Ten Philosophical Problems in Deontic Logic

Jörg Hansen[1], Gabriella Pigozzi[2] and Leendert van der Torre[2]

[1] University of Leipzig, Institut für Philosophie
04107, Leipzig, Beethovenstraße 15, Germany
jhansen@uni-leipzig.de

[2] University of Luxembourg, Computer Science and Communications (CSC)
1359, Luxembourg, 6 rue Richard Coudenhove Kalergi, Luxembourg
{gabriella.pigozzi,leon.vandertorre}@uni.lu

Abstract. The paper discusses ten philosophical problems in deontic logic: how to formally represent norms, when a set of norms may be termed 'coherent', how to deal with normative conflicts, how contrary-to-duty obligations can be appropriately modeled, how dyadic deontic operators may be redefined to relate to sets of norms instead of preference relations between possible worlds, how various concepts of permission can be accommodated, how meaning postulates and counts-as conditionals can be taken into account, and how sets of norms may be revised and merged. The problems are discussed from the viewpoint of input/output logic as developed by van der Torre & Makinson. We argue that norms, not ideality, should take the central position in deontic semantics, and that a semantics that represents norms, as input/output logic does, provides helpful tools for analyzing, clarifying and solving the problems of deontic logic.

Keywords. Deontic logic, normative systems, input/output logic

Introduction

Deontic logic is the field of logic that is concerned with normative concepts such as obligation, permission, and prohibition. Alternatively, a deontic logic is a formal system that attempts to capture the essential logical features of these concepts. Typically, a deontic logic uses Ox to mean that it is obligatory that x, (or it ought to be the case that x), and Px to mean that it is permitted, or permissible, that x. The term 'deontic' is derived from the ancient Greek *déon*, meaning that which is binding or proper.

So-called Standard Deontic Logic (SDL) is a normal propositional modal logic of type KD, which means that it extends the propositional tautologies with the axioms $K : O(x \to y) \to (Ox \to Oy)$ and $D : \neg(Ox \wedge O\neg x)$, and it is closed under the inference rules *modus ponens* $x, x \to y/y$ and Necessitation x/Ox. Prohibition and permission are defined by $Fx = O\neg x$ and $Px = \neg O\neg x$. SDL is an unusually simple and elegant theory. An advantage of its modal-logical setting is that it can easily be extended with other modalities like epistemic or temporal operators and modal accounts of actions.

Dagstuhl Seminar Proceedings 07122
Normative Multi-agent Systems
http://drops.dagstuhl.de/opus/volltexte/2007/941

Not surprisingly for such a highly simplified theory, there are many features of actual normative reasoning that SDL does not capture. Notorious are the so-called 'paradoxes of deontic logic', which are usually dismissed as consequences of the simplifications of SDL. E.g. Ross's paradox [48], the counterintuitive derivation of "you ought to mail or burn the letter" from "you ought to mail the letter", is typically viewed as a side effect of the interpretation of 'or' in natural language. Many researchers seem to believe that the subject of deontic logic may be more or less finished, and we can focus on the use of deontic logic in computer science and agent theory, since there is nothing important left to add to it. In our view, this is far from the truth. On the contrary, there is a large number of important open problems in this field of research.

In this paper we discuss ten philosophical problems in deontic logic. All of these problems have been discussed in previous literature, and solutions have been offered, but we believe that all of them should be considered open and thus meriting further research. These problems are how deontic logic relates or applies to given sets of norms (imperatives, rules, aims) (sec. 1), what it means that a set of norms should be coherent (sec. 2), how conflicts of norms can be taken into account (sec. 3), how deontic logic should react to contrary-to-duty situations in which some norms are invariably violated (sec. 4), how to interpret dyadic deontic operators that formalize 'it ought to be that x on conditions α' as $O(x/\alpha)$ (sec. 5), how explicit permissions relate to, and change, an agent's obligations (sec. 6), how meaning postulates – norms that define legal terms – and constitutive norms, that create normative states of affairs, can be modeled (sec. 7 and 8), and how normative systems may be revised (sec. 9) and merged (sec. 10). Our choice is motivated by our aim at providing ourselves with models of normative reasoning of actual agents which may be human beings or computers, but the list of open problems is by no means final. Other problems may be considered equally important, such as how a hierarchy of norms (or of the norm-giving authorities) is to be respected, or how general norms relate to individual obligations, but we hope that our discussion provides the tools, and encourages the reader, to take a fresh look at these other problems, too.

To illustrate the problems, we use Makinson & van der Torre's input/output logic as developed in [42], [43], [44], and we therefore assume familiarity with this approach (cf. [45] for a good introduction). Input/output logic takes a very general view at the process used to obtain conclusions (more generally: outputs) from given sets of premises (more generally: inputs). While the transformation may work in the usual way, as an 'inference motor' to provide logical conclusions from a given set of premises, it might also be put to other, perhaps non-logical uses. Logic then acts as a kind of secretarial assistant, helping to prepare the inputs before they go into the machine, unpacking outputs as they emerge, and, less obviously, coordinating the two. The process as a whole is one of logically assisted transformation, and is an inference only when the central transformation is so. This is the general perspective underlying input/output logic. It is one of logic at work rather than logic in isolation; not some kind of non-classical logic, but a way of using the classical one.

1 Jørgensen's dilemma

While normative concepts are the subject of deontic logic, it is quite difficult how there can be a logic of such concepts at all. Norms like individual imperatives, promises, legal statutes, moral standards etc. are usually not viewed as being true or false. E.g. consider imperative or permissive expressions such as "John, leave the room!" and "Mary, you may enter now": they do not describe, but demand or allow a behavior on the part of John and Mary. Being non-descriptive, they cannot meaningfully be termed true or false. Lacking truth values, these expressions cannot – in the usual sense – be premise or conclusion in an inference, be termed consistent or contradictory, or be compounded by truth-functional operators. Hence, though there certainly exists a logical study of normative expressions and concepts, it seems there cannot be a logic of norms: this is Jørgensen's dilemma ([30], cf. [41]).

Though norms are neither true nor false, one may state that *according to the norms*, something ought to be (be done) or is permitted: the statements "John ought to leave the room", "Mary is permitted to enter", are then true or false descriptions of the normative situation. Such statements are sometimes called normative statements, as distinguished from norms. To express principles such as the principle of conjunction: $O(x \wedge y) \leftrightarrow (Ox \wedge Oy)$, with Boolean operators having truth-functional meaning at all places, deontic logic has resorted to interpreting its formulas Ox, Fx, Px not as representing norms, but as representing such normative statements. A possible logic of normative statements may then reflect logical properties of underlying norms – thus logic may have a "wider reach than truth", as von Wright [54] famously stated.

Since the truth of normative statements depends on a normative situation, like the truth of the statement "John ought to leave the room" depends on whether some authority ordered John to leave the room or not, it seems that norms must be represented in a logical semantics that models such truth or falsity. But semantics used to model the truth or falsity of normative statements mostly fail to include norms. Standard deontic semantics evaluates deontic formulas with respect to sets of worlds, in which some are ideal or better than others – Ox is then defined true if x is true in all ideal or the best reachable worlds. In our view, norms, not ideality, should take the central position from which normative statements are evaluated. Then the following question arises, pointedly asked by D. Makinson in [41]:

Problem 1. How can deontic logic be reconstructed in accord with the philosophical position that norms are neither true nor false?

In the older literature on deontic logic there has been a veritable 'imperativist tradition' of authors that have, deviating from the standard approach, in one way or other, tried to give truth definitions for deontic operators with respect to given sets of norms.[3] The reconstruction of deontic logic as logic about imperatives

[3] Cf. among others S. Kanger [32], E. Stenius [53], T. J. Smiley [51], Z. Ziemba [62], B. van Fraassen [15], Alchourrón & Bulygin [1] and I. Niiniluoto [47].

has been the project of one of the authors beginning with [19]. Makinson & van der Torre's input/output logic [42] is another reconstruction of a logic of norms in accord with the philosophical position that norms direct rather than describe, and are neither true nor false. Suppose that we have a set G (meant to be a set of conditional norms), and a set A of formulas (meant to be a set of given facts). The problem is then: how may we reasonably define the set of propositions x making up the output of G given A, which we write $out(G, A)$? In particular, if we view the output as descriptions of states of affairs that ought to obtain given the norms G and the facts A, what is a reasonable output operation that enables us to define a deontic O-operator that describes the normative statements that are true given the norms and the facts, we say: the normative consequences given the situation? One such definition is the following:

$$G, A \models Ox \text{ iff } x \in out(G, A)$$

So Ox is true iff the output of G under A includes x. Note that this is rather a description of how we think such an output should or might be interpreted, whereas 'pure' input/output logic does not discuss such definitions. For a simple case, let G include a conditional norm that states that if a is the case, x should obtain (we write $(a, x) \in G$).[4] If a can be inferred from A, i.e. if $a \in Cn(A)$, and z is logically implied by x, then z should be among the normative consequences of G given A. An operation that does this is simple-minded output out_1:

$$out_1(G, A) = Cn(G(Cn(A)))$$

where $G(B) = \{y \mid (b, y) \in G \text{ and } b \in B\}$. So in the given example, Oz is true given $(a, x) \in G$, $a \in Cn(A)$ and $z \in Cn(x)$.

