A_n$ then $\bigwedge C \not\Rightarrow \bigvee S$, for every $S \subseteq V$.*

*Proof.* By Def. 6 each $\neg A_i$ is a generalized illicit relative to $N$ and by Prop. 1 a disjunction of elementary illicits (minimal formulas that trigger violations). If there exists $S$ s.t. $\bigwedge C \Rightarrow \bigvee S$, there are generalized illicits $\neg A_1, \ldots \neg A_m$ such that $C \models \neg A_1 \vee \cdots \vee \neg A_m$. $\square$

## 4. (Un)Obeyability of Normative Codes

Kelsen pointed out that a normative code may establish requirements (cf. [15, p.25]) that cannot be jointly complied with: "within [...] a normative order the same behaviour may be [...] commanded and forbidden at the same time [...]. This is the case if a certain conduct is the condition of a sanction and at the same time the omission of this conduct is also the condition of a sanction." We formalize this intuition through two notions, absolute and contextual unobeyability.

**Definition 10** (Absolute (un)obeyability)**.** *A code $N$ is absolutely unobeyable iff $t \Rightarrow \bigvee V$, and it is absolutely obeyable otherwise.*

**Example 8.** *An absolutely unobeyable code is* $\{\text{Speed} \Rightarrow v_1; \neg\text{Speed} \Rightarrow v_2\}$. *Indeed by* $(Weakening)$ *and* $(Or)$ *we get $t \Rightarrow v_1 \vee v_2$.*

Absolutely unobeyable codes are rare, as they involve norms that always establish sanctions. A weaker and more common notion is that of contextual unobeyability. A code $N$ is unobeyable in a context $C$ iff $C$ entails that that a disjunction of alternative violations will be committed, not specifying which ones will be. Thus, the agent faces a predicament: possible violations can only be avoided by incurring in other violations.

**Definition 11** (Contextual unobeyability)**.** *A code $N$ is unobeyable in a context $C$ iff there is a set of violations $V_i \subseteq V$ such that:* (1) $\bigwedge C \Rightarrow \bigvee V_i$, *and* (2) *for each* $v \in V_i$, $\bigwedge C \not\Rightarrow v$.

**Example 9.** *The code* $N := \{\text{Speed} \Rightarrow v_1; \text{motorway} \wedge \neg\text{Speed} \Rightarrow v_2\}$ *is unobeyable in context* $\{\text{Motorway}\}$ *while being obeyable in context* $\{\}$ *(through* $\neg$Speed*). Since for no set of violations* $V_i \subseteq V$, $t \Rightarrow_N \bigvee V_i$, *N is absolutely obeyable.*

When a code $N$ is unobeyable in a context $C$, violations cannot be avoided in that context and the addressees of $N$ are forced to deliberate on which not yet settled violation to commit. They will have to face a "tragic dilemma", as in the biblical story below.

**Example 10** (from the Book of the Judges)**.** *Jephthah promised to God that if he was given victory in a battle he would sacrifice (kill and dedicate to God) the first human being that he encountered coming home. After winning the battle, he first encountered his daughter (rather than an animal, as he may have assumed). Thus, he faced a hard choice: either violate his promise to God, or violate the moral prohibition to kill his daughter. We can model this situation through the code* $N = \{\text{Win(j)} \wedge \text{Encounter(j,d)} \wedge \neg\text{Kill(j,d)} \Rightarrow v_1; \text{Kill(j,d)} \Rightarrow v_2\}$*, which is unobeyable in context* $C = \{\text{Win(j)}, \text{Encounter}(j,d)\}$.

Note that in a context of unobeyability, the agent has the duty to prevent each fact causing a new violation. However, he has no contextual duty to prevent the tautological disjunction of all such facts (i.e. to realise the contradictory conjunction of them). In fact such a disjunction is not a c-illicit, being entailed by the context (Remark 2). Thus, in the above example, in context $C$ Jephthah has duties $\mathbf{O}Kill(j,d)$ and $\mathbf{O}\neg Kill(j,d)$, but no duty $\mathbf{O}(Kill(j,d) \wedge \neg Kill(j,d))$.

## 5. The Logic of Duties

In this section we analyse the logic(s) of the $\mathbf{O}_N$ and $\mathbf{O}_{(N,C)}$ modalities. We consider the main principles discussed in the deontic logic literature, and check whether they hold. We also use some of the best-known deontic paradoxes as benchmarks.

