

Aggregative Deontic Detachment for Normative Reasoning

Xavier Parent and Leendert van der Torre

University of Luxembourg
6, rue Richard Coudenhove-Kalergi
L-1359 Luxembourg

Abstract

Aggregative deontic detachment is a new form of deontic detachment that keeps track of previously detached obligations. We argue that it handles iteration of successive detachments in a more principled manner than the traditional systems do. To study this new form of deontic detachment, we introduce a ‘minimal’ logic for aggregative deontic detachment, and we discuss various properties of the logic.

Aggregative Deontic Detachment

The *Handbook of Deontic Logic and Normative Systems* (Gabbay et al. 2013) reports on-going development of new standards for the formal study of normative reasoning. In his chapter “Alternative semantics for deontic logic”, Hansson writes:

“Deontic logic has ... a strong semantic tradition. Using possible worlds and orderings (preference relations) as their major building-blocks, deontic logicians have constructed models that can be used to determine the validity of deontic sentences. ... But, unfortunately, deontic logic also has a long history of getting stuck in semantic principles that support blatantly implausible deontic postulates. This is probably a major reason why the influence of deontic logic ... has been rather weak.” (Hansson 2013, p. 447)

In the past, deontic logic seems to have been heavily, perhaps one-sidedly, influenced by the modal logic paradigm, to which Hansson alludes here. However, given all the difficulties the traditional systems face, deontic logic is currently undergoing a paradigm shift from modal logic to what may be called “norm-based semantics” (Hansen 2014). The core idea is to explain the truths of deontic logic not by some set of possible worlds among which some are ideal or at least better than others, but with reference to an explicit set of given norms or existing moral standards. The present short paper embraces such a paradigm shift.

To study aggregative deontic detachment we introduce a new logic for normative reasoning. It builds on so-called input/output (I/O) logic – one of the candidate systems

for a new standard described in the *Handbook* (Parent and van der Torre 2013). Detachment (or modus-ponens) is the core mechanism of the semantics being used. The law of detachment (or modus-ponens) is well-known from propositional logic: from a conditional statement and its antecedent, the consequent of the conditional statement is inferred. In the deontic logic literature, this is referred to as “factual” detachment. The rule may be given the following form, where the conditional obligation for x given a is written as $\bigcirc(x \mid a)$:

$$\text{FD} \frac{\bigcirc(x \mid a) \quad a}{\bigcirc x}$$

Van Eck asked, “How can we take seriously a conditional obligation if it cannot, by way of detachment, lead to an unconditional obligation?” (van Eck 1982, p. 263). Thus, it is no wonder that factual detachment is accepted as valid in many traditional systems of deontic logic, either in its plain form or in a qualified one. However, the traditional systems make some extra assumptions that are potentially controversial.

This can be illustrated with the Danielsson-Hansson-Lewis logic for conditional obligation (Danielsson 1968; Hansson 1969; van Fraassen 1972; Lewis 1973), which until recently was mainstream. First, it shares with the so-called classical theory of rational choice the assumption that an individual has (well-defined) preferences, and that a normative judgment is based on a maximization process. Advocates of –as Simon (1957) terms it– “bounded rationality” have argued that an approach to rationality in terms of maximization is not realistic, because human beings lack the cognitive resources to optimize. Second, such a logic takes the so-called trichotomy thesis for granted. It is the assumption that comparable items (worlds) can only be better than, worse than, or equal to each other in overall value. The trichotomy view has been challenged more than once. Some have argued that these three value relations do not exhaust the space of possibilities. The best-known proposal for a fourth *sui generis* relation is Chang’s argument for “on a par” –see Chang (2002).

The best way to avoid potential objections is to make as few assumptions as possible. We believe that the assumption that conditionals obey the detachment rule is one that can hardly be challenged. Obligations and permissions are

contextual and vary based on the setting. Consequently, a norm always takes the form of a conditional statement. Some philosophers like Boghossian (2000) think (rightly, in our view) that the disposition to reason according to detachment is constitutive of the possession of the concept of conditional, and thus of the concept of norm. The idea is that, if some agent says “if a then x ”, and if he truly means it, then he commits himself to detaching x given a . If this agent refuses to acknowledge that he is justified in employing detachment, this will be good evidence that he fails to understand what is meant by “if ... then”. Accepting detachment and acquiring an implication are two sides of the same coin.

