Norms, Logics and Information Systems. New Studies in Deontic Logic and Computer Science, edited by Paul McNamara and Henry Prakken (Amsterdam: IOS Press, Series: Frontiers in Artificial Intelligence and Applications, Volume: 49, 1999, ISBN 9051994273) pp 29-53.

On a Fundamental Problem of Deontic Logic

David Makinson

Les Etangs B2, La Ronce, 92410 Ville d'Avray, France
Email: d.makinson@unesco.org

Abstract

The usual presentations of deontic logic, whether axiomatic or semantic, treat norms as if they could bear truth-values. A fundamental problem of deontic logic, we believe, is to reconstruct it in accord with the philosophical position that norms direct rather than describe, and are neither true nor false.

(Alchourrón and Bulygin 1981) have indeed made such a construction, refining an earlier one of (Stenius 1963), based on the distinction between a norm and a proposition about norms. However it has the limitation that it does not deal with conditional norms. These are covered by an extension of (Alchourrón 1993), but with certain shortcomings. Our purpose is to extend the basic 1981 construction in another manner which, we suggest, provides a more satisfactory and sensitive analysis of conditional norms within the same philosophical perspective.

The approach takes seriously the warning: no logic of norms without attention to a system of which they form part. It is based on the notion of the *iterative development of output* of an explicitly presented normative code, under a given condition. It is neither axiomatic in style nor formulated in terms of a semantics of "possible worlds". It develops output by repeated detachment rather than by consequence (so as not to lose the directionality of conditional norms), and it distinguishes between gross and net output (so as to deal adequately with conditions that are "contrary-to-duty").

The investigation also provides new perspectives on some well known problems. In particular, it throws light on the way in which explicit obligations may have a part in generating permissions, and how explicit permissions may limit obligations. It also helps pin down a distinction between "substantive" and "technical" defeasibility of conditional norms.

1. Introduction: the problem

There is a singular tension between the philosophy of norms and the formal work of deontic logicians.

On the philosophical level, it is widely accepted that a distinction may be drawn between norms on the one hand, and declarative statements or propositions on the other. Declarative statements may bear truth-values, i.e. are capable of being true or false, whilst norms are items of another kind. They assign obligations, permissions, and prohibitions. They may be applied or not, respected or not, and may also be subject to evaluation from the standpoint of other norms, as for example when a legal norm is judged from the point of view of a moral code. But it makes no sense to describe norms as true or as false.

Of course, one may say with truth that such and such a norm is (or is not) part of (or implied by, consistent with, etc.) such and such a normative code. For example one may say that the British driving code requires vehicles, in normal circumstances, to be driven on the left hand side of public thoroughfares, while the American code requires otherwise. But such statements are not themselves norms. They report on the presence or absence (or more complex status) of a norm in a given normative system. To mark the difference, they are sometimes called "normative propositions" or "propositions about norms".

On the formal level, however, work goes on as if such a distinction had never been heard of. In axiomatic presentations of systems of deontic propositional logic, the truth-functional connectives "and", "or" and most spectacularly "not" are routinely applied to items construed as norms, forming compound norms out of elementary ones. But as (Dubislav 1937) and (Jørgensen 1937-8) already observed, if norms lack truth-values then it is not clear what could be meant by such compounding. For example, the negation of a declarative statement is understood to be true iff the item negated is false, and false iff the latter is true; but we cannot meaningfully formulate such a rule for norms. In semantic presentations, models are constructed that blithely assign truth-values to norms in possible worlds, and define validity in a model as truth in all its worlds.

There are a few deontic logicians such as (Kalinowski 1972) who take the view that norms *do*, after all, bear truth-values. Such a position has at least the merits of forthrightness and consistency – provided it is applied to *all* norms, great and small, moral, legal or aesthetic, personal or communal, that are taken to fall within the scope of deontic logic. But most deontic logicians accept that there is a fundamental distinction to be made, and find themselves in the uncomfortable position – indeed the unprincipled and close to inconsistent one – of conceding that norms do not bear truth-values, but hoping that for the purposes of logic they may, for some mysterious reason, be treated as if they did.

It is thus a central problem – we would say, a fundamental problem – of deontic logic to reconstruct it in accord with the philosophical position that norms are devoid of truth-values. In other words: to explain how deontic logic is possible on a positivistic philosophy of norms.

A natural first reaction is to regard the problem as trivial. On the syntactic level, all we need to do is re-read deontic formulae as propositions about norms, rather than as being themselves norms. As suggested for example by (Hansson 1969, section 3), we need only read a formula $O\alpha$ or $P\alpha$ as saying that α is obligatory, or permitted, according to some fixed system N of norms. On the semantic level, it is enough to take one's preferred possible-worlds semantics and reinterpret its deontic components as relativised to a given normative system N. In particular, we might interpret the "betterness" relation between possible worlds by putting one world better than another iff it violates only a subset of the explicit obligations of the system N, that are violated by the other.

Such a reaction is perfectly in order – so far as it goes. But it waves a hand in the direction of a solution, rather than supply one.

On the syntactic level, it remains to provide a precise formulation of the new reading of deontic formulae as reporting the norms that hold "according to" a particular code. The problem is to identify the norms that are implicit in a code on the basis of those that it presents explicitly, without appealing to some already given deontic logic. For unconditional obligation this is relatively easy to do, although it is also easy to go wrong even at that level as we shall see in section 2.1 below. But for conditional norms it is quite elusive. In effect, the present paper attempts to carry out that task.

On the semantic level, the problem is quite different. The definition of a "betterness" relation between possible worlds suggested above, in terms of the set of explicit promulgations of the code that are violated in each world, is perfectly precise; and such a relation permits a semantic account of conditional obligation.

However, such a construction faces at least three difficulties. One of these is validation of formulae $O(\beta/\beta)$ whose intuitive standing is open to question.[1] Another is that it is rather difficult to construct an account of what is known as strong or positive permission along the same lines; the only work in this area known to the author is that of (van der Torre 1997, section 2.4.1). Positive permission appears to be needed by real-life normative systems that change over time, as a device for limiting the interpretation of obligations and preventing their proliferation. In particular, as pointed out by (Alchourrón and Bulygin 1984), in legal contexts it is needed to limit the authority of subordinate instances to create new norms. Finally, the above semantic construction gives an unintuitive result in examples such as the following, whose name is suggested by its shape.

> *Example 1 (Möbius strip).* Consider a code with just three explicit norms, stating that α is obligatory given β, that γ is obligatory given α, and that $\neg\beta$ is obligatory given γ. Intuitively, is natural to conclude that under condition β, each of α and γ is obligatory (even though we may not want to conclude for $\neg\beta$ under the same condition, as we shall see when discussing contrary-to-duty conditions). But under the above possible worlds account, defined as above modulo the code, both $O(\alpha/\beta)$ and $O(\gamma/\beta)$ come out false. Consider the world w that puts β true but α and γ false. This world is minimal among all β-worlds under the relation defined : it violates only one of the three explicit norms (the first one), and clearly no β-world w' violates none of them, for when w' satisfies α so as not to violate the first one, it will violate the third or second according to whether it satisfies γ or not.

For readers indifferent to the philosophical motivation, the work of this paper may still have some formal interest. Deontic logic has fallen into ruts. The older rut is the axiomatic approach, with its succession of propositional and occasionally quantified calculi. The newer one is possible-worlds semantics, with endless minor variations in the details. We depart from these confines, with an approach which, whilst syntactic rather than semantic, is not at all axiomatic in character. It could be called *iterative*.

To be sure, any really adequate representation of norms requires much more than truth-functional and deontic operators. It needs to represent the instantiation of variables and indeed full quantification, as well as the passage of time, human agency, competences and powers, bearers and counterparties of obligations, and so on. Any formal language lacking the means to handle such elements will inevitably yield counter-intuitive results in certain applications. Nevertheless, our policy in this paper is to work with the simplest possible syntactic apparatus, reserving complex machinery until the exact limits of the more spartan one are clear – and only in so far as it is confirmed that its essential ideas are indeed "on the right track".

2. Background: constructions of Stenius and Alchourrón/Bulygin

2.1. A first move by Stenius 1963

In a little-known paper, Stenius (1963) made what is perhaps the first move to deal with our "fundamental problem". We recall the essential ideas. Stenius works with a propositional language with boolean connectives that may be applied to any formulae, and the deontic operator for obligation, which is allowed to operate only on purely boolean formulae – so that embedded deontic connectives are not allowed. A normative code is identified with a set A of formulae of the form $O\alpha$, where α is boolean, that is closed under an appropriate consequence operation. Stenius adds to the usual evaluation rules for the truth-functional connectives one for obligation, putting $v_A(O\alpha) = 1$ iff $O\alpha \in A$.

At the time of its formulation, this simple construction represented a major, and rather isolated break with the tradition that had set in since (von Wright 1951). It considers norms modulo a normative code, fixed as parameter. Although Stenius does not use the phrase, we can say that his approach heeds the warning: no logic of norms without attention to the normative systems in which they occur.

However, the construction faces three basic difficulties – an apparent circularity and two technical gaps. The former appears as soon as we ask: what is the consequence operation under which A is taken to be closed? It cannot be *classical* consequence, as the elements of A are not purely boolean, but are of the form $O\alpha$. It must be an *already given* consequence operation over at least atomic *deontic formulae* – which is what we are trying to reconstruct in a philosophically principled manner. One technical gap is that the construction has no means of handling conditional obligations and permissions – which, it is generally accepted, are not reducible to unconditional ones. The other is that it makes no attempt to handle positive permissions, whether conditional or unconditional.