Simple-minded output may, however, not be strong enough. Sometimes, legal argumentation supports reasoning by cases: if there is a conditional norm (a, x) that states that an agent must bring about x if a is the case, and a norm (b, x) that states that the same agent must also bring about x if b is the case, and $a \vee b$ is implied by the facts, then we should be able to conclude that the agent must bring about x. An operation that supports such reasoning is basic output out_2:

$$out_2(G, A) = \cap\{Cn(G(V)) \mid v(A) = 1\}$$

where v ranges over Boolean valuations plus the function that puts $v(b) = 1$ for all formulae b, and $V = \{b \mid v(b) = 1\}$. It can easily be seen that now Ox is true given $\{(a, x), (b, x)\} \subseteq G$ and $a \vee b \in Cn(A)$.

It is quite controversial whether reasoning with conditional norms should support 'normative' or 'deontic detachment', i.e. whether it should be accepted that if one norm (a, x) commands an agent to make x true in conditions a, and another norm (x, y) directs the agent to make y true given x is true, then the agent has an obligation to make y true if a is factually true. Some would argue that as long as the agent has not in fact realized x, the norm to bring about y is not 'triggered'; others would maintain that obviously the agent has an obligation to make $x \wedge y$ true given that a is true. If such detachment is viewed

[4] As has become usual, an unconditional norm that commits the agent to realizing x is represented by a conditional norm $(\top, x)$, where $\top$ means an arbitrary tautology.

as permissible for normative reasoning, then one might use reusable output out_3 that supports such reasoning:

$$out_3(G, A) = \cap\{Cn(G(B)) \mid A \subseteq B = Cn(B) \supseteq G(B)\}$$

An operation that combines reasoning by cases with deontic detachment is then reusable basic output out_4:

$$out_4(G, A) = \cap\{Cn(G(V)) : v(A) = 1 \text{ and } G(V) \subseteq V\}$$

Finally, it is often required to reconsider the facts when drawing conclusions about what an agent must do: suppose there is an unconditional norm $(\top, x \vee y)$ to bring about $x \vee y$, but that the agent cannot realize x as the facts include $\neg x$. We would like to say that then the agent must bring about y, as this is the only possible way left to satisfy the norm. To do this, one may use the throughput versions out_n^+ of any of the output operations $out_1, out_2, out_3, out_4$,

$$out_n^+(G, A) = out_n(G^+, A),$$

where $G^+ = G \cup I$ and I is the set of all pairs (a, a) for formulae a. The choice of the throughput versions might appear questionable, since each makes Ox true in case $x \in Cn(A)$, i.e. it makes the unalterable facts obligatory.

It may turn out that further modifications of the output operation are required in order to produce reasonable results for normative reasoning. Also, the proposal to employ input/output logic to reconstruct deontic logic may lead to competing solutions, depending on what philosophical views as to what transformations should be acceptable one subscribes to. All this is what input/output logic is about. However, it should be noted that input/output logic succeeds in representing norms as entities that are neither true nor false, while still permitting normative reasoning about such entities.

2 Coherence

Consider norms which on one hand require you to leave the room, while on the other requiring you not to leave the room at the same time. In such cases, we are inclined to say that there is something wrong with the normative system. This intuition is captured by the SDL axiom $D : \neg(Ox \wedge O\neg x)$ that states that there cannot be co-existing obligations to bring about x and to bring about $\neg x$, or, using the standard cross-definitions of the deontic modalities: x cannot be both, obligatory and forbidden, or: if x is obligatory then it is also permitted. But what does this tell us about the normative system?

Since norms do not bear truth values, we cannot, in any usual sense, say that such a set of norms is inconsistent. All we can consider is the consistency of the output of a set of norms. We like to use the term *coherence* with respect to a set of norms with consistent output, and define:

(1) A set of norms G is coherent iff $\perp \notin out(G, A)$.

However, this definition seems not quite sufficient: one might argue that one should be able to determine whether a set of norms G is coherent or not regardless of what arbitrary facts A might be assumed. A better definition would be $(1a)$:

(1a) A set of norms G is coherent iff there exists a set of formulas A such that $\bot \notin out(G, A)$.

For (1a) it suffices that there exists a situation in which the norms can be, or could have been, fulfilled. However, consider the set of norms $G = \{(a, x), (a, \neg x)\}$ that requires both x to be realized and $\neg x$ to be realized in conditions a: it is immediate that e.g. for all output operations $out_n^{(+)}$, we have $\bot \notin out_n^{(+)}(G, \neg a)$: no conflicting demands arise when $\neg a$ is factually assumed. Yet something seems wrong with a normative system that explicitly considers a fact a only to tie to it conflicting normative consequences. The dual of (1a) would be

(1b) A set of norms G is coherent iff for all sets of formulas A, $\bot \notin out(G, A)$.

Now a set G with $G = \{(a, x), (a, \neg x)\}$ would no longer be termed coherent. (1b) makes the claim that for no situation A, two norms $(a, x), (b, y)$ would ever come into conflict, which might seem too strong. We may wish to restrict A to sets of facts that are consistent, or that are not in violation of the norms. The question is, basically, how to distinguish situations that the norm-givers should have taken care of, from those that describe misfortune of otherwise unhappy circumstances. A weaker claim than (1b) would be (1c):

(1c) A set of norms G is coherent iff for all a with $(a, x) \in G$, $\bot \notin out(G, a)$.

By this change, consistency of output is required just for those factual situations that the norm-givers have foreseen, in the sense that they have explicitly tied normative consequences to such facts. Still, (1c) might require further modification, since if a is a foreseen situation, and so is b, then also $a \vee b$ or $a \wedge b$ might be counted as foreseen situations for which the norms should be coherent.

However, there is a further difficulty: let G contain a norm $(a, \neg a)$ that, for conditions in which a is unalterably true, demands that $\neg a$ be realized. We then have $\neg a \in out_n(G, a)$ for the principal output operations out_n, but not $\bot \in out_n(G, a)$. Certainly the term 'incoherent' should apply to a normative system that requires the agent to accomplish what is – given the facts in which the duty arises – impossible. But since not every output operation supports 'throughput', i.e. the input is not necessarily included in the output, neither (1) nor its variants implies that the agent can actually realize all propositions in the output, though they might be logically consistent. We might therefore demand that the output is not consistent *simpliciter*, but consistent with the input:

(2) A set of norms G is coherent iff $\bot \notin out(G, A) \cup A$.

But with definition (2) we obtain the questionable result that for any case of norm-violation, i.e. for any case in which $(a, x) \in G$ and $(a \wedge \neg x) \in Cn(A)$, G must be termed incoherent – Adam's fall would only indicate that there was something wrong with God's commands. One remedy would be to leave aside all those norms that are invariably violated, i.e. instead of $out(G, A)$ consider $out(\{(a, x) \in G \mid (a \wedge \neg x) \notin Cn(A)\}, A)$ – but then a set G such that $(a, \neg a) \in G$ would not be incoherent. It seems it is time to formally state our problem:

Problem 2. When is a set of norms to be termed 'coherent'?

As can be seen from the discussion above, input/output logic provides the tools to formally discuss this question, by rephrasing the question of coherence of the norms as one of consistency of output, and of output with input. Both notions have been explored in the input/output framework as 'output under constraints':

Definition (Output under constraints) *Let G be a set of conditional norms and A and C two sets of propositional formulas. Then G is coherent in A under constraints C when $out(G, A) \cup C$ is consistent.*

Future study must define an output operation, determine the relevant states A, and find the constraints C, such that any set of norms G would be appropriately termed coherent or incoherent by this definition.

3 Normative conflicts and dilemmas

There are essentially two views on the question of normative conflicts: in the one view, they do not exist. In the other view, conflicts and dilemmas are ubiquitous.

According to the view that normative conflicts are ubiquitous, it is obvious that we may become the addressees of conflicting normative demands at any time. My mother may want me to stay inside while my brother wants me to go outside with him and play games. I may have promised to finish a paper until the end of a certain day, while for the same day I have promised a friend to come to dinner – now it is late afternoon and I realize I will not be able to finish the paper if I visit my friend. Social convention may require me to offer you a cigarette when I am lighting one for myself, while concerns for your health should make me not offer you one. Legal obligations might collide - think of the recent case where the SWIFT international money transfer program was required by US anti-terror laws to disclose certain information about its customers, while under European law that also applied to that company, it was required not to disclose this information. Formally, let there be two conditional norms (a, x) and (b, y): unless we have that either $(x \rightarrow y) \in Cn(a \wedge b)$ or $(y \rightarrow x) \in Cn(a \wedge b)$ there is a possible situation $a \wedge b \wedge \neg(x \wedge y)$ in which the agent can still satisfy each norm individually, but not both norms collectively. But to assume the former for any two norms (a, x) and (b, y) is clearly absurd.[5] So any logic about norms must take into account possible conflicts. But standard deontic logic SDL includes D: $\neg(Ox \wedge O\neg x)$ as one of its axioms, and it is not quite immediate how deontic reasoning could accommodate conflicting norms. The problem is thus:

Problem 3a. How can deontic logic accommodate possible conflicts of norms?