The present paper focuses on the question of how to handle iterations of successive detachments. The standard I/O systems handle them, by injecting the following rule, known as “deontic detachment”:

$$\text{DD} \frac{\frac{\circ(y \mid x) \quad \circ x}{\circ y}}{\circ y}$$

That is fully in line with the tradition in deontic logic. For instance, the Danielsson-Hansson-Lewis semantics for conditional obligation validates such a law. However, potential counterexamples to DD may be found in the literature (McLaughlin 1955; Hansson 1997; Makinson 1999). They all rely on the intuition that the obligation of y ceases to hold when the obligation of x is violated. This can be illustrated with the following example, due to Broome (2013, §7.4):

You ought to exercise hard everyday
 If you exercise hard everyday, you ought to eat heartily
 ?* You ought to eat heartily

Intuitively, the obligation to eat heartily no longer holds, if you do not take exercise. Like the others, this counterexample suggests an alternative (call it “aggregative”) form of detachment, which keeps track of what has been previously detached:

$$\text{ADD} \frac{\frac{\circ(y \mid x) \quad \circ x}{\circ(x \wedge y)}}{\circ(x \wedge y)}$$

This form of detachment has been overlooked in the literature.

To study aggregative deontic detachment we define a system supporting ADD, but not DD. Accepting ADD, but not DD, implies rejecting the rule known as “weakening”. This is the rule, where \vdash stands for the deducibility relation in propositional logic:

$$\text{W} \frac{\frac{\circ(x \mid a) \quad x \vdash y}{\circ(y \mid a)}}{\circ(y \mid a)}$$

ADD and W together imply DD.¹

We provide not only a suitable semantics, but also a sound and complete axiomatization for the new logic. Due to space limitations we do not give the soundness and completeness proofs here, but we do present some properties of the new logic.

¹(van der Torre and Tan 1998) define a system PDL validating ADD, but not DD. However, PDL has a preference-based semantics.

Logic for Aggregative Deontic Detachment

In input/output logic, a normative code is a set N of conditional obligations. A conditional obligation is a pair (a, x) , where a and x are two formulae of classical propositional logic. We use this notation instead of $\circ(x \mid a)$, because the latter has distinct interpretations in the literature. In the notation (a, x) , the first element a is called the body of the rule, and is thought of as an input, representing some condition or situation. The second element x is called the head of the rule, and is thought of as an output, representing what the norm tells us to be obligatory in that situation. We use the standard notation (\top, x) for the unconditional obligation of x , where \top is a zero-place connective standing for ‘tautology’.

Some further notation. \mathcal{L} is the set of all formulae of classical propositional logic. Given an input $A \subseteq \mathcal{L}$, and a set of generators N , $N(A)$ denotes the image of N under A , i.e., $N(A) = \{x : (a, x) \in N \text{ for some } a \in A\}$. $Cn(A)$ denotes the set $\{x : A \vdash x\}$, where \vdash is the deducibility relation used in classical propositional logic. The notation $x \dashv\vdash y$ is short for $x \vdash y$ and $y \vdash x$. We use PL as an abbreviation for (classical) propositional logic. For the sake of conciseness, we denote by $h(M)$ the set of all the heads of elements of M . Moreover, \mathcal{B}_A^M denotes the set of all B such that $A \subseteq B = Cn(B) \supseteq M(B)$. Intuitively, \mathcal{B}_A^M gathers all the B s that contain A and are closed under both Cn and M .

Definition 1 (Semantics). $x \in \mathcal{O}(N, A)$ if and only if there is some finite $M \subseteq N$ such that,

- $M(Cn(A)) \neq \emptyset$, and
- for all B , if $B \in \mathcal{B}_A^M$, then $x \dashv\vdash \bigwedge M(B)$.

We call such a M an A -witness for x . \mathcal{O} is a variation of the operation out_3 defined by Makinson and van der Torre (2000). We have $out_3(N, A) = \bigcap \{Cn(N(B)) : B \in \mathcal{B}_A^N\}$.

While out_3 validates DD, \mathcal{O} does not, as illustrated below.