### 5.1. Properties of Duties

For ease of readability, we consider first the version of a given property for noncontextual duty (i.e. unary obligation), when available.

Extensionality   For non-contextual duty it takes the form $(RE)$ "If $A \equiv B$, then $\mathbf{O}_N(A) \equiv \mathbf{O}_N(B)$". There are two versions for conditional duty: "If $\wedge C \equiv \wedge C'$, then $\mathbf{O}_{(N,C)}(A) \equiv \mathbf{O}_{(N,C')}(A)$" and "If $A \equiv B$, then $\mathbf{O}_{(N,C)}(A) \equiv \mathbf{O}_{(N,C)}(B)$". All versions trivially hold.

And introduction   If $\mathbf{O}_N A$ and $\mathbf{O}_N B$, then $\mathbf{O}_N(A \wedge B)$. Assume $\mathbf{O}_N A$ and $\mathbf{O}_N B$. So $A \equiv \neg A'$, and $B \equiv \neg B'$ for some $A', B'$ generalised illicits. We have $(A \wedge B) \equiv (\neg A' \wedge \neg B') \equiv \neg(A' \vee B')$. Furthermore, $A' \vee B'$ is a generalised illicit by Corollary 1. This suffices for $\mathbf{O}_N(A \wedge B)$. The analogue principle for contextual duties does not hold unrestrictedly. E.g., assume that $\mathbf{O}_{(N,C)}(A)$ and $\mathbf{O}_{(N,C)}(\neg A)$. $\mathbf{O}_{(N,C)}(A \wedge \neg A)$ is not a contextual duty as the negation of its content is equivalent to $t$, which is not a c-illicit (cond. 3 in Def. 8 fails)

**Remark 4.** *The logic of* $\mathbf{O}_N$ *duties contains the non-normal modal logic EC [16]. EC is obtained by adding the C axiom* $(\mathbf{O}_N A \wedge \mathbf{O}_N B) \rightarrow \mathbf{O}_N(A \wedge B)$ *to the system E of so-called classical modal logic, consisting of the sole rule of extensionality* $(RE)$.

<u>*Monotonicity wrt context*</u> It is the analogue of the "Strengthening of the antecedent" principle in conditional logic. It fails in its usual form "If $\mathbf{O}_{(N,C)}(A)$, and $C \subseteq C'$ then $\mathbf{O}_{(N,C')}(A)$", as hinted in Remark 3. The following counter-example shows that its failure is in line with the idea that in the context of deliberation one puts aside the moral status of the facts which are settled. Let $N = \{\neg A \Rightarrow v_1\}$. We have $\mathbf{O}_{(N,\emptyset)}A$, while $\mathbf{O}_{(N,\{\neg A\})}A$ no longer holds due to the failure of (the first half of) condition 3 in Def. 8.

<u>*Factual detachment*</u> (= detachment via modus ponens [17]) In our framework it might be expressed as: If $\mathbf{O}_{(N,C)}(A \to B)$ then $\mathbf{O}_{(N,C\cup\{A\})}B$, for $A$ being a conjunction of literals. It holds under the hypotheses: $(*)$ $C \cup \{A\} \not\models \neg A_1 \vee \cdots \vee \neg A_n$, for all duties $\mathbf{O}_N A_1, \ldots, \mathbf{O}_N A_n$, and $(**)$ there is no formula $D$ weaker than $\neg B$ s.t. $A \wedge D \models \neg B$. We show that $\neg B$ is a c-illicit in context $C \cup \{A\}$ so that $\mathbf{O}_{(N,C\cup\{A\})}B$. By $\mathbf{O}_{(N,C)}(A \to B)$ follows that there is $S \subseteq V$ s.t. $\bigwedge(C \cup \{A \wedge \neg B\}) \Rightarrow \bigvee S$, that is $\bigwedge(C \cup \{A\} \cup \{\neg B\}) \Rightarrow \bigvee S$. Hence Def. 8.1 holds for $\neg B$ relative to $C \cup \{A\}$. Condition 2 follows from the fact that if there is a $D'$ weaker than $\neg B$ such that $\bigwedge(C \cup \{A\} \cup \{D'\}) \Rightarrow \bigvee S$, then $A \wedge \neg B$ is not a c-illicit ($C \cup \{A \wedge D'\} \Rightarrow \bigvee S$ and from $\neg B \models D'$ follows $A \wedge \neg B \models A \wedge D'$, and from $(**)$ that $A \wedge D' \not\models A \wedge \neg B$) contradicting the hypothesis. Condition 3 also holds: the hypothesis $(*)$ guarantees by Lemma 1 that $\bigwedge(C \cup \{A\}) \not\Rightarrow \bigvee V$, where $V$ is the set of all violations, and hence $\bigwedge(C \cup \{A\}) \not\Rightarrow \bigvee S$; $C \cup \{A\} \not\models B$ follows from $C \not\models A \to B$.