In an input/output setting one could say that there exists a conflict whenever $\bot \in Cn(out(G, A) \cup A)$, i.e. whenever the output is inconsistent with the input: then the norms cannot all be satisfied in the given situation. There appear to be two ways to proceed when such inconsistencies cannot be ruled out.[6]. For both, it is necessary to recur to the the notion of a *maxfamily*(G, A, A), i.e. the

[5] Nevertheless, Lewis' [36], [37] and Hansson's [24] deontic semantics imply that there exists a 'system of spheres', in our setting: a sequence of boxed contrary-to-duty norms $(\top, x_1), (\neg x_1, x_2), (\neg x_1 \wedge \neg x_2, x_3), \ldots$ that satisfies this condition.

[6] For the concepts underlying the 'some-things-considered' and 'all-things-considered' O-operators defined below cf. Horty [28] and Hansen [20], [21]

family of all maximal $H \subseteq G$ such that $out(H, A) \cup A$ is consistent. On this basis, input/output logic defines the following two output operations $out^{\cup}$ and $out^{\cap}$:

$$\begin{array}{rcl} out^{\cup}(G, A) & = & \bigcup\{out(H, A) \mid H \in maxfamily(G, A, A)\} \\ out^{\cap}(G, A) & = & \bigcap\{out(H, A) \mid H \in maxfamily(G, A, A)\} \end{array}$$

Note that $out^{\cup}$ is a non-standard output operation that is not closed under consequences, i.e. we do not generally have $Cn(out^{\cup}(G, A)) = out^{\cup}(G, A)$. Finally we may use the intended definition of an O-operator

$$G, A \models Ox \text{ iff } x \in out(G, A)$$

to refer to the operations $out^{\cup}$ and $out^{\cap}$, rather than the underlying operation $out(G, A)$ itself, and write $O^{\cup}x$ and $O^{\cap}x$ to mean that $x \in out^{\cup}(G, A)$ and $x \in out^{\cap}(G, A)$, respectively. Then the 'some-things-considered', or 'bold' O-operator $O^{\cup}$ describes x as obligatory given the set of norms G and the facts A if x is in the output of some $H \in maxfamily(G, A, A)$, i.e. if some subset of non-conflicting norms, or: some coherent normative standard embedded in the norms, requires x to be true. It is immediate that neither the SDL axiom $D : \neg(Ox \wedge O\neg x)$ nor the agglomeration principle $C : Ox \wedge Oy \rightarrow O(x \wedge y)$ holds for $O^{\cup}$, as there may be two competing standards demanding x and $\neg x$ to be realized, while there may be none that demands the impossible $x \wedge \neg x$. On the other hand, the 'all-things-considered', or 'sceptic', O-operator $O^{\cap}$ describes x as obligatory given the norms G and the facts A if x is in the outputs of all $H \in maxfamily(G, A, A)$, i.e. it requires that x must be realized according to all coherent normative standards. Note that by this definition, both SDL theorems D and C are validated.

The opposite view, that normative conflicts do not exist, appeals to the very notion of obligation: it is essential for the function of norms to direct human behavior that the subject of the norms is capable of following them. To state a norm that cannot be fulfilled is a meaningless use of language. To state two norms which cannot both be fulfilled is confusing the subject, not giving him or her directions. To say that a subject has two conflicting obligations is therefore a misuse of the term 'obligation'. So there cannot be conflicting obligations, and if things appear differently, a careful inspection of the normative situation is required that resolves the dilemma in favor of the one or other of what only appeared both to be obligations. In particular, this inspection may reveal a priority ordering of the apparent obligations that helps resolve the conflict (this summarizes viewpoints prominent e.g. in Ross [49], von Wright [59], [60], and Hare [25]). The problem that arises for such a view is then how to determine the 'actual obligations' in face of apparent conflicts, or, put differently, in the face of conflicting 'prima facie' obligations.

Problem 3b. How can the resolution of apparent conflicts be semantically modeled?

Again, both the $O^{\cup}$ and the $O^{\cap}$-operator may help to formulate and solve the problem: $O^{\cup}$ names the conflicting *prima facie* obligations that arise from a set of norms G in a given situation A, whereas $O^{\cap}$ resolves the conflict by telling the agent to do only what is required by all maximal coherent subsets of the norms: so there might be conflicting 'prima facie' $O^{\cup}$-obligations, but no conflicting 'all things considered' $O^{\cap}$-obligations. The view that a priority ordering helps to resolve conflicts seems more difficult to model. A good approach appears to be to let the priorities help us to select a set $\mathscr{P}(G, A, A)$ of preferred maximal subsets $H \in \mathit{maxfamily}(G, A, A)$. We may then define the $O^{\cap}$-operator not with respect to the whole of $\mathit{maxfamily}(G, A, A)$, but only with respect to its selected preferred subsets $\mathscr{P}(G, A, A)$. Ideally, in order to resolve all conflicts, the priority ordering should narrow down the selected sets to $\mathit{card}(\mathscr{P}(G, A, A)) = 1$, but this generally requires a strict ordering of the norms in G. The demand that all norms can be strictly ordered is itself subject of philosophical dispute: some moral requirements may be incomparable (this is Sartre's paradox, where the requirement that Sartre's student stays with his ailing mother conflicts with the requirement that the student joins the resistance against the German occupation), while others may be of equal weight (e.g. two simultaneously obtained obligations towards identical twins, of which only one can be fulfilled). The difficult part is then to define a mechanism that determines the preferred maximal subsets by use of the given priorities between the norms. There have been several proposals to this effect, not all of them successful, and the reader is referred to the discussions in Boella & van der Torre [8] and Hansen [22], [23].

4 Contrary-to-duty reasoning

Suppose we are given a code G of conditional norms, that we are presented with a condition (input) that is unalterably true, and asked what obligations (output) it gives rise to. It may happen that the condition is something that should not have been true in the first place. But that is now water under the bridge: we have to "make the best out of the sad circumstances" as B. Hansson [24] put it. We therefore abstract from the deontic status of the condition, and focus on the obligations that are consistent with its presence. How to determine this in general terms, and if possible in formal ones, is the well-known problem of contrary-to-duty conditions as exemplified by the notorious contrary-to-duty paradoxes. Chisholm's paradox [13] consists of the following four sentences:

(1) It ought to be that a certain man go to the assistance of his neighbors.
(2) It ought to be that if he does go, he tell them he is coming.
(3) If he does not go then he ought not to tell them he is coming.
(4) He does not go.

Furthermore, intuitively, the sentences derive (5):

(5) He ought not to tell them he is coming.

Chisholm's paradox is a contrary-to-duty paradox, since it contains both a primary obligation to go, and a secondary obligation not to call if the agent does not go. Traditionally, the paradox was approached by trying to formalize each of the sentences in an appropriate language of deontic logic, and then consider the sets $\{Ox, O(x \rightarrow z), O(\neg x \rightarrow \neg z), \neg x\}$, or $\{Ox, x \rightarrow Oz, \neg x \rightarrow O\neg z, \neg x\}$, or $\{Ox, O(x \rightarrow z), \neg x \rightarrow O\neg z, \neg x\}$ or $\{Ox, x \rightarrow Oz, O(\neg x \rightarrow \neg z), \neg x\}$. But whatever approach is taken, it turned out that either the set of formulas is traditionally inconsistent or inconsistent in SDL, or one formula is a logical consequence – by traditional logic or in SDL – of another formula. Yet intuitively the natural-language expressions that make up the paradox are consistent and independent from each other: this is why it is called a paradox. Though the development of dyadic deontic operators as well as the introduction of temporally relative deontic logic operators can be seen as a direct result of Chisholm's paradox, the paradox seems so far unsolved. The problem is thus:

Problem 4. How do we reason with contrary-to-duty obligations which are in force only in case of norm violations?

In the input/output logic framework, the strategy for eliminating excess output is to cut back the set of generators to just below the threshold of yielding excess. To do that, input/output logic looks at the maximal non-excessive subsets, as described by the following definition:

Definition (Maxfamilies) *Let G be a set of conditional norms and A and C two sets of propositional formulas. Then $maxfamily(G, A, C)$ is the set of maximal subsets $H \subseteq G$ such that $out(H, A) \cup C$ is consistent.*

For a possible solution to Chisholm's paradox, consider the following output operation $out^{\cap}$:

$$out^{\cap}(G, A) \;=\; \bigcap\{out(H, A) \mid H \in maxfamily(G, A, A)\}$$

So an output x is in $out^{\cap}(G, A)$ if it is in output $out(H, A)$ of all maximal norm subsets $H \subseteq G$ such that $out(H, A)$ is consistent with the input A. Let a deontic O-operator be defined in the usual way with regard to this output:

$$G, A \models O^{\cap}x \;\text{ iff }\; x \in out^{\cap}(G, A)$$

Furthermore, tentatively, and only for the task of shedding light on Chisholm's paradox, let us define an entailment relation between norms as follows:

Definition (Entailment relation) *Let G be a set of conditional norms, and (a, x) be a norm whose addition to G is under consideration. Then (a, x) is entailed by G iff for all sets of propositions A, $out^{\cap}(G \cup \{(a, x)\}, A) = out^{\cap}(G, A)$.*

So a (considered) norm is entailed by a (given) set of norms if its addition to this set would not make a difference for any set of facts A. Finally, let us use the following cautious definition of 'coherence from the start' (also called 'minimal coherence' or 'coherence per se'):

A set of norms G is 'coherent from the start' iff $\bot \notin out(G, \top)$.