2.2. The construction of Alchourrón/Bulygin 1981

Alchourrón and Bulygin (1981) took Stenius' idea further, avoiding the circularity and plugging one of the two gaps – positive permission, but not conditionality; see also Alchourrón (1993, especially section 3.2.4). We recall their construction, reformulating it somewhat to bring out more clearly the formal structure. [2]

Again, one works with a propositional language with boolean connectives operating on any formulae, and deontic connectives applicable only to purely boolean formulae, so that embedded deontic connectives are not admitted. The deontic connectives involved are now both $O\alpha$ and $P\alpha$, for obligation and positive permission.

An *unconditional normative code* is defined to be a pair N = (A,B) where A, B are sets of purely boolean formulae. Intuitively they represent, respectively, the states of affairs that the code *explicitly requires* to come into effect, and those that it *explicitly permits* to do so.

There is thus a small, but immensely significant step compared to the sketch of Stenius (1963). Alchourrón and Bulygin appear to have been the first to realize the liberating effect of taking the set of promulgations of a normative code to be made up of purely boolean formulae. At the same time, they consider explicit permissions along with promulgations.

An *unconditional valuation structure* is a pair (N,v) where N = (A,B) is a normative code, and v is an assignment of truth-values to elementary letters. Intuitively, the latter is understood as representing the truth-values of propositions in the real world. Given a valuation structure, one defines an assignment v_N of truth-values to all formulae by means of the following rules, where $\vdash$ is classical logical consequence:

For elementary letters, v_N agrees with v
For boolean connectives, v_N behaves in a boolean manner
$v_N(O\alpha) = 1$ iff $A \vdash \alpha$
$v_N(P\alpha) = 1$ iff $\chi \vdash \alpha$ for some $\chi \in B$.

A formula χ is called *valid* iff $v_N(\chi) = 1$ for every valuation structure.

2.3. Immediate formal features

There is more to these rules than meets the eye, and we comment on their formal and philosophical features.

The last two rules are well-defined, because A,B are sets of purely boolean formulae and also the connectives O,P may be applied only to purely boolean formulae, so that for all formulae $O\alpha$ and $P\alpha$ it is meaningful whether $A \vdash \alpha$ and whether $\chi \vdash \alpha$ for some $\chi \in B$, where $\vdash$ is classical consequence. Of course, another way of writing these conditions is $\alpha \in Cn(A)$, $Cn^{-1}(\alpha) \cap B \neq \emptyset$, where Cn is classical consequence expressed as an operation. We shall move freely from one notation to the other, according to ease of reading.

Clearly, for purely boolean formulae, the truth-value of α does not depend on what the normative code N is, but only on what the truth-values of elementary letters are under v. In contrast, the truth-values of $O\alpha$, $P\alpha$ do not depend on what the truth-values of elementary letters are, but only on what A, B, α are. In other words, $O\alpha$ and $P\alpha$ are treated as records of the status of α in the code N, i.e. as propositions about norms, as desired. Evidently for mixed formulae, e.g. $\beta \wedge \neg O\beta$ where β is purely boolean, the value will in general depend on both v and N.

The rule for permission expresses a notion of positive or strong permission, as contrasted with negative or weak permission P_w, which requires only that $A \vdash \neg\alpha$ fails, and for whose valuation the second set B is not relevant. It tells us that $P\alpha$ is true in a given valuation structure iff α is a classical consequence of some explicitly permitted proposition. This is as it should be, since intuitively each of several actions may be permitted without thereby permitting their conjunction. It contrasts with the rule for obligation, which tells us that $O\alpha$ is true in a given valuation structure iff α is a classical consequence of the entire set of explicitly promulgated propositions.

The rule for strong permission may also be motivated in the following way. We understand an explicit strong permission of α, as a commitment not to allow the set of explicit promulgations to grow in such a way as to render α forbidden. So we put $v_N(P\alpha) = 1$ iff $\neg\alpha \notin Cn(X)$ for every set X of explicit promulgations such that $\neg\beta \notin Cn(X)$ for all $\beta \in B$. Clearly this is equivalent to the above definition, and it will be useful when we come to consider conditional norms.

2.4. Independent versus dependent rules

In the construction of Alchourrón and Bulygin, permissions are determined entirely by the set B, and promulgations have no role in generating them. Likewise, obligations are determined entirely by the set A of promulgations, and explicit permissions have no effect in limiting them. The former feature yields the invalidity of the "should implies may" principle $O\alpha \rightarrow P\alpha$ as well as $(O\alpha\wedge P\beta) \rightarrow P(\alpha\wedge\beta)$, and the latter gives rise to the invalidity of $P\alpha \rightarrow \neg O\neg\alpha$ (which of course could also be written contrapositively as $O\alpha \rightarrow \neg P\neg\alpha$, or again as $P\alpha \rightarrow P_w\alpha$ where P_w is weak permission).

> *For the invalidity of $O\alpha \rightarrow P\alpha$.* Consider the code where A = $\{\alpha\}$ and B = $\{\beta\}$. Here (and in all examples) we take $\alpha,\beta,...$ to be distinct elementary letters. Then $v_N(O\alpha) = 1$ since $\alpha \vdash \alpha$, but $v_N(P\alpha) = 0$ since $\beta \vdash \alpha$ fails, and thus $v_N(O\alpha \rightarrow P\alpha) = 0$. *For the invalidity of $(O\alpha\wedge P\beta) \rightarrow P(\alpha\wedge\beta)$.*Use the same example. Since $\alpha \vdash \alpha$ we have $v_N(O\alpha) = 1$, and since $\beta \vdash \beta$ we have $v_N(P\beta) = 1$, so that $v_N(O\alpha\wedge P\beta) = 1$. But $\beta \vdash \alpha\wedge\beta$ fails, so $v_N(P(\alpha\wedge\beta)) = 0$. *For the invalidity of $P\alpha \rightarrow \neg O\neg\alpha$.* Put A = $\{\neg\alpha\}$ and B = $\{\alpha\}$. Then clearly $v_N(P\alpha) = 1$, and also $v_N(O\neg\alpha) = 1$ so that $v_N(\neg O\neg\alpha) = 0$.

We note in passing that if P is read as weak permission then the construction validates trivially the third of these principles, validates the second without restriction, and validates the first under the hypothesis that A is classically consistent.

It is of interest to observe that a variation on the evaluation rule for permissions allows the set of explicit promulgations to have a part in generating them. Simply replace the rule by the following, which brings A into play alongside B.

$$v_N(P\alpha) = 1 \text{ iff } \{\chi\}\cup A \vdash \alpha \text{ for some } \chi \in B.$$

That is, $P\alpha$ is true in the valuation structure iff it is a consequence of some explicitly permitted proposition taken together with all the explicitly promulgated propositions. We call this the *promulgation-dependent* version of the rule for permission, in contrast with Alchourrón and Bulygin's *independent* version. Then the formula $(O\alpha\wedge P\beta) \rightarrow P(\alpha\wedge\beta)$ becomes valid; and although strictly speaking $O\alpha \rightarrow P\alpha$ remains invalid, we have $v_N(O\alpha \rightarrow P\alpha) = 1$ whenever $B \neq \emptyset$.

> *For $(O\alpha\wedge P\beta) \rightarrow P(\alpha\wedge\beta)$.* Suppose $v_N(O\alpha) = 1$ and $v_N(P\beta) = 1$. By the former, $A \vdash \alpha$ and by the latter, $\{\chi\}\cup A \vdash \beta$ for some $\chi \in B$. Hence $\{\chi\}\cup A \vdash \alpha\wedge\beta$ and so $v_N(P(\alpha\wedge\beta)) = 1$. *For $O\alpha \rightarrow P\alpha$.* Observe that whenever $v_N(O\alpha) = 1$ we have $A \vdash \alpha$ so that $\{\chi\}\cup A \vdash \alpha$ for all boolean formulae χ and in particular for some $\chi \in B$ provided B is non-empty, so that $v_N(P\alpha) = 1$.

Conversely, one may wish to allow explicit permissions to influence obligations and, more particularly, to limit them. The idea is to trim the explicit promulgations when there is conflict, giving priority to generated permissions. It is somewhat akin to the operation of belief contraction. As Sven Ove Hansson has remarked (personal communication) it can also be seen as introducing a special kind of defeasibility: a permission can prevent an explicit promulgation from creating an obligation. However, as one would expect, there are a number of different ways of defining such operations, each with its attractions and blemishes. The simplest procedure would be to put $v_N(O\alpha) = 1$ iff $A \vdash \alpha$ and there is no $\chi \in B$ with $\chi \vdash \neg\alpha$, i.e. iff $A \vdash \alpha$ and $v_N(P\neg\alpha) = 0$. Another approach is to put $v_N(O\alpha) = 1$ iff $X \vdash \alpha$ for every maximal subset $X \subseteq A$ that does not imply the negation of any element of B. We do not set out systematically the behaviour of such *permission-dependent* accounts of obligation, nor attempt to adjudicate between them.

It is also natural to ask whether one can formulate, in a sensible way, a fourth version of the valuation rules in which they are *mutually* dependent, with explicit promulgations creating permissions and explicit permissions limiting obligations. In the author's view, this is a particularly delicate task, even after settling on a particular account of permission-dependence. We do not attempt it here.

2.5. Alchourrón's 1993 extension

As remarked in section 2.2, the construction of Alchourrón Bulygin (1981) has no apparatus for handling conditional obligations and permissions, and our main purpose in this paper is to develop it to do so.

An attempt in this direction was made by Alchourrón (1993, section 3.4). He extends the set of formulae as one would expect, by introducing two-place connectives for conditional obligation and permission, read intuitively as α is obligatory (permitted) on condition β. Like the unary deontic connectives, they may be applied only to purely boolean formulae, whilst boolean connectives may be applied to arbitrary formulae.