Example 1 (Deontic Detachment). Consider the two norms $N = \{(\top, a), (a, x)\}$. We have obligation $a \in \mathcal{O}(N, \top)$, since $M = \{(\top, a)\}$ is a \top -witness for a . We also have $x \in \mathcal{O}(N, a)$, since $M = \{(a, x)\}$ is an a -witness for x . But we do not have $x \in \mathcal{O}(N, \top)$. To see this, it suffices to go through all the non-empty subsets of N , and check that, for each of them, the smallest relevant B does not deliver heads whose conjunction is equivalent to x :

M	B	$M(B)$
$\{(\top, a)\}$	$Cn(a)$	$\{a\}$
$\{(a, x)\}$	$Cn(\emptyset)$	\emptyset
$\{(\top, a), (a, x)\}$	$Cn(a, x)$	$\{a, x\}$

We define $\mathcal{O}(N) = \{(A, x) : x \in \mathcal{O}(N, A)\}$. This definition leads to an axiomatic characterization that is much alike those used for conditional logic.

Given a set R of rules, (a, x) is said to be derivable from N if (a, x) is in the least superset of N that is closed under the rules in R . This may be written as $(a, x) \in \mathcal{D}(N)$, or equivalently $x \in \mathcal{D}(N, a)$. The specific rules of interest here are described below. They are formulated for a singleton input set A (for such an input set, curly brackets will be omitted). SI, EQ and ACT abbreviate “strengthening of the input,” “equivalence” and “aggregative cumulative transitivity.”

Definition 2 (Proof system). $(a, x) \in \mathcal{D}(N)$ if and only if (a, x) is derivable from N using the rules $\{SI, EQ, ACT\}$.

$$SI \frac{(a, x) \quad b \vdash a}{(b, x)} \quad EQ \frac{(a, x) \quad x \dashv\vdash y}{(a, y)}$$

$$ACT \frac{(a, x) \quad (a \wedge x, y)}{(a, x \wedge y)}$$

ACT yields ADD as a special case (a is \top).

Where A is a set of formulae, $(A, x) \in \mathcal{D}(N)$ means that $(a, x) \in \mathcal{D}(N)$, for some conjunction a of elements in A . Moreover, $\mathcal{D}(N, A)$ is $\{x : (A, x) \in \mathcal{D}(N)\}$.

Theorem 1 (Completeness). $\mathcal{O}(N, A) = \mathcal{D}(N, A)$

Below is an illustration.

Example 2 (“Change your mind!”). Hansson (1997) gives the following example, with credit to Pörn:

“Consider the hoary example of the man who ought to go to a meeting on August 5 and who ought to send, on August 2, a note explaining his absence, if and only if he is in fact going to be absent.” (1997, p.425-6)

The example is structurally identical to the Chisholm example (Chisholm 1963). The norms involved may be rendered as $N = \{(\top, m), (m, s)(\neg m, \neg s)\}$, where m and s are for attending the meeting and sending a note, respectively. Given input \top , $m \wedge s$ is outputted, but not s . This is as it should be. The obligation of s will not be triggered unless the agent is going to fulfil his primary obligation of m . In the violation context $\neg m$, $m \wedge s$ is still outputted. If not, then the following intuitive deontic reasoning pattern would not be supported:

“August 2 arrives, and though he is able to attend the meeting, he has no intention of doing so. He argues: ‘I ought to change my mind, forbear note-writing, and attend the meeting. ... My present fulfillment of this obligation will help make up for my sinfully staying at home on the fifth!’.” (Hansson 1997, p. 426)

In the remainder of this paper, we consider some properties of this new logic.

Properties

Deontic properties

Exact factual detachment (EFD) and violation detection (VD) characterise what is special about *deontic* logic.

$$EFD \quad (x, y) \in N \Rightarrow y \in \mathcal{O}(N, x)$$

$$VD \quad (A, y) \Rightarrow (A \cup \{\neg y\}, y)$$

Exact factual detachment represents that if there is a norm with precisely the context as body, then the output contains the head. On the one hand, exact factual detachment is relatively weak, as the context has to be *precisely* the body. The much stronger factual detachment principle (FD) discussed in the introduction imposes detachment when the body is *implied* by the context. On the other hand, exact factual detachment principle is already quite strong, as in context a from the norm (a, \perp) the contradiction \perp is detached, and

in case of a dilemma of (a, x) and $(a, \neg x)$, in context a both x and $\neg x$ are detached.

Violation detection says that violations are explicit. The distinctive feature of norms and obligations with respect to other types of rules and modalities is that they can be violated. For example, dilemma examples arise because some obligation has to be violated, and contrary-to-duty examples arise because some obligation has been violated. Modal logic offers a useful representation for violations. An obligation for x has been violated if and only if we have $\neg x \wedge \bigcirc x$. In our notation with explicit norms, this is $x \in \mathcal{O}(N, A)$ with $\neg x \in Cn(A)$.