<u>*No conflict*</u> $\neg(\mathbf{O}_N A \wedge \mathbf{O}_N \neg A)$ holds for a normative system $N$ that is absolutely obeyable in the sense of Definition 10. Suppose $N$ has this property. Assume by contradiction that both $\mathbf{O}_N A$ and $\mathbf{O}_N \neg A$ hold. Hence $A \equiv \neg B_1$, and $\neg A \equiv \neg B_2$ for some $B_1, B_2$ generalised illicits. Hence $B_1 \vee B_2 \Rightarrow \bigvee(V_1 \cup V_2)$, and so $t \Rightarrow \bigvee(V_1 \cup V_2)$, since $B_1 \equiv \neg A$ and $A \equiv B_2$. By $(Weakening)$, $t \Rightarrow \bigvee V$, where $V$ is the set of all violations. Contradiction.

The contextual version of the principle –$\neg(\mathbf{O}_{(N,C)}A \wedge \mathbf{O}_{(N,C)}\neg A)$– holds for normative systems that are contextually obeyable in $C$. Assume by contradiction that so is $N$ and that $\mathbf{O}_{(N,C)}A$ and $\mathbf{O}_{(N,C)}\neg A$ hold. Being $A$ and $\neg A$ contextual illicits w.r.t. $C$, there are $V_1$ and $V_2$ s.t. $(*)$ $\bigwedge(C \cup \{A\}) \Rightarrow \bigvee V_1$ and $\bigwedge(C \cup \{\neg A\}) \Rightarrow \bigvee V_2$, and $(**)$ $\wedge C \not\Rightarrow \bigvee V_1$ and $\wedge C \not\Rightarrow \bigvee V_2$. From $(*)$ it follows that $\bigwedge C \Rightarrow \bigvee V_1 \cup V_2$ (i.e., condition (1) of Def. 11), and by $(**)$ that for each $v \in V_1 \cup V_2$, $\bigwedge C \not\Rightarrow v$ (i.e. condition (2) of Def. 11), thus establishing that $N$ is unobeyable in context $C$. This property echoes the fact that for Kelsen normative systems can be conflicting. Thus, for him, and in our framework, the "no conflict" axiom (also known as D axiom in modal logic) only holds for normative systems that are either absolutely or contextually obeyable.

### *5.2. Solutions to Deontic Paradoxes*

We analyse the behaviour of our logic w.r.t. the main deontic paradoxes that have beset SDL, which corresponds to the deontic logic obtained by Anderson using his reduction schema. Deontic paradoxes are intended here in a broad sense as (un)derivable theorems that are counter-intuitive in a common-sense reading.

Recall that unconstrained I/O logic [10], which is at the base of our framework, is unable to handle many of the considered paradoxes. Two paths have been followed in order to fix it: using constraints to filter excess outputs [18], or suitably weakening the logic. In particular, (Weakening) and axiom (Truth) go, and at the same time a consistency check restrains the application of some rules (like AND introduction). Although based on the strongest system of unconstrained I/O logic (Def. 2), our framework eliminates those paradoxes due to condition 2. in Definitions 5 and 8.

*Contrary-to-duty obligation (Forester's and Chisholm's) paradoxes* As seen before, in our framework we can overcome such paradoxes. In Forrester's case (see Ex. 6) in context $\{kill\}$, $\mathbf{O}\mathit{KillGently}$ holds, and the conflicting obligation $\mathbf{O}\neg \mathit{Kill}$ does not hold. The treatment of Chisholm's scenario is similar (see Ex. 7).

*Ross' paradox* The paradox consists in the derivation of (a) You should mail the letter or burn it, from (b) You should mail the letter. Introduced in 1941 by Ross [19], this paradox has been a discussion topic ever since. It does not appear in our logics. E.g., relative to a normative system $N = \{\neg \mathit{PostLetter} \Rightarrow v_1\}$, we have the obligation $\mathbf{O}\mathit{PostLetter}$, but we do not have $\mathbf{O}(\mathit{PostLetter} \vee \mathit{BurnLetter})$, due to condition 2. in Def. 5.