Alchourrón's construction is intended to model only indefeasible conditional norms. The idea is to work with an indexed family of unconditional normative codes $N_u = (A_u, B_u)$, one code for each assignment u of truth-values to propositional letters. If $\underline{N}$ is such a family, then a conditional obligation $O(\alpha/\beta)$ is counted as true in $\underline{N}$ iff α is a classical consequence of every set of promulgations associated with an assignment satisfying β. In other words, $v_{\underline{N}}(O(\alpha/\beta)) = 1$ iff for every assignment u, if $u(\beta) = 1$ then $A_u \vdash \alpha$.

However, this construction has certain shortcomings. On the one hand, there are principles that it validates "too easily". In particular, the principle of monotony $O(\alpha/\beta) \rightarrow O(\alpha/\beta\wedge\gamma)$, also known as "strengthening the antecedent", holds without restriction, simply because every assignment satisfying $\beta\wedge\gamma$ satisfies β. Although we are considering norms intended as indefeasible, we shall see that there are reasons to believe that when $\beta\wedge\gamma$ is "contrary-to-duty", monotony requires qualification. In general terms, Alchourrón's definitions take no heed of the subtleties arising for contrary-to-duty conditions.

On the other hand, there are principles that it fails "too radically". The construction takes as "given" a function associating an unconditional code with each "possible world". But it leaves quite arbitrary which codes are associated with what worlds. As a result, transitivity for conditional obligation, $O(\alpha/\beta)\wedge O(\beta/\gamma) \rightarrow O(\alpha/\gamma)$, fails. Roughly speaking, the left hand side tells us that whenever $u(\gamma) = 1$ then $A_u \vdash \beta$ and whenever $u(\beta) = 1$ then $A_u \vdash \alpha$. But, unless we make the unmotivated assumption that the code A_u contains only items that are true under u, $A_u \vdash \beta$ does not imply $u(\beta) = 1$ and so we cannot combine the two pieces of information. We do not suggest that the transitivity principle is acceptable without restraint. Indeed, we shall see that it also needs to be qualified in contexts involving "contrary-to-duty" conditions, as well as those involving the passage of time or variation in the bearer of the obligation. But in Alchourrón's construction the principle fails without residue.

For these reasons, we shall follow a more analytical approach that replaces the family of unconditional codes by a single conditional code. Like Alchourrón (1993), we intend our construction for substantively indefeasible conditionals, although we shall see that even for them a certain kind of "technical defeasibility" arises from contrary-to-duty conditions, so that monotony, transitivity and some other principles hold only under certain hypotheses.

3. The iterative approach

3.1. Preliminaries and guiding ideas

We consider the same formulae as for Alchourrón (1993). We define a *conditional normative code* to be a pair N = (C,D) where C,D are sets of ordered pairs (φ,ψ) of purely boolean formulae.[3] A *conditional valuation structure* is a pair (N,v) where N is a conditional normative code, and v is an assignment of truth-values to elementary letters. As before, we take a formula χ to be valid iff $v_N(\chi) = 1$ for every valuation structure (N,v).

Intuitively, $(\varphi,\psi) \in C$ means that the normative system explicitly requires φ in the situation ψ, and $(\varphi,\psi) \in D$ means that it explicitly permits φ in that situation. Of course, the direction of reading (φ,ψ), whether from right to left or inversely, is of no significance as long as it is fixed; we use the right-to-left reading to harmonize with that of formulae $O(\varphi/\psi)$. The formula φ is called the *head* or *consequent* of (φ,ψ) and ψ is called its *body* or *antecedent*.

For simplicity we focus on developing "independent" rules for the operators of conditional obligation and permission, indicating only briefly in section 3.5 how they might be adapted to obtain "promulgation-dependent" and "permission-dependent" versions.

To appreciate what is going on, it may help to bear in mind the general policies guiding the technicalities. They will be explained more fully as they come into play.

- As indicated above, we work with a *single conditional code* rather than with an indexed family of unconditional ones;
- When developing the explicitly given elements of a code, we use *iterated detachment* instead of consequence, in order to not to lose the directionality of conditional norms;
- At the same time, we distinguish between *gross* and *net* output, the latter moderating the former in order not to drown conditional norms whose bodies are "contrary-to-duty".

3.2. Conditional obligation

<u>*3.2.1. Consequence versus detachment*</u>

Let N = (C,D) be a conditional normative code. Let β be a purely boolean formula, which we will be considering in the role of condition. Let v be any assignment of truth-values to elementary statements. We generalize in a natural way the Alchourrón/Bulygin valuation rule for unconditional obligation, putting:

$$v_N(O(\alpha/\beta)) = 1 \text{ iff } \alpha \in \text{out}(C,\beta)$$

where out(N,β) is the "output" of the set C under the condition β. The question is: how to define out(C,β)?

At first sight, the answer may appear to be obvious: just put out(C,β) = Cn($\{\beta\}\cup\{\psi\rightarrow\varphi$: $(\varphi,\psi) \in C\}$), i.e. the classical consequences of the set consisting of condition β taken together with the material implications associated with the explicit promulgations in C. However, such a definition would be far from the mark. For one thing, it puts $\beta \in$ out(C,β). For another, it permits contraposition. This is not due to its use of material implication $\rightarrow$, but rather to the way in which classical consequence is deployed. Indeed, as Lloyd Humberstone has remarked (personal communication), for any function f on promulgation sets C, if we define out(C,β) = Cn($\{\beta\}\cup$f(C)) then we immediately have $\alpha \in$ out(C,β) iff $\alpha \in$ Cn($\{\beta\}\cup$f(C)) iff $\neg\beta \in$ Cn($\{\neg\alpha\}\cup$f(C)) iff $\neg\beta \in$ out(C,$\neg\alpha$).

On the other hand as is widely recognised, contraposition is quite undesirable from the point of view of the intended reading. For example, suppose that the normative system contains a pair (φ,ψ) where φ is "I show the inspector my train ticket" and ψ is "The inspector asks to see my train ticket". The normative system requires that I show my ticket in the case that the inspector so requests, but it certainly need not require that the inspector be silent if I do not show it. For an example with no distracting change of agent and little trace of the passage of time, consider the rule that if you change address you should inform the local post office; this hardly implies that if you do not inform the post office, then you should not change address. There is a profound asymmetry or directionality in conditional norms, which is lost in brutal application of the consequence operation.

For this reason, we need to use a quite different technique in defining the output of a code under a condition, and the one that seems most appropriate is *iteration of successive detachments*. The exact construction is quite subtle and we shall proceed in two steps. In the first step we formulate a concept of "gross output", serving as a first approximation. The second step nuances gross output in order to deal adequately with "contrary-to-duty" conditions, giving the desired notion of "net output". However, the time spent on gross output is not wasted, as it remains an essential tool for analysing the final construction.

3.2.2. Gross output

Let N = (C,D) be a conditional normative code. For any boolean formula β, we define the *gross output* of N under the condition β, written out*(N,β) or more usually as out*(C,β) as its definition does not depend on the set D of explicit permissions, to be the product of iterated detachments on the explicit promulgations of the code, starting from β, and allowing also for secondary operations along the way. To be precise, we put $\varphi \in$ out*(C,β) whenever any of the following six inductive clauses hold:

- *Basis*: $(\varphi,\beta) \in C$

- *Iterated detachment*: $\varphi_1 \in$ out*(C,β) and $\varphi \in$ out*(C,$\beta\wedge\varphi_1$)
- *Body strengthening*: $\varphi \in$ out*(C,β_1) and $\beta \vdash \beta_1$
- *Head weakening*: $\varphi_1 \in$ out*(C,β) and $\varphi_1 \vdash \varphi$
- *Head conjunction*: $\varphi_i \in$ out*(C,β) for all $i \leq n$ and $\varphi = \wedge\varphi_i$
- *Body disjunction (or proof by cases)*: $\varphi \in$ out*(C,β_i) for all $i \leq n$ and $\beta = \vee\beta_i$.

In the last two clauses, it is understood that $n \geq 1$. When the identity of C is fixed and clear we will also write out*(C,β), alias out*(N,β), simply as out*(β).

Just as for the unary counterpart, the value of $v_N(O(\alpha/\beta))$ in a valuation structure (N,v) depends on the formulae α,β and the normative code N (more specifically, on its positive ingredient C), but not on the "real world" truth-assignment v. We are attributing truth-values to deontic formulae, but read as propositions about norms.

We draw attention to the formulation of iterated detachment. It conjoins the condition β with its output φ_1 in order to activate another condition. Transitivity is what we would have if only φ_1 could be used. Clearly, transitivity follows from iterated detachment with body strengthening.

It is important to notice that *in general we do not have* $\beta \in$ out*(β). For example, put C = {(α,β)}; then $\beta \notin$ out*(β) = Cn(α). Indeed, it is easy to check by induction that out*(β) is always a subset of the closure under classical consequence of the set of all heads of elements of C.

> *Example 2 (Joined forces).* Put C = {($\neg\alpha,\tau$), (γ,β)}. Here (and throughout the paper) τ is a tautology. Then out*($\alpha\vee\beta$) contains $\neg\alpha$ (by base and body strengthening) and also γ (by base, body strengthening and iterated detachment), but it contains neither β nor $\alpha\vee\beta$.

3.2.3. Shortcoming

Gross output runs into difficulty when the condition under consideration is inconsistent with the heads of norms that it activates.