Theorem 2. \mathcal{O} satisfies EFD, FD and VD.

Logical properties

Substitution (SUB), replacements of logical equivalents (RLE), implication (IMP) and paraconsistency (PC) say something about *logic*.

$$SUB \quad x \in \mathcal{O}(N, A) \Rightarrow x[\sigma] \in \mathcal{O}(N[\sigma], A[\sigma])$$

$$RLE \quad N \approx M \Rightarrow \mathcal{O}(N) \subseteq \mathcal{O}(M)$$

$$IMP \quad \mathcal{O}(N, A) \subseteq Cn(m(N) \cup A)$$

$$PC \quad x \in \overline{V}(N, A) \Rightarrow \exists M \subseteq N : x \in \mathcal{O}(M, A)$$

and $\mathcal{O}(M, A) \cup A$ consistent

Substitution is well known from classical propositional logic. We use the notation $x[\sigma]$ to denote a substitution instance of x . Thus, $x[\sigma]$ is obtained from x by replacing uniformly, in x , all occurrences of a propositional letter by the same propositional formula. $A[\sigma]$ and $N[\sigma]$ extend the notion of substitution instance to sets of formulae, and sets of norms in the straightforward way.

Replacement for logically equivalent expressions expresses a principle of irrelevance of syntax. For $N \approx M$, read “ N is (logically) equivalent to M ”. The simplest way to define \approx is as follows: $N \approx M$ whenever M is obtained from N , by replacing each $(b, y) \in N$ with some (c, z) such that b is equivalent with c , and y is equivalent with z .

Implication makes use of the so-called materialisation of a normative system, which means that each norm (a, x) is interpreted as a material conditional $a \rightarrow x$, i.e. as the propositional sentence $\neg a \vee x$. The implication property says that if the materializations of N , written as $m(N)$, do not imply $a \rightarrow x$, then $(a, x) \notin \mathcal{O}(N)$. This represents the idea that we cannot derive more than we can derive in propositional logic. In general, implication in the base logic is the upper bound.

To prevent explosion we do not want to derive the whole language, unless maybe in pathological cases in which the normative system contains a norm for each propositional formula. A consequence relation may be said to be paraconsistent if it is not explosive, though there are various ways to make this formal.

To define our paraconsistency property, we distinguish obligations representing violations from other obligations. That is, we decompose an operator $\mathcal{O}(N, A)$ into two operators $V(N, A)$ and $\overline{V}(N, A)$, such that

- $V(N, A) = \{x \in \mathcal{O}(N, A) \mid \neg x \in Cn(A)\}$
- $\overline{V}(N, A) = \mathcal{O}(N, A) \setminus V(N, A)$

Trivially, we have

$$\mathcal{O}(N, A) = V(N, A) \cup \bar{V}(N, A)$$

The basic idea of our paraconsistency property is that obligations in \bar{V} can be derived from a set of norms M in N , such that this set of norms M does not explode.

The underlying intuition to restrict to a set of norms is raised in the following example. If we can derive $\bigcirc(y \wedge z)$ from $\bigcirc(x \wedge y)$ and $\bigcirc(\neg x \wedge z)$, and we have substitution and replacements of logical equivalents, then we also derive $\bigcirc y$ from $\bigcirc x$ and $\bigcirc \neg x$, in other words, we have deontic explosion. This can be verified by replacing y by $x \vee y$ and z by $\neg x \vee y$. Therefore, we restrict the set of norms we use to a set of norms which is in some sense “consistent” with the input A .

Theorem 3. \mathcal{O} satisfies SUB, RLE, IMP and PC.

Norm change properties

The next two properties are: norm monotony (NM) and norm induction (NI). We call them “norm change properties”, because the N is no longer held constant.

$$\begin{array}{l} \text{NM} \quad \mathcal{O}(N) \subseteq \mathcal{O}(N \cup M) \\ \text{NI} \quad M \subseteq \mathcal{O}(N) \Rightarrow \mathcal{O}(N) = \mathcal{O}(N \cup M) \end{array}$$

The most interesting property is norm induction. It says that if there is an output x for an input a , and we add the norm (a, x) to the normative system, then for all inputs, the output of the normative system stays the same. This is called “norm induction”, because the norm is induced from the relation between facts and obligations.

Theorem 4. \mathcal{O} satisfies NM and NI.

Together, exact factual detachment, norm monotony and norm induction are equivalent to saying that \mathcal{O} is a closure operator.