Now consider a 'Chisholm norm set' $G = \{(\top, x), (x, z), (\neg x, \neg z), \}$, where $(\top, x)$ means the norm that the man must go to the assistance of his neighbors, (x, z) means the norm that it ought to be that if he goes he ought to tell them he is coming, and $(\neg x, \neg z)$ means the norm that if he does not go he ought not to tell them he is coming. It can be easily verified that the norm set G is 'coherent from the start' for all standard output operations $out_n^{(+)}$, since for these either $out(G, \top) = Cn(\{x\})$ or $out(G, \top) = Cn(\{x, z\})$, and both sets $\{x\}$ and $\{x, z\}$ are consistent. Furthermore, it should be noted that all norms in the norm set G

are independent from each other, in the sense that no norm $(a, x) \in G$ is entailed by $G \setminus \{(a, x)\}$ for any standard output operation $out_n^{(+)}$: for $(\top, x)$ we have $x \in out^{\cap}(G, \top)$ but $x \notin out^{\cap}(G \setminus \{(\top, x)\}, \top)$, for (x, z) we have $z \in out^{\cap}(G, x)$ but $z \notin out^{\cap}(G \setminus \{(x, z)\}, x)$, and for $(\neg x, \neg z)$ we have $\neg z \in out^{\cap}(G, \neg x)$ but $\neg z \notin out^{\cap}(G \setminus \{(\neg x, \neg z)\}, \top)$. Finally consider the 'Chisholm fact set' $A = \{\neg x\}$, that includes as an assumed unalterable fact the proposition $\neg x$, that the man will not go to the assistance of his neighbors: we have $maxfamily(G, A, A) = \{G \setminus \{(\top, x)\}\} = \{\{(x, z), (\neg x, \neg z), \}\}$ and either $out(G \setminus \{(\top, x)\}, A) = Cn(\{\neg z\})$ or $out(G \setminus \{(\top, x)\}, A) = Cn(\{\neg x, \neg z\})$ for all standard output operations $out_n^{(+)}$, and so $O^{\cap}\neg z$ is true given the norm and fact sets G and A, i.e. the man must not tell his neighbors he is coming.

5 Descriptive dyadic obligations

Dyadic deontic operators, that formalize e.g. 'x ought to be true under conditions a' as $O(x/a)$, were introduced over 50 years ago by G. H. von Wright [56]. Their introduction was due to Prior's paradox of derived obligation: often a primary obligation Ox is accompanied by a secondary, 'contrary-to-duty' obligation that pronounces y (a sanction, a remedy) as obligatory if the primary obligation is violated. At the time, the usual formalization of the secondary obligation would have been $O(\neg x \to y)$, but given Ox and the axioms of standard deontic logic SDL, $O(\neg x \to y)$ is derivable for any y. A bit later, Chisholm's paradox showed that formalizing the secondary obligation as $\neg x \to Oy$ produces similarly counterintuitive results. So to deal with such contrary-to-duty conditions, the dyadic deontic operator $O(x/a)$ was invented.

The perhaps best-known semantic characterization of dyadic deontic logic is Bengt Hansson's [24] system *DSLD*3, axiomatized by Spohn [52]. Hansson's idea was that the circumstances (the conditions a) are something which has actually happened (or will unalterably happen) and which cannot be changed afterwards. Ideal worlds in which $\neg a$ is true are therefore excluded. But some worlds may still be better than others, and there should then be an obligation to make "'the best out of the sad circumstances". Consequently, Hansson presents a possible worlds semantics in which all worlds are ordered by a preference (betterness) relation. $O(x/a)$ is then defined true if x is true in the best a-worlds. Here, we intend to employ semantics that do not make use of any prohairetic betterness relation, but that models deontic operators with regard to given sets of norms and facts, and the question is then

Problem 5. How to define dyadic deontic operators with regard to given sets of norms and facts?

Input/output logic assumes a set of (conditional) norms G, and a set of invariable facts A. The facts A may describe a situation that is inconsistent with the output $out(G, A)$: suppose there is a primary norm $(\top, a) \in G$ and a secondary norm $(\neg a, x) \in G$, i.e. $G = \{(\top, a), (\neg a, x)\}$, and $A = \{\neg a\}$. Though

$a \in out(G, A)$, it makes no sense to describe a as obligatory since a cannot be realized any more in the given situation – no crying over spilt milk. Rather, the output should include only the consequent of the secondary obligation x – it is the best we can make out of these circumstances. To do so, we return to the definitions of $maxfamily(G, A, A)$ as the set of all maximal subsets $H \subseteq G$ such that $out(H, A) \cup A$ is consistent, and the set $out^{\cap}(G, A)$ as the intersection of all outputs from $H \in maxfamily(G, A, A)$, i.e. $out^{\cap}(G, A) = \bigcap\{out(H, A) \mid H \in maxfamily(G, A, A)\}$. We may then define:

$$G \models O(x/a) \text{ iff } x \in out^{\cap}(G, \{a\})$$

Thus, relative to the set of norms G, $O(x/a)$ is defined true if x is in the output under a of all maximal sets H of norms such that their output under $\{a\}$ is consistent with a. In the example where $G = \{(\top, a), (\neg a, x)\}$ we therefore obtain $O(x/\neg a)$ but not $O(a/\neg a)$ as being true, i.e. only the consequent of the secondary obligation is described as obligatory in conditions $\neg a$.

In the above definition, the antecedent a of the dyadic formula $O(x/a)$ makes the inputs explicit: the truth definition does not make use of any facts other than a. This may be unwanted; one might consider an input set A of *given* facts, and employ the antecedent a only to denote an additional, *assumed* fact. Still, the output should contradict neither the given nor the assumed facts, and the output should include also the normative consequences x of a norm (a, x) given the assumed fact a. This may be realized by the following definition:

$$G, A \models O(x/a) \text{ iff } x \in out^{\cap}(G, A \cup \{a\})$$

So, relative to a set of norms G and a set of facts A, $O(x, a)$ is defined true if x is in the output under $A \cup \{a\}$ of all maximal sets H of norms such that their output under $A \cup \{a\}$ is consistent with $A \cup \{a\}$.

Hansson's description of dyadic deontic operators as describing defeasible obligations that are subject to change when more specific, namely contrary-to-duty situations emerge, may be the most prominent view, but it is by no means the only one. Earlier authors like von Wright [57] [58] and Anderson [4] have proposed more normal conditionals, which in particular support 'strengthening of the antecedent' SA $O(x/a) \rightarrow O(x/a \wedge b)$. From an input/output perspective, such operators can be accommodated by defining

$$G, A \models O(x/a) \text{ iff } x \in out(G, A \cup \{a\})$$

It is immediate that for all standard output operations $out_n^{(+)}$ this definition validates SA. The properties of dyadic deontic operators that are, like the above, semantically defined within the framework of input/output logic, have not been studied so far. The theorems they validate will inevitably depend on what output operation is chosen (cf. [23] for some related conjectures).

6 Permissive norms

In formal deontic logic, permission is studied less frequently than obligation. For a long time, it was naively assumed that it can simply be taken as a dual of obligation, just as possibility is the dual of necessity in modal logic. Permission is then defined as the absence of an obligation to the contrary, and the modal operator P defined by $Px =_{def} \neg O \neg x$. Today's focus on obligations is not only in stark contrast how deontic logic began, for when von Wright [55] started modern deontic logic in 1951, it was the P-operator that he took as primitive, and defined obligation as an absence of a permission to the contrary. Rather, more and more authors have come to realize how subtle and multi-faceted the concept of permission is. Much energy was devoted to solving the problem of 'free choice permission', where one may derive from the statement that one is permitted to have a cup of tea or a cup of coffee that it is permitted to have a cup of tea, and it is permitted to have a cup of coffee, or for short, that $P(x \vee y)$ implies Px and Py (cf. [31]. Von Wright, in his late work starting with [61], dropped the concept of inter-definability of obligations and permissions altogether by introducing P-norms and O-norms, where one may call something permitted only if it derives from the collective contents of some O-norms and at most one P-norm. This concept of 'strong permission' introduced deontic 'gaps': whereas in standard deontic logic SDL, $O\neg x \vee Px$ is a tautology, meaning that any state of affairs is either forbidden or permitted, von Wright's new theory means that in the absence of explicit P-norms only what is obligatory is permitted, and that nothing is permitted if also O-norms are missing. Perhaps most importantly, Bulygin [12] observed that an authoritative kind of permission must be used in the context of multiple authorities and updating normative systems: if a higher authority permits you to do something, a lower authority can no longer prohibit it. Summing up, the understanding of permission is still in a less satisfactory state than the understanding of obligation and prohibition. The problem can be phrased thus:

Problem 6. How to distinguish various kinds of permissions and relate them to obligations?