Recall that since (Chisholm 1963), a condition for an obligation is informally called "contrary-to-duty" if it is itself in violation of some obligation. For example, one may have the responsibility for repairing the fender of a neighbour's car in the (contrary-to-duty) case that one has carelessly backed into it in the parking lot. It is true that this example involves implicit reference to the passage of time; indeed that is a feature of most real-life examples, as noticed by (Loewer and Belzer 1983). But it does not appear to be an intrinsic feature. As observed by (Jones and Pörn 1985), there are also real-life examples from which time is absent. For instance, as remarked by (Prakken and Sergot 1996), regulations for a group of holiday cottages may require the dwellings to be fenceless, but also require fences to be white in the contrary-to-duty situation that they exist. It is also important to bear in mind, as again emphasized by Prakken and Sergot (1996), that contrary-to-duty conditions arise for indefeasible obligations as readily as for defeasible ones.

A simple formal example, deriving ultimately from Chisholm (1963), shows how our gross output may behave counter-intuitively under a contrary-to-duty condition.

> *Example 3 (Chisholm).* Take C = {(γ,β), ($\neg\beta,\tau$), ($\neg\gamma,\neg\beta$)} and consider condition β. The code thus promulgates $\neg\beta$ under the tautologous condition,

and contains two other conditional promulgations, one with body β and the other with body $\neg\beta$. Consider β as condition. Intuitively, we would like to get $O(\gamma/\beta)$, but not $O(\neg\beta/\beta)$, nor $O(\neg\gamma/\beta)$. But clearly by applications of the base clause, body strengthening and iterated detachment, we get in $out^*(\beta)$ all of γ, $\neg\beta$, $\neg\gamma$, and thus by head conjunction and weakening also every boolean formula.

3.2.4. Net output

Why is gross output unsuitable for contrary-to-duty conditions? When we entertain a condition β, we want to look beyond the question of whether that condition should never have been realized, and consider only what should be done given that it *is* realized. So, if the code yields the negation of the condition as an output, we need to make adjustments.

In this spirit, we make a distinction between *gross* and *net* output. The essential idea of net output is that given a condition β, we use only those pairs in C whose heads are consistent with β.

It is tempting to do this by putting $out(C,\beta) = out^*(C_\beta,\beta)$ where C_β is a maximal subset $C' \subseteq C$ such that $out^*(C',\beta)$ is classically consistent with β. However, this maxichoice approach (which the author followed for some time) runs into at least two serious technical difficulties.

In the first place, it fails to guarantee certain inferences that intuitively should be accepted. Indeed, this is illustrated by the "Möbius strip" example, already employed in section 1 in connection with a "possible worlds" approach.

> *Example 4 (Möbius strip).* Let $C = \{(\alpha,\beta), (\gamma,\alpha), (\neg\beta,\gamma)\}$. Intuitively, we would like to have $\alpha,\gamma \in out(\beta)$. But there are three possible values of the "maxichoice" set C_β , namely the three two-element subsets of C. For only two of those three values of C_β do we get $\alpha \in out(\beta)$, and for only one is $\gamma \in out(\beta)$.

In the second place, the maxichoice approach runs into difficulty with disjunction. Specifically, it seems able to give an adequate account of "proof by cases" only in the case of a finite language, and there in an exceedingly complex manner.

> Note that the definition of $out^*(C,\beta)$ makes use of elements of $out^*(C,\beta_i)$ for arbitrary β_i with $\beta = \vee\beta_i$. Hence we should look at not only those $C' \subseteq C$ with $out^*(C',\beta)$ consistent with β but also, for each such β_i, the $C' \subseteq C$ with $out^*(C',\beta_i)$ consistent with β_i. In the finite case it may be possible to sort this out inductively, by first considering values of β that are atoms and then working up the boolean lattice step by step. But this would be quite complex and disappointingly restricted.

A better procedure, we believe, is to control for "consistency with the condition" in a more piecemeal way, using a labelled deductive system in the sense of (Gabbay 1996), adapting ideas in particular from van der Torre (1997). The elements of $out(C,\beta)$ are no longer taken to be plain formulae φ, but *labelled formulae* φ: L. There is nothing mysterious about these. They are simply pairs (φ,L); we use a colon merely for visual relief. The purpose of L is to serve as a record of the heads of all explicit promulgations used in getting φ. When there is no application of reasoning by cases, L can be taken to be a boolean formula, that grows by conjunction as more promulgations are used, and for heuristic purposes it is advisable to keep this simple case in mind. But in general, to cover the parallel tracks created through reasoning by cases, we need to consider labels that are sets of boolean formulae.

Formally we define out(C,β) by putting φ: L $\in$ out(C,β) whenever any of the following inductive clauses hold, *provided that β is consistent with each element of L*. The six clauses parallel those for gross output (differing only in the presence and specification of labels) and we call them by the same names.

- (φ,β) $\in$ C and L = {φ}
- φ_1: $L_1 \in$ out(N,β) and φ: $L_2 \in$ out(N,$\beta_1 \wedge \varphi_1$), and L = {$\lambda_1 \wedge \lambda_2$: $\lambda_i \in L_i$}
- φ: L $\in$ out(N,β_1) and $\beta \vdash \beta_1$
- φ_1: L $\in$ out(N,β) and $\varphi_1 \vdash \varphi$
- φ_i: $L_i \in$ out(N,β) for all $i \leq n$ and $\varphi = \wedge\varphi_i$, and L = {$\wedge\lambda_i$: $\lambda_i \in L_i$}
- φ: $L_i \in$ out(N,β_i) for all $i \leq n$ and $\beta = \vee\beta_i$, and L = $\cup${L_i: $i \leq n$}.

To understand the label L = {$\wedge\lambda_i$: $\lambda_i \in L_i$} in the clause for head conjunction (and likewise the label in the clause for iterated detachment) remember that for each i, L_i consists of a number of formulae, one for each case in the derivation of φ_i . The set of cases involved in the derivation of the conjunction is the Cartesian product of the sets of cases involved in the derivations of the conjuncts. The label of the conjunction is formed accordingly: each element of L is a conjunction with one conjunct from each of the labels L_i. Evidently, when a derivation contains no applications of body disjunction (proof by cases) then each label L_i contains a single formula, and L will consist of the conjunction of those formulae. Again, if all the L_i are singletons except for L_1 which contains two formulae λ_{11} and λ_{12}, then L will contain two formulae, namely $\lambda_{11} \wedge \lambda_2 \wedge \ldots \wedge \lambda_n$ and $\lambda_{12} \wedge \lambda_2 \wedge \ldots \wedge \lambda_n$.

When a label consists of a single formula, we shall usually simplify notation and write the label as that formula. As labels are used only for consistency checks, we will also take the liberty of writing any formula in a label in its simplest classically equivalent form. Thus, for example, the label {$\tau \wedge \varphi$} is written as φ.

The purpose of the labels is to control the production of elements of the net output, rather than to differentiate them within that output. For this reason, we use the "flattening" definition for unlabelled formulae, that puts $\varphi \in$ out(C,β) iff φ: L $\in$ out(C,β) for some label L. We shall also express the latter as: $\varphi \in$ out(β) with label L.

We illustrate the definition with three examples. The first is the "window" example of (von Wright 1964), and illustrates the situation where the condition under consideration activates promulgations with incompatible heads. The other two are the Chisholm example and our "Möbius strip", already considered in connection with gross output. They illustrate "contrary-to-duty" situations, where the condition under consideration is incompatible with the head of one of the explicit promulgations that are activated. As has been emphasized by Prakken and Sergot (1996) and by van der Torre (1997), these are quite different kinds of conflict. Whereas consistency checks may be used to handle contrary-to-duty conditions, the standard procedure for resolving conflicts between incompatible heads is prioritisation of the explicit promulgations. Such prioritisation could be integrated into our iterative account, but in this paper we consider only the unprioritised case.

Example 5 (von Wright). Put $C = \{(\alpha,\beta), (\neg\alpha,\gamma)\}$, and consider condition $\beta\wedge\gamma$. Intuitively, one may think of α as representing "the window is closed", β as "the sun is shining", γ as "it is raining", noting that the latter two are mutually consistent. We have $\alpha,\neg\alpha \in \text{out}(\beta\wedge\gamma)$ with labels $\alpha,\neg\alpha$ respectively, but the consistency check on the clause for head conjunction prevents us getting $\alpha\wedge\neg\alpha \in \text{out}(\beta\wedge\gamma)$, since its label $\alpha\wedge\neg\alpha$ is inconsistent and so inconsistent with the condition $\beta\wedge\gamma$.

Example 6 (Chisholm). We compare the net output under condition β with the gross one in the contrary-to-duty Chisholm example $C = \{(\neg\beta,\tau), (\gamma,\beta), (\neg\gamma,\neg\beta)\}$. Clearly $\gamma \in \text{out}(\beta)$ with label γ, since γ is consistent with β. But although $(\neg\beta,\tau) \in C$, we cannot conclude that $\neg\beta \in \text{out}(\beta)$ since its label $\neg\beta$ is inconsistent with β. And so we cannot go on to get $\neg\gamma \in \text{out}(\beta)$ with label $\neg\beta\wedge\neg\gamma$.

Example 7 (Möbius strip). Recall that in this example $C = \{(\alpha,\beta), (\gamma,\alpha), (\neg\beta,\gamma)\}$. Intuitively, we would like to have $\alpha,\gamma \in \text{out}(\beta)$, but $\neg\beta \notin \text{out}(\beta)$. Clearly, gross output puts each of α, $\gamma,\neg\beta \in \text{out}^*(\beta)$. As we have seen, a "maxichoice" approach to net output can give three different results according to which we choose among the three maximal subsets of C whose gross output under β is consistent with β. Under the labelled definition of net output we get α, $\gamma \in \text{out}(\beta)$ with labels α and $\alpha\wedge\gamma$ respectively but, as desired, $\neg\beta \notin \text{out}(\beta)$ since its label $\alpha\wedge\gamma\wedge\neg\beta$ is inconsistent with β.