*Deontic detachment [17]* Deontic detachment is paradoxical in it usual form: $\mathbf{O}A$ and $\mathbf{O}(A \to B)$ then $\mathbf{O}B$. This can be illustrated with Broome [20]'s counter-example: let $N = \{\neg \mathit{Exercise} \Rightarrow v_1; \mathit{Exercise} \wedge \neg \mathit{EatMore} \Rightarrow v_2\}$ (one should exercise and if one exercises, one should eat more). We have $\mathbf{O}\mathit{Exercise}$ and $\mathbf{O}(\mathit{Exercise} \to \mathit{EatMore})$. But it should not be the case that $\mathbf{O}\mathit{EatMore}$, since, intuitively, the obligation to eat more holds only if the first obligation is fulfilled. The correct inference is captured by the following form of deontic detachment, called "aggregative" in [21], which keeps track of what has been previously detached: from $\mathbf{O}A$ and $\mathbf{O}(A \to B)$, infer $\mathbf{O}(A \wedge B)$. In our logic only the latter deontic detachmens holds due to *AND introduction* and *Extensionality*. Hence from $N$ we can infer $\mathbf{O}(\mathit{Exercise} \wedge \mathit{EatMore})$. The contextual version of the paradoxical deontic detachment (if $\mathbf{O}_{(N)}A$ and $\mathbf{O}_{(N,A)}B$ then $\mathbf{O}_{(N)}B$), does not hold either in our logic. For a counter-example, let $N = \{\neg a \Rightarrow v_1, a \wedge \neg b \Rightarrow v_2\}$. We have $\mathbf{O}a$ and $\mathbf{O}_a b$, but not $\mathbf{O}b$, as $\neg b$ is not a generalized illicit (condition 1 in Def. 5 fails).

*The alternative service paradox* This scenario, proposed by Horty in 1994 [22], is handled similarly to the previous case, by changing the disjunctive obligation in the original formulation into a conditional obligation: from (1) You should fight in the army or perform alternative service ($\mathbf{O}(\neg A \to B)$), and (2) You should not fight in the army ($\mathbf{O}\neg A$), we derive that (3) You should not fight and perform alternative service ($\mathbf{O}(\neg A \wedge B)$).

*Necessitation* $\mathbf{O}t$ does not hold. Take any non-empty set of norms, e.g., $N = \{\mathit{Speed} \Rightarrow v_1\}$. $\mathbf{O}(\mathit{Speed} \vee \neg \mathit{Speed})$ does not hold, because $\mathit{Speed} \wedge \neg \mathit{Speed}$ is not a generalised illicit, due to condition 2. in Def. 5.

*Reflexivity* This is the law $\mathbf{O}_{(N,C)} \wedge C$. It fails, because of condition 3 in Def 8 ($C \models \wedge C$, see Remark 2) and so $\neg \wedge C$ cannot be a $c$-illicit relative to $C$.

*And elimination* It is the principle "If $\mathbf{O}(A \wedge B)$ then $\mathbf{O}A$", known in the non-normal modal logic literature as axiom M (see [16]). As suggested by several authors, when $A$ and $B$ are not separable, such a principle is counter-intuitive, so that (to quote Hansen) "failing a part [of the order] means that satisfying the remainder no longer makes sense. E.g. if I am to satisfy the imperative 'buy apples and walnuts', and the walnuts [...] and the apples [are meant to] land in a Waldorf salad, then it might be unwanted and a waste of money to buy the walnuts if I cannot get the apples" [23, p. 91]. Put $N = \{\neg \mathit{Apples} \vee \neg \mathit{Walnuts} \Rightarrow v\}$ (or equivalently $\{\neg \mathit{Apples} \Rightarrow v, \neg \mathit{Walnuts} \Rightarrow v\}$). In our logic $\mathbf{O}(\mathit{Apples} \wedge \mathit{Walnuts})$ does not entail $\mathbf{O}(\mathit{Apples})$, due to condition (2) in Def. 5.

## 6. Future Work

We have developed a Kelsenian deontic logic that defines obligation in terms of violation and causality. The behaviour of the resulting obligation has been analysed using as benchmarks well known properties and paradoxes from the deontic logic literature.