¿From the viewpoint of input/output logic, one may first try to define a concept

of negative permission in the line of the classic approach. Such a definition is the following:

$$G, A \models P^{neg}x \;\text{ iff }\; \neg x \notin out(G, A)$$

So something is permitted by a code iff its negation is not obligatory according to the code and in the given situation. As innocuous and standard as such a definition seems, questions arise as to what output operation *out* may be used. Simple-minded output out_1 and basic output out_2 produce counterintuitive results: consider a set of norms G of which one norm $(work, tax)$ demands that if I am employed then I have to pay tax. For the default situation $A = \{\top\}$ then $P^{neg}(a \wedge \neg x)$ is true, i.e. it is by default permitted that I am employed and do not pay tax. Stronger output operations out_3 and out_4 that warrant reusable output exclude this result, but their use in deontic reasoning is questionable for other reasons.

In contrast to a concept of negative permission, one may also define a concept of 'strong' or 'positive permission'. This requires a set P of explicit permissive norms, just as G is a set of explicit obligations. As a first approximation, one may say that something is positively permitted by a code iff the code explicitly presents it as such. But this leaves a central logical question unanswered as to how explicitly given permissive and obligating norms may generate permissions that – in some sense – follow from the explicitly given norms. In the line of von Wrights later approach, we may define:

$$G, P \models P^{stat}(x/a) \ \text{ iff } \ x \in out(G \cup \{(b, y)\}, a) \text{ for some } (b, y) \in P \cup \{(\top, \top)\}$$

So there is a permission to realize x in conditions a if x is generated under these conditions either by the norms in G alone, or the norms in G together with some explicit permission (b, y) in P. We call this a 'static' version of strong permission. For example, consider a set G consisting of the norm $(work, tax)$, and a set P consisting of the sole license $(18y, vote)$ that permits all adults to take part in political elections. Then all of the following are true: $P^{stat}(tax/work)$, $P^{stat}(vote/18y)$, $P^{stat}(tax/work \wedge male)$ and also $P^{stat}(vote/\neg work \wedge 18y)$ (so even unemployed adults are permitted to vote).

Where negative permission is liberal, in the sense that anything is permitted that does not conflict with ones obligations, the concept of static permission is quite strict, as nothing is permitted that does not explicitly occur in the norms. In between, one may define a concept of 'dynamic permission' that defines something as permitted in some situation a if forbidding it for these conditions would prevent an agent from making use of some explicit (static) permission. The formal definition reads:

$$G, P \models P^{dyn}(x/a) \ \text{ iff } \ \neg y \in out(G \cup \{(a, \neg x)\}, b) \text{ for some } y \text{ and conditions } b \text{ such that } G, P \models P^{stat}(y/b)$$

Consider the above static permission $P^{stat}(vote/\neg work \wedge 18y)$ that even the unemployed adult populations is permitted to vote, generated by the sets $P = \{(18y, vote)\}$ and $G = \{(work, tax)\}$. We might also like to say, without reference to age, that the unemployed are protected from being forbidden to vote, and in this sense are permitted to vote, but $P^{stat}(vote/\neg work)$ is not true. And we might like to say that adults are protected from being forbidden to vote unless they are employed, and in this sense are permitted to be both unemployed and take part in elections, but also $P^{stat}(\neg work \wedge vote/18y)$ is not true. Dynamic permissions allow us to express such protections, and make both $P^{dyn}(vote/\neg work)$ and $P^{dyn}(\neg work \wedge vote/18y)$ true: if either $(\neg work, \neg vote)$ or $(18y, (\neg work \rightarrow \neg vote))$ were added to G we would obtain $\neg vote$ as output in conditions $\neg work \wedge 18y)$ in spite of the fact that, as we have seen, $G, P \models P^{stat}(vote/\neg work \wedge 18y)$.

There are, ultimately, a number of questions for all these concepts of permissions that have been further explored in [44]. Other kinds of permissions have been discussed from an input/output perspective in the literature, too, for example permissions as exceptions of obligations [8]. But it seems input/output logic is able to help clarify the underlying concepts of permission better than traditional deontic semantics.

7 Meaning postulates and intermediate concepts

To define a deontic operator of individual obligation seems straightforward if the norm in question is an individual command or act of promising. For example, if you are the addressee α of the following imperative sentence

(1) You, hand me that screwdriver, please.

and you consider the command valid, then what you ought to do is to hand the screwdriver in question to the person β uttering the request. In terms of input/output logic, let x be the proposition that α hands the screwdriver to β: with the set of norms $G = \{(\top, x)\}$, the set of facts $A = \{\top\}$, and the truth definition Ox iff $x \in out(A, G)$: then we obtain that Ox is true, i.e. it is true that it ought to be that α hands the screwdriver to β.

Norms that belong to a legal system are more complex, and thus more difficult to reason about. Consider, for example

(2) An act of theft is punished by a prison sentence not exceeding 5 years or a fine.

Things are again easy if you are a judge and you know that the accused in front of you has committed an act of theft – then you ought to hand out a verdict that commits the accused to pay a fine or to serve a prison sentence not exceeding 5 years. But how does the judge arrive at the conclusion that an act of theft has been committed? 'Theft' is a legal term that is usually accompanied by a legal definition such as the following one:

(3) Someone commits an act of theft if that person has taken a movable object from the possession of another person into his own possession with the intention to own it, and if the act occurred without the consent of the other person or some other legal authorization.

It is noteworthy that (3) is not a norm in the strict sense – it does not prescribe or allow a behavior – but rather a stipulative definition, or, in more general terms, a *meaning postulate* that constitutes the legal meaning of theft. Such sentences are often part of the legal code. They share with norms the property of being neither true nor false. The significance of (3) is that it decomposes the complex legal term 'theft' into more basic legal concepts. These concepts are again the subject of further meaning postulates, among which may be the following:

(4) A person in the sense of the law is a human being that has been born.
(5) A movable object is any physical object that is not a person or a piece of land.
(6) A movable object is in the possession of a person if that person is able to control the uses and the location of the object.
(7) The owner of an object is – within the limits of the law – entitled to do with it whatever he wants, namely keep it, use it, transfer possession or ownership of the object to another person, and destroy or abandon it.

Not all of definitions (4)-(7) may be found in the legal statutes, though they may be viewed as belonging to the normative system by virtue of having been accepted in legal theory and judicial reasoning. They constitute 'intermediate concepts': they link legal terms (person, movable object, possession etc.) to words describing natural facts (human being, born, piece of land, keep an object etc.).

Any proper representation of legal norms must include means of representing meaning postulates that define legal terms, decompose legal terms into more basic legal terms, or serve as intermediate concepts that link legal terms to terms that describe natural facts. But for deontic logic, with its standard possible worlds semantics, a comprehensive solution to the problem of representing meaning postulates is so far lacking (cf. Lindahl [39]). The problem is thus:

Problem 7. How can meaning postulates and intermediate terms be modeled in semantics for deontic logic reasoning?

The representation of intermediate concepts is of particular interest, since such concepts arguably reduce the number of implications required for the transition from natural facts to legal consequences and thus serve an economy of expression (cf. Lindahl & Odelstad [40]). Lindahl & Odelstad use the term 'ownership' as an example to argue as follows: let $F_1, ..., F_p$ be descriptions of some situations in which a person α acquires ownership of an object γ, e.g. by acquiring it from some other person β, finding it, building it from owned materials, etc., and let $C_1, ..., C_n$ be among the legal consequences of α's ownership of γ, e.g. freedom to use the object, rights to compensation when the object is damaged, obligations to maintain the object or pay taxes for it etc. To express that each fact F_i has the consequence C_j, $p \times n$ implications are required. The introduction of the term $\mathit{Ownership}(x, y)$ reduces the number of required implications to $p + n$: there are p implications that link the facts $F_1, ..., F_p$ to the legal term $\mathit{Ownership}(x, y)$, and n implications that link the legal term $\mathit{Ownership}(x, y)$ to each of the legal consequences $C_1, ..., C_n$. The argument obviously does not apply to all cases: one implication $(F_1 \vee ... \vee F_p) \rightarrow (C_1 \wedge ... \wedge C_n)$ may often be sufficient to represent the case that a variety of facts $F_1, ..., F_p$ has the same multitude of legal consequences $C_1, ..., C_n$. However, things may be different when norms that link a number of factual descriptions to the same legal consequences stem from different normative sources, may come into conflict with other norms, can be overridden by norms of higher priority, or be subject of individual exemption by norms that grant freedoms or licenses: in these cases, the norms must be represented individually. So it seems worthwhile to consider ways to incorporate intermediate concepts into a formal semantics for deontic logic.

In an input/output framework, a first step could be to employ a separate set T of theoretical terms, namely meaning postulates, alongside the set G of norms. Let T consists of intermediates of the form (a, x), where a is a factual sentence (e.g. that β is in possession of γ, and that α and β agreed that α should have γ, and that β hands γ to α), and x states that some legal term obtains (e.g. that α is now owner of γ). To derive outputs from the set of norms G, one may then use $A \cup out(T, A)$ as input, i.e. the factual descriptions together with the legal statements that obtain given the intermediates T and the facts A.