The Möbius strip example illustrates how in the labelling approach we just go on detaching until we hit trouble. Here "trouble" means inconsistency of the condition under consideration with the conjunction of all heads so far used in the derivation – or, in the general case that reasoning by cases is involved, in some case of the derivation. The purpose of the label is to carry along sufficient traces of the derivation to be able to make the consistency check locally.

We end this section by noting some formal properties of net output, that will also be useful in section 7.

(1) *Relation with gross output*. Clearly, whenever $\varphi: L \in \text{out}(C,\beta)$ then $\varphi \in \text{out}^*(N,\beta)$ so that by the flattening convention, $\text{out}(C,\beta) \subseteq \text{out}^*(C,\beta)$. For this reason, although we work with net output in preference to gross, the latter remains an important tool of analysis.

(2) *Consistency proviso: partial redundancy*. In the definition of net output, each of the six clauses is qualified with the consistency proviso that β is consistent with each element of L. In fact, the consistency check is redundant for two of the clauses: head weakening and (as observed by van der Torre: personal communication) body disjunction. This is easily verified by an inductive argument, noting that in the head weakening clause neither the condition nor the label changes, whilst in the body disjunction clause the induction hypothesis tells us that for all $i \leq n$ we have each β_i consistent with each element of L_i, so that $\vee\beta_i$ is consistent with each element of $L = \cup\{L_i : i \leq n\}$ and the proviso is satisfied.

(3) *Consistency properties for output elements*. It is immediate from the presence of the consistency proviso that whenever $\varphi: L \in \text{out}(C,\beta)$, β is consistent with each element of L. It is also clear by induction that whenever $\varphi: L \in \text{out}(C,\beta)$, we have $\lambda \vdash \varphi$ for each

element λ of L. Putting these together and using the flattening convention, it follows that whenever $\varphi \in$ out(C,β), then β is consistent with φ, and so φ is itself consistent.

(4) *Preservation of equivalence and monotony properties: right argument*. We have preservation of equivalence in the right argument: out(C,β) = out(C,β') whenever Cn(β) = Cn(β'). This is by straightforward induction, noting that β will satisfy a consistency proviso iff β' does. However, we do not in general have monotony in the right argument: we do not have out(C,β) $\subseteq$ out(C,β') whenever Cn(β) $\subseteq$ Cn(β'), for it is blocked by the consistency proviso on the clause for body strengthening.

(5) *Preservation of equivalence and monotony properties: left argument*. On the one hand, monotony under inclusion is immediate for the left argument: out(C,β) $\subseteq$ out(C',β) whenever C $\subseteq$ C'. However that we do not have monotony under any natural notion of *consequence* in that argument. In fact, we do not even have preservation of equivalence, for net output is very sensitive to the formulation of the explicit conditional promulgations in C.

> *Example*. Compare the value of out(C,β) when C = {($\alpha\wedge\neg\alpha$,β)} with its value when C = {(α,β), ($\neg\alpha$,β)}. In the former case, out(C,β) = $\emptyset$ whereas in the latter, out(C,β) contains each of α,$\neg\alpha$ but not their conjunction.

It is debatable whether this dependence of the value of net output on the manner of presentation of the explicit promulgations should be regarded as a virtue or a vice. On the one hand, independence from syntax simplifies structure, and is generally a desirable feature in a formal representation, to be abandoned with reluctance. On the other hand, as observed by (Prakken and Sergot 1997, sections 7-8), consideration of the "gentle murderer" example of (Forrester 1984) and variations on it suggest that in the present context such dependence may be intuitively desirable. Indeed, as noted by (Carmo and Jones 1997, section 8), this is particularly evident in the well-known "Reykjavik scenario" of (Belzer 1987).

We recall that in the Reykjavik scenario, the explicitly presented promulgations are three: neither Reagan nor Gorbatchov should be told the secret, if Reagan is told then Gorbatchov should be told, and if Gorbatchov is told then Reagan should be told. Intuitively, under the condition that Gorbatchov is told, we would like to be able to conclude that Reagan should be told, without also deriving its negation. On the other hand, if the first item is broken down into two parts, saying separately that Reagan should not be told and that Gorbatchov should not be told, so that there are four explicitly presented promulgations, then the conclusion to draw seems intuitively more ambiguous, depending on whether promulgations are prioritised according to, say, specificity of their antecedents. In the absence of prioritisation we seem authorised to conclude both that Reagan should be told and that he should not; with prioritisation, only the former. The iterative account, which we have formulated for non-prioritised codes, fits well with these intuitions.

> Take the first representation as {($\neg\alpha\wedge\neg\beta$,$\tau$), ($\alpha$,$\beta$), ($\beta$,$\alpha$)} and the second as {($\neg\alpha$,$\tau$), ($\neg\beta$,$\tau$), ($\alpha$,$\beta$), ($\beta$,$\alpha$)}. For the first representation we have that $\alpha \in$ out(C,β) with label α, but $\neg\alpha \notin$ out(C,β) since the only available derivation introduces the label $\neg\alpha\wedge\neg\beta$, inconsistent with β. But for the second representation we have both $\alpha \in$ out(C,β) with label α, and $\neg\alpha \in$ out(C,β) with label $\neg\alpha$, although $\alpha\wedge\neg\alpha \notin$ out(C,β).

3.2.5. Limitations

We note two aspects of our construction that could be questioned. One concerns the role of detachment, and the other concerns the "irreversibility" of conditions. We suggest that both of these indicate severe limitations to its applicability to real life examples, arising from its total abstraction from questions of time and of bearer.

Iteration of detachment. The definition of gross output authorises iteration of detachment, as also does net output under the consistency proviso. But as Sven Ove Hansson has reminded the author (personal communication), there are simple examples in which we would not want this to be so.

> *Example*. Let C = $\{(\gamma,\alpha), (\alpha,\beta)\}$ where β is "John owes Peter $1000", α is "John pays Peter $1000", and γ is "Peter gives John a receipt for $1000". Then $\gamma \in$ out(β) so that $v_N(O(\gamma/\beta) = 1$. On the other hand, intuitively we would say that in the circumstance that John owes Peter $1000, considered alone, Peter has no obligation to write any receipt. That obligation arises only when John fulfils *his* obligation.

There appear to be two principal sources of difficulty here. One concerns the passage of time, and the other concerns bearers of the obligations.

We recall that our representation of norms abstracts entirely from the question of time. Evidently, this is a major limitation of scope, and leads to discrepancies with real-life examples, where there is almost always an implicit time element. This may be transitive, as when we say "when β holds then α should eventually hold", or "...should simultaneously hold". But it may be intransitive, as when we say "when β holds then α should hold within a short time" or "...should be treated as a matter of first priority to bring about". Clearly, iteration of detachment can be legitimate only when the implicit time element is either nil or transitive.

Our representation also abstracts from the question of bearer, that is, who (if anyone) is assigned responsibility for carrying out what is required. This too can lead to discrepancies. Iteration of detachment becomes questionable as soon as some promulgations have different bearers from others, or some are impersonal (i.e. without bearer) while others are not. Only when the locus of responsibility is held constant can such an operation take place.

Sven Ove Hansson's example above involves both of these factors. When neither is involved, iteration of detachment does appear to be appropriate, as in the following example, based on instructions to authors preparing manuscripts.

> *Example*. Let C be the same as above, but where β is "The text area is 25 by 15 cm", α is "The font size for the main text is 12 points", and γ is "The font size for the list of references is 10 points".

Irreversibility of conditions. The consistency proviso for the base clause of the definition of net output prevents detachment of the head of an explicit promulgation when it is inconsistent with the body. We are thus in effect excluding the possibility that what should be the case under condition β, is its opposite $\neg\beta$.

As remarked by (Bengt Hansson 1969, section 13), in non-temporal contexts this is surely part of what is involved in taking a condition seriously. On the other hand, it does make it awkward to deal with examples saying that if condition β holds at a certain time, $\neg\beta$ should be hold subsequently. An adequate representation of such examples would have to go beyond our framework and make explicit reference to the passage of time.

As mentioned in the Introduction, our policy is to sail closer to the Scylla of oversimplification than the Charybdis of excessive complexity. In our view, formal machinery to handle time and bearers (e.g. by indices) should be introduced after the simpler structure is well understood and it becomes clear that its essential ideas are on the right track. We do not do so in this paper.

4. Benchmark Examples

In the following table we summarize the performance of the iterative account in a number of well-known examples, four of which (von Wright's "windows", Chisholm's "help and inform", Belzer's "Reykjavik scenario" and our "Möbius strip") have been mentioned in the preceding discussion. In order not to overburden the presentation, we assume a passing acquaintance with the examples, and use letters mnemonically for the statements involved. Thus in the example from Chisholm (1963), h is for "you go to help your neighbours" and i for "you inform them that you are coming". However, in the last three examples, the letters have no mnemonic value. For uniformity of font, we use t for the tautology, rather than the τ elsewhere in the paper.

The first column gives the original source of the example and its usual name. The first four have been discussed extensively in the literature; the last two are presented here for the first time.

In the second column we write the explicit promulgations of the example, using the notation of this paper. The original presentations were written in various notations, most commonly with a conditional obligation operator $O(\neg f/t)$ where we have $(\neg f,t)$. In two of the examples ("Reykjavik scenario", "apples and pears") we consider variant sets of explicit promulgations.

The third column contains the conditions under consideration. In most of the examples we consider more than one condition.