It has often been argued that CTDs involve different senses of "ought", and that a fully adequate treatment of them must be able to capture those nuances. In our analysis of contextual duties we have focused on deliberative duties, where one puts aside the moral or legal status of the facts which are taken as settled and one must decide what to do. Our framework allows, however, a finer-grained analysis of duties –to be left to future work– which could not only shed more light on the various paradoxes, but also capture different types of "ought"-statements. Among them, we plan to investigate contextual (evaluative) duties that are appropriate for what is commonly referred to as the "context of judgement" [24,25], where one assesses the moral or legal status of settled facts through backward looking or *post-eventum* judgments [26, p. 157].

Furthermore, our logic is an ought-to-be deontic logic, as obligations may cover any cause of violations. We plan to develop an ought-to-do deontic logic, reflecting Kelsen's notion of a delict, by carving out agentive elements within illicits and "plugging" in a suitable logic of action in our base logic. Other topics for future research include the question of extending the framework to support defeasible reasoning, reasoning about exceptions, and the question of how to axiomatise and automatise the logic.[2]

## References

[1] Kelsen H. Reine Rechtslehre Einleitung in die rechtswissenschaftliche Problematik. Mohr Siebeck; 1934.
[2] Kelsen H. Reine Rechtslehre. Franz Deuticke; 1960.
[3] Kelsen H. Introduction to the Problems of Legal Theory. Clarendon; 1992.
[4] Kelsen H. The Pure Theory of Law. University of California Press; 1967.
[5] Anderson AR. The Logic of Norms. Logique et Analyse. 1958;1:84–91.
[6] Wang PH. Anderson's reduction and Kelsen's Normativism. In: Pluralism and Law. Proceedings of the 20th IVR World Congress 2001. Vol. 4: Legal Reasoning. Franz Steiner Verlag; 2001. p. 104–110.
[7] Ross WD. Foundations of Ethics. Clarendon; 1939.
[8] Kelsen H. General Theory of Norms. Oxford University Press; 1990.
[9] Forrester JW. Gentle Murder, or the adverbial Samaritan. J of Philosophy. 1984;(81):193–197.
[10] Makinson D, van der Torre LWN. Input/Output Logics. J Phil Log. 2000;29(4):383–408.
[11] Bochman A. On Laws and Counterfactuals in Causal Reasoning. In: Proc. 16 Intern. Conf. on Principles of Knowledge Representation and Reasoning (KR 2018); 2018. p. 494–503.
[12] Kelsen H. General Theory of Norms. Clarendon; 1991.
[13] von Wright GH. Deontic logic. Mind. 1951;60(237):1–15.
[14] Chisholm RM. Contrary-to-Duty Imperatives and Deontic Logic. Analysis. 1963;24:33–6.
[15] Kelsen H. On the pure theory of law. Israel Law Review. 1966:193–215.
[16] Chellas B. Modal Logic. Cambridge University Press; 1980.
[17] Greenspan PS. Conditional Oughts and Hypothetical Imperatives. Journal of Philosophy. 1975;72(10):259–276.
[18] Makinson D, van der Torre LWN. Constraints for Input/Output Logics. J Phil Log. 2001;30(2):155–185.
[19] Ross A. Imperatives and Logic. Theoria. 1941;7(1):53.
[20] Broome J. Rationality through Reasoning. West Sussex, U: Wiley-Blackwell; 2013.
[21] Parent X, van der Torre LWN. "Sing and Dance!" - Input/Output Logics without Weakening. In: Cariani F, et Al, editors. DEON 2014, Proceedings. vol. 8554 of LNCS. Springer; 2014. p. 149–165.
[22] Horty J. Moral Dilemmas and Nonmonotonic Logic. Journal of Philosophical Logic. 1994;23(1):35–65.
[23] Hansen J. Imperatives and Deontic logic. University of Leipzig; 2008.
[24] Thomason RH. In: Hilpinen R, editor. Deontic Logic and the Role of Freedom in Moral Deliberation. Dordrecht: Springer Netherlands; 1981. p. 177–186.
[25] Asher N, Bonevac D. *Prima Facie* Obligation. Stud Logica. 1996;57(1):19–45.
[26] Hare R. The Language of Morals. Clarendon; 1991.

[2]G. Sartor has been supported by the H2020 ERC Project "CompuLaw" (G.A. 833647), A. Ciabattoni and X. Parent by the WWTF project MA16-028. We thank two anonymous referees for their comments.