It may be of particular interest to see that such a set of intermediates may help resolve possible conflicts in the law. Let $(\top, \neg \mathit{dog})$ be a statute that forbids dogs on the premises, but let there also be a higher order principle that no blind person may be required to give up his or her guide dog. Of course the conflict may be solved by modifying the statute (e.g. add a condition that the dog in question is not a guide dog), but then modifying a statute is usually not something a judge, faced with such a norm, is allowed to do: the judge's duty is solely to consider the statute, interpret it according to the known or supposed will of the norm-giver, and apply it to the given facts. The judge may then come to the conclusion that a fair and considerate norm-giver would not have meant the statute to apply to guide dogs, i.e. the term "dog" in the statute is a theoretical term whose extension is smaller than the natural term. So the statute must be re-interpreted as reading $(\top, \neg tdog)$ with the additional intermediate $(dog \wedge \neg guidedog, tdog) \in T$, and thus no conflict arises for the case of blind persons that want to keep their guide dog. While this seems to be a rather natural view of how judicial conflict resolution works (the example is taken from an actual court case), the exact process of creating and modifying theoretical terms in order to resolve conflicts must be left to further study.

8 Constitutive norms

Constitutive norms like counts-as conditionals are rules that create the possibility of or define an activity. For example, according to Searle [50], the activity of playing chess is constituted by action in accordance with these rules. Chess has no existence apart from these rules. The institutions of marriage, money, and promising are like the institutions of baseball and chess in that they are systems of such constitutive rules or conventions. They have been identified as the key mechanism to normative reasoning in dynamic and uncertain environments, for example to realize agent communication, electronic contracting, dynamics of organizations, see, e.g., [9].

Problem 8. How to define counts-as conditionals and relate them to obligations and permissions?

For Jones and Sergot [29], the counts-as relation expresses the fact that a state of affairs or an action of an agent "is a sufficient condition to guarantee that the institution creates some (usually normative) state of affairs". They formalize this introducing a conditional connective $\Rightarrow_s$ to express the "counts-as" connection holding in the context of an institution s. They characterize the logic of $\Rightarrow_s$ as a conditional logic, with axioms for agglomeration $((x \Rightarrow_s y) \,\&\, (x \Rightarrow_s z)) \supset (x \Rightarrow_s (y \wedge z))$, left disjunction $((x \Rightarrow_s z) \,\&\, (y \Rightarrow_s z)) \supset ((x \vee y) \Rightarrow_s z)$ and transitivity $((x \Rightarrow_s y) \,\&\, (y \Rightarrow_s z)) \supset (x \Rightarrow_s z)$. The flat fragment can be phrased as an input/output logic as follows [7].

Definition 1. *Let L be a propositional action logic with $\vdash$ the related notion of derivability and Cn the related consequence operation $Cn(x) = \{y \mid x \vdash y\}$. Let CA be a set of pairs of L, $\{(x_1, y_1), \ldots, (x_n, y_n)\}$, read as '$x_1$ counts as y_1', etc.*

Moreover, consider the following proof rules conjunction for the output (AND), disjunction of the input (OR), and transitivity (T) defined as follows:

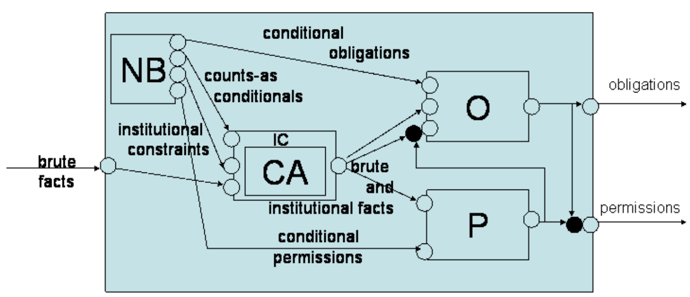
$$\frac{(x,y_1),(x,y_2)}{(x,y_1\wedge y_2)}AND \qquad \frac{(x_1,y),(x_2,y)}{(x_1\vee x_2,y)}OR \qquad \frac{(x,y_1),(y_1,y_2)}{(x,y_2)}T$$

For an institution s, the counts-as output operator out_{CA} is defined as closure operator on the set CA using the rules above, together with a silent rule that allows replacement of logical equivalents in input and output. We write $(x,y)\in out_{\mathrm{CA}}(CA,s)$. Moreover, for $X\subseteq L$, we write $y\in out_{\mathrm{CA}}(CA,s,X)$ if there is a finite $X'\subseteq X$ such that $(\wedge X',y)\in out_{\mathrm{CA}}(CA,s)$, indicating that the output y is derived by the output operator for the input X, given the counts-as conditionals CA of institution s. We also write $out_{\mathrm{CA}}(CA,s,x)$ for $out_{\mathrm{CA}}(CA,s,\{x\})$.

Example 1. If for some institution s we have $CA=\{(a,x),(x,y)\}$, then we have $out_{CA}(CA,s,a)=\{x,y\}$.

There is presently no consensus on the logic of counts-as conditionals, probably due to the fact that the concept is not studied in depth yet. For example, the adoption of the transitivity rule T for their logic is criticized by Artosi *et al.* [5]. Jones and Sergot say that "we have been unable to produce any counter-instances [of transitivity], and we are inclined to accept it".[7]

The main issue in defining constitutive norms like counts-as conditionals is defining their relation with regulative norms like obligations and permissions. Boella and van der Torre [7] use the notion of a logical architecture combining several logics into a more complex logical system, also called logical input/output nets (or lions).



The notion of logical architecture naturally extends the input/output logic framework, since each input/output logic can be seen as the description of a 'black box'. In the above figure there are boxes for counts-as conditionals (CA), institutional constraints (IC), obligating norms (O) and explicit permissions (P). The norm base (NB) component contains sets of norms or rules, which are used in the other components to generate the component's output from its input. The

[7] Neither of these authors considers replacing transitivity by cumulative transitivity (CT): $((x\Rightarrow_s y)\,\&\,(x\wedge y\Rightarrow_s z))\supset(x\Rightarrow_s z)$, that characterizes operations out_3, out_4 of input/output logic.

figure shows that the counts-as conditionals are combined with the obligations and permissions using iteration, that is, the counts-as conditionals produce institutional facts, which are input for the norms. Roughly, if we write $out(CA, G, A)$ for the output of counts-as conditionals together with obligations, $out(G, A)$ for obligations as before, then $out(CA, G, A) = out(G, out_{CA}(CA, A))$.

There are many open issues concerning constitutive norms, since their logical analysis has not attracted much attention yet. How to distinguish among various kinds of constitutive norms? How are constitutive norms (x counts as y) distinguished from classifications (x is a y)? What is the relation with intermediate concepts?

9 Revision of a set of norms

In general, a code G of regulations is not static, but changes over time. For example, a legislative body may want to introduce new norms or to eliminate some existing ones. A different (but related) type of change is the one induced by the fusion of two (or more) codes as it is addressed in the next section.

Little work exists on the logic of the revision of a set of norms. To the best of our knowledge, Alchourrón and Makinson were the first to study the changes of a legal code [2,3]. The addition of a new norm n causes an enlargement of the code, consisting of the new norm plus all the regulations that can be derived from n. Alchourrón and Makinson distinguish two other types of change. When the new norm is incoherent with the existing ones, we have an *amendment* of the code: in order to coherently add the new regulation, we need to reject those norms that conflict with n. Finally, *derogation* is the elimination of a norm n together with whatever part of G implies n.

In [2] a "hierarchy of regulations" is assumed. Few years earlier, Alchourrón and Bulygin [1] already considered the *Normenordnung* and the consequences of gaps in this ordering. For example, in jurisprudence the existence of precedents is an established method to determine the ordering among norms.

However, although Alchourrón and Makinson aim at defining change operators for a set of norms of some legal system, the only condition they impose on G is that it is a non-empty and finite set of propositions. In other words, a norm x is taken to be simply a formula in propositional logic. Thus, they suggest that "the same concepts and techniques may be taken up in other areas, wherever problems akin to inconsistency and derogation arise" ([2], p. 147).

This explains how their work (together with Gärdenfors' analysis of counterfactuals) could ground that research area that is now known as *belief revision*. Belief revision is the formal studies of how a set of propositions changes in view of a new information that may cause an inconsistency with the existing beliefs. Expansion, revision and contraction are the three belief change operations that Alchourrón, Gärdenfors and Makinson identified in their approach (called AGM) and that have a clear correspondence with the changes on a system of norms we mentioned above. Hence, the following question needs to be addressed:

Problem 9. How to revise a set of regulations or obligations? Does belief revision offer a satisfactory framework for norms revision?

Some of the AGM axioms seem to be rational requirements in a legal context, whereas they have been criticized when imposed on belief change operators. An example is the *success* postulate, requiring that a new input must always be accepted in the belief set. It is reasonable to impose such a requirement when we wish to enforce a new norm or obligation. However, it gives rise to irrational behaviors when imposed to a belief set, as observed for instance in [16].