Source and name	**Explicit promulgations**	**Condition**	**No**	**Yes**
von Wright 1964: window closing	(c,r), (¬c,s)	r∧s	c∧¬c	c, ¬c
Chisholm 1963: help and inform	(h,t), (i,h), (¬i,¬h)	1. t 2. ¬h	 h, i	h, i ¬i
Forrester 1984: gentle murderer	(¬k,t), (g,k)	1. t 2. k	g ¬k	¬k g
Belzer 1986: Reykjavik scenario	1. (¬r∧¬g,t), (r,g), (g,r) 2. (¬r,t), (¬g,t), (g,r), (r,g)	r r	¬g	g g, ¬g
Prakken/Sergot 1996: cigarettes for killer	(¬k,t), (¬c,t), (c,k)	k	¬k	c, ¬c
Prakken/Sergot 1996: white fence and dog	(¬f,t), (w∧f,f), (w∧f,d)	1. f 2. d 3. f∧d	¬f ¬f	w∧f w∧f, ¬f w∧f
van der Torre 1997: apples and pears	(a∨p,t), (¬a,t) 4. (a∨p,t) 5. (a,¬p), (p,¬a)	1. t 2. ¬a 3. a ¬a ¬a	 ¬a, p ¬a, p	¬a, p ¬a, p a∨p a∨p p
van der Torre 1997: joining paths	(a,t), (b,t)	¬a∨¬b	a∧b	a, b
Makinson 1998: Möbius strip	(¬b,c), (c,a), (a,b)	b	¬b	a, c
Makinson 1998: exclusive options	(a∧x,b), (a∧¬x,¬b)	t		a

The fourth column singles out some salient items that are not in the output, while the fifth mentions some that are. In all the examples considered, these columns appear to correspond reasonably well with "informed intuition", remembering that we are taking the explicit promulgations as substantively indefeasible and as unprioritised. Of course, "informed intuition" remains debatable.

5. Conditional Permission

As in the unconditional case, one can formulate a concept of weak permission of α under condition β, as $\neg O(\neg\alpha/\beta)$ i.e. the absence of obligation of the negation of α under the same condition. We are also interested, however, in formulating a concept of strong conditional permission. Working backwards, we put the rule as follows, where D is the set of explicit permissions of a code N = (C,D):

$$v_N(P(\alpha/\beta)) = 1 \text{ iff } \alpha \in \text{perm}(D,\beta).$$

It remains to define perm(D,β), the *permission set* of the code under β. Our construction is guided by an intuitive understanding of an explicit strong permission of α in condition β, as a commitment not to allow the set of explicit promulgations to grow in such a way that α is forbidden under condition β. So it is natural to take $\alpha \in$ perm(D,β) iff $\neg\alpha \notin$ out*(X,β) for every set X of explicit promulgations such that $\neg\varphi \notin$ out*(X,ψ) for all $(\varphi,\psi) \in$ D.

So understood, it is reasonable to require $\varphi \in$ perm(D,β) whenever any of the following three inductive clauses hold:

- *Basis*: $(\varphi,\beta) \in$ D
- *Head weakening*: $\varphi_1 \in$ perm(D,β) and $\varphi_1 \vdash \varphi$
- *Body weakening*: $\varphi \in$ perm(D,β_1) and $\beta_1 \vdash \beta$.

Note that the third clause is one of body weakening – not body strengthening. Given that it is permitted to mow the lawn on Sunday between 10h00 and 12h00, I may conclude that it is permitted to do so on Sunday, but I may not conclude from the latter that it is permitted to mow on Sunday afternoon. It also makes sense in terms of the above intuitive understanding of positive permission, for if $\neg\varphi \in$ out*(X,β) and $\beta_1 \vdash \beta$ then it follows that $\neg\varphi \in$ out*(X, β_1).

In the above, we have thought in terms of gross rather than net output, and the author suspects that it is the better option. To be sure, if net output is used, body weakening should be dropped or suitably qualified. Conversely, it is possible that there are further closure properties that are reasonable under our intuitive understanding and so should be added. But in what follows, we take the above three clauses as constituting a tentative inductive definition of perm(D,β), which we will write simply as perm(β) when the identity of D is fixed and clear.

6. Dependent Rules for the Conditional Operators

If we are considering a normative code N = (C,D), we may wish to allow the set out(C,β) to help generate permissions. One way of doing this might be by adding to the definition of perm(D,β) the following fourth clause:

- *Output dependence*: $\varphi_1 \in$ perm(D,β) and $\psi \in$ out(C,β) and $\varphi_1 \wedge \psi \vdash \varphi$.

Of course, the presence of this clause renders head weakening redundant (provided the output set is non-empty).

Conversely, we may wish to make the rule for conditional obligation permission-dependent, so that permissions limit obligations. In principle, this could be done by using maximal non-implying subsets of the output set, but that goes against the spirit of "keep going until you hit trouble" that underlies our general approach. Perhaps the most appropriate move would be simply to add a second proviso to the definition of net output: *provided* $\neg\varphi \notin perm(D,\beta)$.

7. Behaviour of the conditional operators

7.1. General picture

In general, the behaviour of the operators under the iterative account appears to agree reasonably well with "informed intuition". We have already looked at behaviour in some benchmark examples (section 4); we now consider general principles. In order to put some order into the review, we group them into four categories.

- Those concerning obligation. We consider some that are valid without restriction (extensionality with respect to both antecedent and consequent, weakening the consequent, disjunction in the antecedent, consistency, monotony with respect to the normative system), some that are valid under a hypothesis (strengthening the antecedent, conjunction in the consequent, transitivity), and some that fail outright (reflexivity, contraposition).
- Those concerning positive permission. We mention some that are valid without restriction (extensionality, weakening the consequent and the antecedent, monotony with respect to the normative system) and some that fail outright (reflexivity, contraposition, conjunction in the consequent, transitivity, consistency).
- Those expressing relations between the conditional operators and an unconditional counterpart.
- Relations between obligation and permission.

We report only on the effect of the independent rules. Recall that $\varphi \in out(C,\beta)$ means that φ: $L \in out(C,\beta)$ for some label L, and that the operation P(./.) of conditional permission is the positive or strong one, defined by the rules of section 5, rather than the weak one defined as $\neg O(\neg\alpha/\beta)$.

7.2. Obligation

7.2.1. Valid without restriction

The principles of this section follow from properties (1) - (5) of net output noted in section 3.2.4.

Extensionality. $v_N(O(\alpha/\beta)) = v_N(O(\alpha'/\beta'))$ whenever $Cn(\alpha) = Cn(\alpha')$ and $Cn(\beta) = Cn(\beta')$. *Verification*. For the antecedent argument β, suppose α: $L \in out(\beta)$. Then as noted in section 3.2.4, β is consistent with every element of L. Hence if β' is classically equivalent to β, it is also consistent with every element of L, and the consistency proviso for applying the clause of body strengthening is satisfied. For the consequent argument

α, the property is immediate from the redundancy, also noted in section 3.2.4, of the consistency check for head weakening in the definition of net output.

Weakening the consequent. $O(\alpha/\beta) \rightarrow O(\alpha\vee\gamma/\beta)$ is valid. This is immediate from the redundancy of the consistency check for head weakening in the definition of net output.

Disjunction in the antecedent. $O(\alpha/\beta)\wedge O(\alpha/\gamma) \rightarrow O(\alpha/\beta\vee\gamma)$ is valid. This is immediate from the redundancy of the consistency check for body disjunction in the definition of net output.

Consistency. $\neg O(\perp/\beta)$ is valid. This is immediate from the fact that whenever $\varphi \in out(\beta)$ then φ is consistent. However, the formula $\neg(O(\alpha/\beta)\wedge O(\neg\alpha/\beta))$ is not valid. As a trivial counterexample, put $C = \{(\alpha,\beta), (\neg\alpha,\beta)\}$. Then we have both α: $\alpha \in out(\beta)$ and $\neg\alpha$: $\neg\alpha \in out(\beta)$, so that $v_N(O(\alpha/\beta)\wedge O(\neg\alpha/\beta)) = 1$, but the consistency check prevents us applying head conjunction to get $\alpha\wedge\neg\alpha$: $\{\alpha,\neg\alpha\} \in out(\beta)$.

Monotony with respect to the normative system. This says that if $v_N(O(\alpha/\beta)) = 1$ then $v_{N'}(O(\alpha/\beta)) = 1$ when N' is formed from N by enlarging the set of explicit promulgations. This is immediate from the fact that $out(C,\beta) \subseteq out(C',\beta)$ whenever $C \subseteq C'$. It should be remarked, however, that if one were to define net output using a maxichoice procedure instead of consistency checks with labels, monotony with respect to C would fail.

7.2.2. Valid under a hypothesis

The observations of this section follow from the fact that net output is always included in gross (section 3.2.4), together with properties evident from the definition of the latter (section 3.2.2).

Strengthening the antecedent. $O(\alpha/\beta) \rightarrow O(\alpha/\beta\wedge\gamma)$ holds whenever $out(C,\beta\wedge\gamma) = out^*(C,\beta\wedge\gamma)$. For suppose $\alpha \in out(\beta)$. Then $\alpha \in out^*(\beta)$ so by the body strengthening clause in the definition of gross output we have $\alpha \in out^*(\beta\wedge\gamma) = out(\beta\wedge\gamma)$ by the hypothesis.

The hypothesis $out(C,\beta\wedge\gamma) = out^*(C,\beta\wedge\gamma)$ is quite strong. A trivial example shows that it cannot be dropped. Put $C = \{(\neg\gamma,\beta)\}$; then clearly $\neg\gamma \in out(\beta)$ but $out(\beta\wedge\gamma) = \emptyset$. Nevertheless, we may weaken it to a more "local" hypothesis: the verification above still goes through assuming only that $\alpha \notin out^*(\beta\wedge\gamma) - out(\beta\wedge\gamma)$. Alternatively, if α: $L \in out(\beta)$ for some label L every element of which is consistent with $\beta\wedge\gamma$.