Reusability properties

The reusability properties relate the system to standard I/O logic. Their formulation appeals to some key notions of so-called constrained input/output logic, developed by Makinson and van der Torre (2001) in order to reason about norm violation:

$$\begin{array}{l} \text{conf}(N, A) = \{N' \subseteq N \mid \text{out}(N', A) \cup A \text{ consistent}\} \\ \text{maxf}(N, A) = \{N' \in \text{conf}(N, A) \mid N' \subseteq \text{-maximal}\} \\ \text{outf}(N, A) = \{\text{out}(N, A) \mid N' \in \text{maxf}(N, A)\} \end{array}$$

The first property listed below represents inclusion in reusable output (IO). The following two properties define redundancy R and strong redundancy SR.

$$\begin{array}{l} \text{IO} \quad \mathcal{O}(N, A) \subseteq \text{out}_3(N, A) \\ \text{R} \quad \text{outf}_3(N, A) = \text{outf}_3(\mathcal{O}(N), A) \\ \text{SR} \quad \text{outf}_3(N \cup M, A) = \text{outf}_3(\mathcal{O}(N) \cup M, A) \end{array}$$

Redundancy is worth a mention, because it makes a bridge to constrained input/output logic.

Summary

This paper has introduced a ‘minimal’ logic for aggregative deontic detachment, with a sound and complete axiomatization. Various properties of the logic have been discussed. Summarizing, the operator $\mathcal{O}(N)$ is a closure operator, which means that it satisfies factual detachment, norm

monotony and norm induction. In addition, it satisfies substitution and replacement of logical equivalents. Finally, the operator satisfies violation detection, implication, and paraconsistency.

References

- Boghossian, P. 2000. Knowledge of logic. In Boghossian, P., and Peacocke, C., eds., *New Essays on the A Priori*. Oxford: Clarendon Press. 229–254.
- Broome, J. 2013. *Rationality Through Reasoning*. West Sussex, UK: Wiley-Blackwell.
- Chang, R. 2002. The possibility of parity. *Ethics* 112:659–688.
- Chisholm, R. 1963. Contrary-to-duty imperatives and deontic logic. *Analysis* 24:33–36.
- Danielsson, S. 1968. *Preference and Obligation: Studies in the Logic of Ethics*. Uppsala: Filosofiska freningen.
- Gabbay, D.; Horty, J.; Parent, X.; van der Meyden, R.; and van der Torre, L., eds. 2013. *Handbook of Deontic Logic and Normative Systems*. London, UK: College Publications.
- Hansen, J. 2014. Reasoning about permission and obligation. In Hansson, S. O., ed., *David Makinson on Classical Methods for Non-Classical Problems*. Springer. 287–333.
- Hansson, B. 1969. An analysis of some deontic logics. *Noûs* 3:373–398.
- Hansson, S. O. 1997. Situationist deontic logic. *Journal of Philosophical Logic* 26(4):423–448.
- Hansson, S. 2013. Alternative semantics for deontic logic. In Gabbay et al. (2013). 445–498.
- Lewis, D. 1973. *Counterfactuals*. Oxford: Blackwell.
- Makinson, D., and van der Torre, L. 2000. Input/output logics. *Journal of Philosophical Logic* 29(4):383–408.
- Makinson, D., and van der Torre, L. 2001. Constraints for input/output logics. *Journal of Philosophical Logic* 30(2):155–185.
- Makinson, D. 1999. On a fundamental problem in deontic logic. In Namara, P. M., and Prakken, H., eds., *Norms, Logics and Information Systems*, Frontiers in Artificial Intelligence and Applications. Amsterdam: IOS Press. 29–54.
- McLaughlin, R. N. 1955. Further problems of derived obligation. *Mind* 64(255):400–402.
- Parent, X., and van der Torre, L. 2013. Input/output logic. In Gabbay et al. (2013). 499–544.
- Simon, H. 1957. *Models of Man, Social and Rational*. New York: John Wiley and Sons.
- van der Torre, L., and Tan, Y.-H. 1998. Prohairesic deontic logic (PDL). In Dix, J.; Fariñas del Cerro, L.; and Furbach, U., eds., *Logics in Artificial Intelligence*. Berlin Heidelberg: Springer. 77–91.
- van Eck, J. 1982. A system of temporally relative modal and deontic predicate logic, and its philosophical applications. *Logique & Analyse* 25:249–290 and 339–381.
- van Fraassen, B. C. 1972. The logic of conditional obligation. *Journal of Philosophical Logic* 1:417–438.