On the other hand, when we turn to a proper representation of norms, like in the input/output logic framework, the AGM principles prove to be too general to deal with the revision of a normative system. For example, one difference between revising a set of propositions and revising a set of regulations is the following: when a new norm is added, coherence may be restored modifying some of the existing norms, not necessarily retracting some of them. The following example will clarify this point:

Example. If we have $\{(\top, a), (a, b)\}$ and we have that c is an exception to the obligation to do b, then we need to retract (c, b). Two possible solutions are $\{(\neg c, a), (a, b)\}$ or $\{(\top, a), (a \wedge \neg c, b)\}$.

Future research must investigate whether general patterns in the revision of norms exist and how to formalize them.

10 Merging sets of norms

In the previous section we have seen that the change over time of a system of norms raises questions that cannot be properly answered within the belief revision framework. We now want to turn to another type of change, that is the aggregation of regulations. This problem has been only recently addressed in the literature and therefore the findings are still very partial.

The first noticeable thing is the lack of general agreement about where the norms that are to be aggregated come from:

1. some works focus on the merging of conflicting norms that belong to the same normative system [14];
2. other works assume that the regulations to be fused belong to different systems [11]; and finally
3. some authors provide patterns of possible rules to be combined, and consider both cases (1) and (2) above [18].

The first situation seems to be more a matter of coherence of the whole system rather than a genuine problem of fusion of norms. However, such approaches have the merit to reveal the tight connections between fusion of norms, non-monotonic logics and defeasible deontic reasoning. The initial motivation for the study of belief revision was the ambition to model the revision of a set of regulations. On the contrary, the generalization of belief revision to *belief merging* is exclusively dictated by the goal to tackle the problem — arising in computer science — of combining information from different sources. The pieces of information are represented in a formal language and the aim is to merge them in an (ideally) unique knowledge base.[8]

Problem 10. Can the belief merging framework deal with the problem of merging sets of norms?

If (following Alchourrón and Makinson) we assume that norms are unconditional, then we could expect to use standard merging operators to fuse sets of norms. Yet, not only once we consider conditional norms, as in the input/output logic framework, problems arise again. But also, most of the fusion procedures proposed in the literature seem to be inadequate for the scope.

To see why this is the case, we need to explain the merging approach in few words. Let us assume that we have a finite number of belief bases $K_1, K_2, \ldots, K_n$ to merge. IC is the belief base whose elements are the integrity constraints (i.e., any condition that we want the final outcome to satisfy). Given a multi-set $E = \{K_1, K_2, \ldots, K_n\}$ and IC, a merging operator $\mathcal{F}$ is a function that assigns a belief base to E and IC. Let $\mathcal{F}_{IC}(E)$ be the resulting collective base from the IC fusion on E.

Fusion operators come in two types: model-based and syntax-based. The idea of a model-based fusion operator is that models of $\mathcal{F}_{IC}(E)$ are models of IC, which are preferred according to some criterion depending on E. Usually the preference information takes the form of a total pre-order on the interpretations induced by a notion of distance $d(w, E)$ between an interpretation w and E.

Syntax-based merging operators are usually based on the selection of some consistent subsets of E [6,34]. The bases K_i in E can be inconsistent and the result does not depend on the distribution of the wffs over the members of the group.[9]

Finally, the model-based aggregation operators for bases of equally reliable sources can be of two sorts. On the one hand, there are majoritarian operators that are based on a principle of distance-minimization [38]. On the other hand, there are egalitarian operators, which look at the distribution of the distances in E [33]. These two types of merging try to capture two intuitions that often guide the aggregation of individual preferences into a social one. One option is to let the majority decide the collective outcome, and the other possibility is to equally distribute the individual dissatisfaction.

Obviously, these intuitions may well serve in the aggregation of individual knowledge bases or individual preferences, but have nothing to say when we try to model the fusion of sets of norms. Hence, for this purpose, syntactic merging operators may be more appealing. Nevertheless, the selection of a coherent subset

[8] See [35] for a survey on logic-based approaches to information fusion.

[9] [34] refers the term ‘combination’ to the syntax-based fusion operators to distinguish them from the model-based approaches.

depends on additional information like an order of priority over the norms to be merged, or some other meta-principles.

As the application of belief merging to the aggregation of sets of norms turned out to be unfeasible, an alternative approach is to generalize existing belief change operators to merging rules. This is the approach followed in [11], where merging operators defined using a consolidation operation and possibilistic logic are applied to the aggregation of conditional norms in an input/output logic framework. However, at this preliminary stage, it is not clear whether such methodology is more fruitful for testing the flexibility of existing operators to tackle other problems than the ones they were created for, or if this approach can really shed some light to the new riddle at hand.

A different perspective is taken in [18]. Here, real examples from the Belgian-French bilateral agreement preventing double taxation are considered. These are fitted into a taxonomy of the most common legal rules with exceptions, and the combination of each pair of norms is analyzed. Moreover, both the situations in which the regulations come from the same system and those in which they come from different ones are contemplated, and some general principles are derived. Finally, a merging operator for rules with abnormality propositions is proposed. A limit of Grégoire's proposal is that only the aggregation of rules with the same consequence is taken into account and, in our opinion, this neglects other sorts of conflicts that may arise, as we see now.

The call for non-monotonic reasoning in the treatment of contradictions is also in Cholvy and Cuppens' [14]. A logic to reason when several contradictory norms are merged is presented. The proposal assumes an order of priority among the norms to be merged and this order is also the way to solve the incoherence. Even though this is quite a strong assumption, Cholvy and Cuppens' work take into consideration a broader type of incoherence than in [18]. In their example, an organization that works with secret documents has two rules. R_1 is "It is obligatory that any document containing some secret information is kept in a safe, when nobody is using this document". R_2 is "If nobody has used a given document for five years, then it is obligatory to destroy this document by burning it". As they observe, in order to deduce that the two rules are conflicting, we need to introduce the constraint that keeping a document and destroying it are contradictory actions. That is, the notion of coherence between norms can involve information that are not norms.

11 Conclusion: Deontic logic in context

In this paper we discussed problems of deontic logic that should be considered open and how input/output logic may be useful for analyzing these problems and finding fresh solutions. Jørgensen's dilemma might be overcome by distinguishing operations with norms, like the output $out(G, A)$ of a set of norms G under conditions A, from truth definitions that define what ought to obtain or be done given these norms and conditions. Coherence of a set of norms might be defined with respect to output under constraints, meaning that the set of norms should not generate output for certain conditions that is inconsistent with these constraints. Normative conflicts may be overcome by considering coherent subsets of norms and their output, or such subsets that are preferred given a priority ordering of the norms. Likewise, contrary-to-duty obligations, that obtain in conditions that represent violations, may be modeled by considering only output that is consistent with the input, i.e. the given conditions. Input/output logic provides two possible definitions of dyadic deontic operators, which reconstruct past discussions on whether such operators should be defeasible (in particular in contrary-to-duty conditions), or support strengthening of the antecedent that derives $O(x/a \wedge b)$ from $O(x/a)$. Input/output logic may take into account not just sets of obligating norms, but also explicit permissions, and thus helps shed light on the distinction between weak (negative) permission, where something is permitted if it does not conflict with the norms, and strong (positive) permission which requires an explicit license by the norm-givers. Meaning postulates and intermediate terms, common in legal reasoning but largely ignored by traditional deontic literature, can be taken into account by considering generators T that link natural facts to theoretical terms occurring in the norms, and for counts-as conditionals we may use a separate set of generators (normative institutions) that models how norms are created given an input of natural facts. Finally the questions of how to revise and merge given sets of norms may be approached by preparing the generators (norms) with the aid of standard revision and merging operators.

Lately, normative systems and deontic logic have received widespread attention in multiagent systems and artificial intelligence. A normative multiagent system is "a multiagent system together with normative systems in which agents can decide whether to follow the explicitly represented norms, and the normative systems specify how and in which extent the agents can modify the norms" [10]. Deontic logic, that attempts to formalize the normative consequences given a set of norms and a given situation, can be a helpful tool for devising such systems. In such a general setting, a setting of 'deontic logic in context', many new problems arise: how do deontic truths feature in agent planning and decision making? how do they interact with agent desires, goals, preferences and intentions? how do they feature in communication? how do we model the change of obligations over time, when agents violate or discharge their obligations, when the underlying norms are modified or retracted or when new norms come into existence? The clarification and solution of the problems outlined above, and others, may serve as a first step to make deontic logic fit to become a working component in such a larger setting.

Acknowledgments

The authors wish to thank David Makinson for helpful discussions in the preparation of this paper.