One may regard the above facts either positively or negatively. Looked at positively, they tell us that strengthening the antecedent holds "in normal conditions". On the other hand, we must admit that it fails in certain cases, even though we are considering conditionals that are intended as indefeasible.

In general terms, we may distinguish between two ways in which the principle of strengthening the antecedent may fail for conditional norms. It may fail *substantively*, upon adding a normatively neutral element to the body (more precisely, upon strengthening the body in a way that leaves it normatively neutral). Or it may fail *technically*, when the strengthened body is contrary-to-duty. Even when the principle holds substantively, as under the present iterative account, it may fail technically.

Failure of strengthening the antecedent is a typical formal symptom of the informal property of defeasibility, c.f. (Makinson 1993). We may say that under the iterative account, even substantively indefeasible conditional obligations are technically defeasible.[4]

Conjunction in the consequent. $O(\alpha/\beta)\wedge O(\gamma/\beta)\rightarrow O(\alpha\wedge\gamma/\beta)$ holds whenever $out(C,\beta) = out^*(C,\beta)$. For if $\alpha,\gamma \in out(\beta) \subseteq out^*(\beta)$ we have by the head conjunction clause for gross output that $\alpha\wedge\gamma \in out^*(\beta) = out(\beta)$ by the hypothesis.

Transitivity. $O(\alpha/\beta)\wedge O(\beta/\gamma)\rightarrow O(\alpha/\gamma)$ holds whenever $out(C,\gamma) = out^*(C,\gamma)$. For if $\alpha \in out(\beta) \subseteq out^*(\beta)$ and $\beta \in out(\gamma) \subseteq out^*(\gamma)$ then by body strengthening and the iterated detachment clauses for gross output we have $\alpha \in out^*(\gamma) = out(\gamma)$ by the hypothesis.

In section 3.1 we regarded failure of transitivity as a shortcoming of the account of conditional obligation in Alchourrón (1993). We now see that transitivity does not fully succeed under our construction either. However, the situation is appreciably different in the two contexts. There, the principle failed utterly; here it holds in all "normal" situations – those where $\alpha \notin out^*(\gamma) - out(\gamma)$.

7.2.3. Outright failure

Reflexivity. $O(\beta/\beta)$ fails, for as noted in section 3.2.2, in general $\beta \notin out(\beta)$, even when $out(\beta) = out^*(\beta)$. We recall the trivial example $C = \{(\alpha,\beta)\}$.

Contraposition. $O(\alpha/\beta)\rightarrow O(\neg\beta/\neg\alpha)$ fails as indeed, recalling the "train ticket" example discussed in section 3.2.1, it should. For a trivial counterexample, put $C = \{(\alpha,\beta)\}$. Then clearly $\alpha \in out(\beta)$ so that $v_N(O(\alpha/\beta)) = 1$. But $\neg\beta \notin out(\neg\alpha) = \varnothing$ and thus $v_N(O(\neg\beta/\neg\alpha)) = 0$.

7.3. Positive Permission

As there are no consistency provisos on the clauses defining positive permission, we have only two categories to consider: principles that are valid without restriction, and those that fail outright.

7.3.1. Valid without restriction

Weakening the consequent. $P(\alpha/\beta) \rightarrow P(\alpha\vee\gamma/\beta)$ is valid, immediately from the head weakening clause in the definition of perm (D,β).

Weakening the antecedent. $P(\alpha/\beta) \rightarrow P(\alpha/\beta\vee\gamma)$ is valid, immediately from the body weakening clause in the definition of perm (D,β).

Extensionality. $v_N(P(\alpha/\beta)) = v_N(P(\alpha'/\beta'))$ whenever $Cn(\alpha) = Cn(\alpha')$ and $Cn(\beta) = Cn(\beta')$. Immediate from the above two.

Monotony with respect to the normative system. If $v_N(P(\alpha/\beta)) = 1$ then $v_{N'}(P(\alpha/\beta)) = 1$ where N' is formed from N by enlarging the set of explicit permissions. This immediate from the fact that $perm(D,\beta)$ is monotonic in D, as is easily verified inductively.

7.3.2. Outright failure

Consistency. $\neg P(\bot/\beta)$ fails, i.e. we may have $v_N(P(\bot/\beta)) = 1$. This is essentially because there are no consistency provisos in the definition of the permission set. The simplest example is the trivial one where $D = \{(\bot/\beta)\}$.

Transitivity. $P(\alpha/\beta)\wedge P(\beta/\gamma)\rightarrow P(\alpha/\gamma)$ fails. Counterexample: put $D = \{(\alpha,\beta), (\beta,\gamma)\}$. Then clearly $\alpha \in perm(\beta)$ and $\beta \in perm(\gamma)$, so that $v_N(P(\alpha/\beta)) = 1 = v_N(P(\beta/\gamma))$. But recalling that there is no iteration of detachment in the definition of perm(β), $\alpha \notin$ perm(γ) so $v_N(P(\alpha/\gamma)) = 0$.

Conjunction in the consequent. $P(\alpha/\beta) \wedge P(\gamma/\beta) \rightarrow P(\alpha \wedge \gamma/\beta)$ fails, as it should. Counterexample: put $D = \{(\alpha,\beta),(\gamma,\beta)\}$. Then $\alpha,\gamma \in$ perm(β) so $v_N(P(\alpha/\beta)) = 1 = v_N(P(\gamma/\beta))$. But as there is no head conjunction clause in the definition of perm(β), $(\alpha \wedge \gamma) \notin$ perm(β) so $v_N(P(\alpha \wedge \gamma/\beta)) = 0$.

Contraposition. $P(\alpha/\beta) \rightarrow P(\neg\beta/\neg\alpha)$ fails. Counterexample: paralleling that for obligation, put $D = \{(\alpha,\beta)\}$. Then $\alpha \in$ perm(β) so that $v_N(P(\alpha/\beta)) = 1$, but perm(¬α) = $\emptyset$ so that $v_N(P(\neg\beta/\neg\alpha)) = 0$.

Reflexivity. $P(\beta/\beta)$ fails. Counterexample: paralleling that for obligation, put $D = \emptyset$. Clearly perm(β) = $\emptyset$ for any condition β.

7.4. Relating conditional and unconditional norms

Unconditional obligation is deceptively ambiguous. It may be seen as a limiting case of conditional obligation where the condition is implicitly set at a familiar value. But the implicit value varies according to context and intention, giving rise to at least three different kinds of unconditional.

On the one hand, as already observed by (von Wright 1968), we may understand "α is obligatory" as saying that it is so under a condition representing *zero* information about the present world, i.e. we may take $O\alpha$ as meaning $O(\alpha/\tau)$ where τ is a tautology. In contrast, we may understand it as saying that α is obligatory under *complete* information about the present world, i.e. we may take $O\alpha$ as meaning $O(\alpha/V)$ where V is the set of all boolean formulae presently true. These are minimal and maximal readings. Between them, we may read $O\alpha$ as saying that α is obligatory under *current information* about the present world, i.e. as $O(\alpha/X)$ where X is a set of boolean formulae representing the propositions known (alternatively: believed) about the world.

As remarked by Alchourrón (1993), formal work in the literature focuses almost exclusively on the "minimal" notion because of the formal simplicity of its definition, which moreover parallels the well-known definition of necessity from strict implication in modal logic. The other two notions have hardly been touched in formal accounts. For the "maximal" notion, Von Wright (1968) suggested introducing into the language a special propositional constant σ standing for the "actual" situation, defining $O\varphi$ as $O(\varphi/\sigma)$, and formulating axioms for σ to reflect its intended interpretation. However, as observed by Alchourrón (1993), a satisfying formulation of such axioms is not immediate. For the "intermediate" notion, the author is not aware of any formal accounts in the literature, despite its evident importance.

Intuitively, the "maximal" (resp. intermediate) notion expresses the idea that given the totality of facts true of the present world (resp. known or believed about the present world), a certain state of affairs should be brought about. It is important to notice the tensed character of the last part. If we ignore it, we not only miss part of the meaning, but we can also be led to the quite counterintuitive result that everything unconditionally obligatory is in fact true (resp. is not known to be false). In our definition of net output, in order to deal adequately with contrary-to-duty norms, we have required that an explicit promulgation may only be activated when its head is consistent with the condition that is being entertained. Evidently, when the condition is a maxiconsistent set, then whatever is consistent with it is already an element of it.

For this reason, in our view, it is not possible to give an even remotely credible account of the very important notions of obligation under complete or current information, without enlarging the formal language to represent the passage of time. Since the machinery of this paper abstracts from temporal considerations, we cannot here deal with those notions. In what follows we shall consider only obligation under zero information, with $O\alpha$ defined as $O(\alpha/\tau)$, whilst recognising that it appears to be infrequent in ordinary usage. We consider two principles: conditionalisation and deontic modus ponens.

Conditionalisation. This is the principle that an unconditional norm implies its conditional counterpart. For obligation it is $O\alpha \rightarrow O(\alpha/\beta)$ and for permission, $P\alpha \rightarrow P(\alpha/\beta)$. Given the definition of the unconditional operators, the status of each follows from that for strengthening the antecedent: the principle for permission fails without qualification, and that for obligation holds whenever $out(\beta) = out^*(\beta)$, indeed when $\alpha \notin out^*(\beta) - out(\beta)$.