References

1. Alchourrón, C. E. and Bulygin, E., "The Expressive Conception of Norms", in [27] 95–124.
2. Alchourrón, C. E. and Makinson, D., "Hierarchies of Regulations and Their Logic", in [27] 125–148.
3. Alchourrón, C. E. and Makinson, D., "On the Logic of Theory Change: Contraction Functions and Their Associated Revision Functions", *Theoria*, **48**, 1982, 14–37.
4. Anderson, A. R., "On the Logic of Commitment", *Philosophical Studies*, **19**, 1959, 23–27.
5. Artosi, A., Rotolo, A. and Vida, S., "On the logical nature of count-as conditionals", in: *Procs. of LEA 2004 Workshop*, 2004.
6. Baral, C., Kraus, S., Minker, J. and Subrahmanian, V. S., "Combining knowledge bases consisting of first-order theories", *Computational Intelligence*, **8**, 1992, 45–71.
7. Boella, G. and van der Torre, L., "A Logical Architecture of a Normative System", in [17] 24–35.
8. Boella, G. and van der Torre, L., "Permissions and Obligations in Hierarchical Normative Systems", in: *Proceedings of the 9th International Conference on Artificial Intelligence and Law, ICAIL 2003, June 24-28, Edinburgh, Scotland, UK*, ACM, 2003, revised version to appear in *Artificial Intelligence and Law*.
9. Boella, G. and van der Torre, L., "Constitutive Norms in the Design of Normative Multiagent Systems", in: *Computational Logic in Multi-Agent Systems, 6th International Workshop, CLIMA VI*, LNCS 3900, Springer, 2006, 303–319.
10. Boella, G., van der Torre, L. and Verhagen, H., "Introduction to normative multiagent systems", *Computational and Mathematical Organization Theory*, **12(2-3)**, 2006, 71–79.
11. Booth, R., Kaci, S. and van der Torre, L., "Merging Rules: Preliminary Version", in: *Proceedings of the NMR'06*, 2006.
12. Bulygin, E., "Permissive Norms and Normative Concepts", in: Martino, A. A. and Socci Natali, F. (eds.), *Automated Analysis of Legal Texts*, Amsterdam: North Holland, 1986, 211–218.
13. Chisholm, R., "Contrary-to-duty imperatives and deontic logic", *Analysis*, **24**, 1963, 3336.
14. Cholvy, L. and Cuppens, F., "Reasoning about Norms Provided by Conflicting Regulations", in [46] 247–264.
15. van Fraassen, B., "Values and the Heart's Command", *Journal of Philosophy*, **70**, 1973, 5–19.
16. Gabbay, D., Pigozzi, G. and Woods, J., "Controlled Revision — An algorithmic approach for belief revision", *Journal of Logic and Computation*, **13**, 2003, 3–22.
17. Goble, L. and Meyer, J.-J. C. (eds.), *Deontic Logic and Artificial Normative Systems. 8th International Workshop on Deontic Logic in Computer Scicence, DEON 2006, Utrecht, July 2006, Proceedings*, Berlin: Springer, 2006.
18. Grégoire, E., "Fusing legal knowledge", in: *Proceedings of the 2004 IEEE INt. Conf. on Information Reuse and Integration (IEEE-IRI'2004)*, 2004, 522–529.
19. Hansen, J., "Sets, Sentences, and Some Logics about Imperatives", *Fundamenta Informaticae*, **48**, 2001, 205–226.
20. Hansen, J., "Problems and Results for Logics about Imperatives", *Journal of Applied Logic*, **2**, 2004, 39–61.
21. Hansen, J., "Conflicting Imperatives and Dyadic Deontic Logic", *Journal of Applied Logic*, **3**, 2005, 484–511.

22. Hansen, J., "Deontic Logics for Prioritized Imperatives", *Artificial Intelligence and Law*, *forthcoming*, 2005.
23. Hansen, J., "Prioritized Conditional Imperatives: Problems and a New Proposal", *Autonomous Agents and Multi-Agent Systems*, 2007, submitted.
24. Hansson, B., "An Analysis of Some Deontic Logics", *Nôus*, **3**, 1969, 373–398, reprinted in [26] 121–147.
25. Hare, R. M., *Moral Thinking*, Oxford: Clarendon Press, 1981.
26. Hilpinen, R. (ed.), *Deontic Logic: Introductory and Systematic Readings*, Dordrecht: Reidel, 1971.
27. Hilpinen, R. (ed.), *New Studies in Deontic Logic*, Dordrecht: Reidel, 1981.
28. Horty, J. F., "Nonmonotonic Foundations for Deontic Logic", in: Nute, D. (ed.), *Defeasible Deontic Logic*, Dordrecht: Kluwer, 1997, 17–44.
29. Jones, A. and Sergot, M., "A formal characterisation of institutionalised power", *Journal of IGPL*, **3**, 1996, 427443.
30. Jørgensen, J., "Imperatives and Logic", *Erkenntnis*, **7**, 1938, 288–296.
31. Kamp, H., "Free Choice Permission", *Proceedings of the Aristotelian Society*, **74**, 1973/74, 57–74.
32. Kanger, S., "New Foundations for Ethical Theory: Part 1", duplic., 42 p., 1957, reprinted in [26] 36–58.
33. Konieczny, S., *Sur la Logique du Changement: Révision et Fusion de Bases de Connaissance*, Ph.D. thesis, University of Lille, France, 1999.
34. Konieczny, S., "On the difference between merging knowledge bases and combinig them", in: *Proceedings of KR'00*, vol. 8, Morgan Kaufmann, 2000, 135–144.
35. Konieczny, S. and Grégoire, E., "Logic-based approaches to information fusion", *Information Fusion*, **7**, 2006, 4–18.
36. Lewis, D., *Counterfactuals*, Oxford: Basil Blackwell, 1973.
37. Lewis, D., "Semantic Analyses for Dyadic Deontic Logic", in: Stenlund, S. (ed.), *Logical Theory and Semantic Analysis*, Dordrecht: Reidel, 1974, 1 – 14.
38. Lin, J. and Mendelzon, A., "Merging databases under constraints", *International Journal of Cooperative Information Systems*, **7**, 1996, 55–76.
39. Lindahl, L., "Norms, Meaning Postulates, and Legal Predicates", in: Garzón Valdés, E. (ed.), *Normative Systems in Legal and Moral Theory. Festschrift for Carlos E. Alchourrón and Eugenio Bulygin*, Berlin: Duncker & Humblot, 1997, 293–307.
40. Lindahl, L. and Odelstad, J., "Intermediate Concepts in Normative Systems", in [17] 187–200.
41. Makinson, D., "On a Fundamental Problem of Deontic Logic", in [46], 29–53.
42. Makinson, D. and van der Torre, L., "Input/Output Logics", *Journal of Philosophical Logic*, **29**, 2000, 383–408.
43. Makinson, D. and van der Torre, L., "Constraints for Input/Output Logics", *Journal of Philosophical Logic*, **30**, 2001, 155–185.
44. Makinson, D. and van der Torre, L., "Permissions from an Input/Output Perspective", *Journal of Philosophical Logic*, **32**, 2003, 391–416.
45. Makinson, D. and van der Torre, L., "What is Input/Output Logic", in: Löwe, B., Malzkom, W. and Räsch, T. (eds.), *Foundations of the Formal Sciences II : Applications of Mathematical Logic in Philosophy and Linguistics (Papers of a conference held in Bonn, November 10-13, 2000)*, Trends in Logic, vol. 17, Dordrecht: Kluwer, 2003, 163–174, reprinted in this volume.
46. McNamara, P. and Prakken, H. (eds.), *Norms, Logics and Information Systems*, Amsterdam: IOS, 1999.

47. Niiniluoto, I., "Hypothetical Imperatives and Conditional Obligation", *Synthese*, **66**, 1986, 111–133.
48. Ross, A., "Imperatives and Logic", *Theoria*, **7**, 1941, 53–71, Reprinted in *Philosophy of Science* **11**:30–46, 1944.
49. Ross, W. D., *The Right and the Good*, Oxford: Clarendon Press, 1930.
50. Searle, J., *Speech Acts: an Essay in the Philosophy of Language*, Cambridge (UK): Cambridge University Press, 1969.
51. Smiley, T. J., "The Logical Basis of Ethics", *Acta Philosophica Fennica*, **16**, 1963, 237–246.
52. Spohn, W., "An Analysis of Hansson's Dyadic Deontic Logic", *Journal of Philosophical Logic*, **4**, 1975, 237–252.
53. Stenius, E., "The Principles of a Logic of Normative Systems", *Acta Philosophica Fennica*, **16**, 1963, 247–260.
54. von Wright, G., *Logical Studies*, London: Routledge and Kegan, 1957.
55. von Wright, G. H., "Deontic Logic", *Mind*, **60**, 1951, 1–15.
56. von Wright, G. H., "A Note on Deontic Logic and Derived Obligation", *Mind*, **65**, 1956, 507–509.
57. von Wright, G. H., "A New System of Deontic Logic", *Danish Yearbook of Philosophy*, **1**, 1961, 173–182, reprinted in [26] 105–115.
58. von Wright, G. H., "A Correction to a New System of Deontic Logic", *Danish Yearbook of Philosophy*, **2**, 1962, 103–107, reprinted in [26] 115–119.
59. von Wright, G. H., *Norm and Action*, London: Routledge & Kegan Paul, 1963.
60. von Wright, G. H., *An Essay in Deontic Logic and the General Theory of Action*, Amsterdam: North Holland, 1968.
61. von Wright, G. H., "Norms, Truth and Logic", in: von Wright, G. H. (ed.), *Practical Reason: Philosophical Papers vol. I*, Oxford: Blackwell, 1983, 130–209.
62. Ziemba, Z., "Deontic Syllogistics", *Studia Logica*, **28**, 1971, 139–159.