Deontic modus ponens. This is the principle that from a conditional norm and some form of affirmation of its antecedent, we may infer its consequent. As is well known, it has two quite distinct forms: that in which the antecedent is affirmed unchanged, and that in which it is normed. In the former case, the principles are $O(\alpha/\beta)\wedge\beta \rightarrow O\alpha$ and $P(\alpha/\beta)\wedge\beta \rightarrow P\alpha$, whereas in the latter case they are $O(\alpha/\beta)\wedge O\beta \rightarrow O\alpha$ and $P(\alpha/\beta)\wedge P\beta \rightarrow P\alpha$.

In the literature these are sometimes referred to as "factual" and "deontic" modus ponens respectively. However, in both cases the first conjunct of the antecedent of the principle remains a conditional norm, so that the principle retains an essentially deontic element. For this reason we prefer to describe both as forms of deontic modus ponens, distinguishing between those "with boolean minor" and those "with normative minor".

Deontic modus ponens with boolean minor. $P(\alpha/\beta)\wedge\beta \rightarrow P\alpha$ holds – but trivially, as a consequence of weakening the antecedent which gives us $P(\alpha/\beta) \rightarrow P(\alpha/\tau)$. On the other hand, $O(\alpha/\beta)\wedge\beta \rightarrow O\alpha$ fails. We can use the same counterexample as for contraposition, adding $v(\beta) = 1$.

Deontic modus ponens with normative minor. These are the principles $O(\alpha/\beta)\wedge O\beta \rightarrow O\alpha$ and $P(\alpha/\beta)\wedge P\beta \rightarrow P\alpha$. Given the definition of the unconditional operator, the status of each follows from that for transitivity. For obligation, we have immediately from transitivity that $O(\alpha/\beta)\wedge O\beta \rightarrow O\alpha$ whenever $out(\tau) = out^*(\tau)$. The local hypothesis $\alpha \notin out^*(\tau) - out(\tau)$ also suffices. On the other hand, for permission this form of deontic detachment fails. We can use the same counterexample as for transitivity, but with $\gamma = \tau$.

7.5. Relating conditional obligation to permission

The relations between the conditional operators parallel exactly those in the unconditional case, described in section 2.4, with essentially the same verifications. The independent rules fail all of $O(\alpha/\beta) \rightarrow P(\alpha/\beta)$, $O(\alpha/\beta)\wedge P(\gamma/\beta) \rightarrow P(\alpha\wedge\gamma/\beta)$, and $P(\alpha/\beta) \rightarrow \neg O(\neg\alpha/\beta)$. The counterexamples are analogous to those in the unconditional case.

8. Open questions

If the basic concepts developed in this paper are viable, then it may be of interest to follow through on a number of questions.

Philosophical

- Are there any anomalies in our definition of net output other than those that can be ascribed to changes in time or bearer?
- Is the dependence of net output on the manner of formulation of the explicit conditional promulgations, a virtue as we have suggested, or a vice?
- Given our intuitive understanding of strong permission (section 5), should we refine further the definition of perm(D,β)?

Comparative

- It may be rewarding to compare our use of labels with that in the "labelled deontic logic" of van der Torre (1997), (van der Torre and Tan 1997), (van der Torre 1998).

Two technical differences may already be noted. The presentation in van der Torre (1997) does not consider the parallel traces needed for dealing with reasoning by cases. Van der Torre (1998) takes account of this, but favours a variant labelling policy. Roughly speaking, rather than check the consistency of the current body β with the current label, van der Torre prefers to put β into the current label and check the latter for consistency. As he observes, this has the effect of making all subsequent consistency checks include the current body. This is illustrated by his example C = $\{(\alpha,\beta), (\alpha,\neg\beta)\}$. Body disjunction gives $\alpha \in$ out(τ) with label $\{\alpha\}$ for us, and with label $\{\alpha\wedge\beta, \alpha\wedge\neg\beta\}$ for van der Torre. Body strengthening then gives $\alpha \in$ out($\beta\leftrightarrow\alpha$) with label $\{\alpha\}$ for us, and this label is consistent with the condition $\beta\leftrightarrow\alpha$. But van der Torre gets the same formula with label $\{\alpha\wedge\beta\wedge(\beta\leftrightarrow\alpha), \alpha\wedge\neg\beta\wedge(\beta\leftrightarrow\alpha)\}$, one element of which is inconsistent, rendering the step illegitimate.

- It may be instructive to consider the relationships between net output and extensions of a Reiter default system.

Evidently, three distinguishing features of net output are that it does not have $\beta \in$ out(N,β), it admits reasoning by cases, and it uses a derivation policy of "keep going until you hit trouble" rather than a fixpoint condition on the end product. In this connection, see also (Horty 1997).

- Is there a possible worlds semantics for the iterative account? In section 1 we have already discussed a natural approach to this problem, and its shortcomings. Is it possible to do better?

In the view of the author, the existence of a possible worlds semantics, or indeed of any kind of semantics, should not be treated as a touchstone. When available, a semantic characterisation can be technically useful and sometimes conceptually enlightening, but its absence is not necessarily a ground for rejection.

If the iterative approach is on the right track

- To minimize overheads, we have been working in a spartan environment: no attention to defeasible norms, instantiation and more generally quantifiers, passage of time, human agency, bearers and counterparties of obligations, distributions of "burden of proof", representation of powers and competences – and whatever other dimensions that a comprehensive approach might require.

From our discussions, it seems that the feature most urgently needing integration is the passage of time. Its intricacies should not be under-estimated. There are choices to be made on at least three levels. Two of these levels are already familiar from temporal logic itself – that of mathematical structure (e.g. whether to take time as discrete or continuous, branching or linear, whether to work with points or with intervals), and that of representation (e.g. whether to use tense operators or temporal indices, whether to operate on the latter with quantifiers or with functions).

The third level, however, is more specific to our present context, and arises from the multiplicity of temporal patterns that may be intended when uttering conditional norms. Suppose for simplicity we are working with discrete time points, and consider an explicit promulgation (α,β). This may be understood in many different ways, for example: at the present moment (or: at any time from now on) if β holds at that point (or: at some subsequent moment, or: at all subsequent moments) then at the next moment (or: some following moment, or: at all following moments, or: at all moments beyond some following one) α is required to hold. This already gives twenty-four different temporal readings! According to the choice made, we will get quite different logics. For example, under the last reading of β above, proof by cases will fail.

Notes

1 Aware of this problem, some recent accounts of conditional obligation in terms of possible worlds introduce addenda into their valuation rules for O(./.) expressly in order to destroy the validity of $O(\beta/\beta)$. However, the addenda are generally quite *ad hoc*. One exception is the possible worlds semantics sketched in section 7.2 of Makinson (1993). However, that account has other limitations. In particular, it validates strengthening of the antecedent, dealing insensitively with contrary-to-duty obligations.

2 In particular, Alchourrón and Bulygin work with a set R of "explicitly rejected" formulae rather than a set B of explicitly permitted ones, and put $v_N(P\alpha) = 1$ iff $\neg\alpha \vdash \chi$ for some $\chi \in R$. But the two are trivially equivalent, taking $B = \{\neg\chi: \chi \in R\}$ and $R = \{\neg\chi: \chi \in B\}$.

3 The approach to conditional norms in this paper is fundamentally different from that in (Hansson and Makinson 1997). That paper works in a "naïve" or "pre-critical" frame, assuming that for the purposes of logic, norms may be treated as if they had truth-values. It takes as given a consequence operation over formulae with deontic operators – whereas our purpose is to construct such a consequence operation in a manner that is consistent with the position that norms are neither true nor false. This basic difference of viewpoint gives rise to technical ones. In particular, in Hansson and Makinson (1997), the definition of a normative code allows deontic operators to appear in the heads (but not the bodies) of conditional promulgations. As a result of this asymmetry, the consequence operation may be used in place of detachment without authorizing contraposition or modus tollens.

4 There may be some connection between this distinction, and one made by van der Torre (1997, chapter 4) between "overriding" and "factual" defeasibility. There may also

be some connection between our distinction of net from gross output, and that made by (van der Torre and Tan 1998) between the contexts of "deliberation" and "justification".

Acknowledgements

Special thanks are due to Leendert van der Torre, whose thesis (1997) and insightful comments have been a source of inspiration at a number of critical junctures. Thanks are also due to Veronica Becher, Eugenio Bulygin, Dov Gabbay, Sven Ove Hansson, Lloyd Humberstone, Carlos Oller, Karl Schlechta, Marek Sergot, and an anonymous referee for their patient and helpful criticisms of drafts. Comments at ECSQARU/FAPR 97, IVR 97 and DEON 98 were also useful.

References

Alchourrón, Carlos (1993). Philosophical foundations of deontic logic and the logic of defeasible conditionals. In Meyer, J.J. and R.J. Wieringa (eds) *Deontic Logic in Computer Science*, 43-84. Wiley: New York.

Alchourrón, Carlos & Eugenio Bulygin (1981). The expressive conception of norms. In Hilpinen, R. (ed.) *New Essays in Deontic Logic*, 95-124. Reidel: Dordrecht.

Alchourrón, Carlos & Eugenio Bulygin (1984). Pragmatic foundations for a logic of norms. *Rechtstheorie* 15: 453-464.

Belzer, Marvin (1987). Legal reasoning in 3-D. In *Proceedings of the First International Conference in Artificial Intelligence and Law*, 155-163. ACM Press: Boston.

Carmo, José & Andrew J.I. Jones (1997). A new approach to contrary-to-duty obligations. In Nute, D. (ed.) *Defeasible Deontic Logic*, 317-344. Kluwer: Dordrecht.

Chisholm, R.M. (1963). Contrary-to-duty imperatives and deontic logic. *Analysis* 24: 33-36.

Dubislav, Walter (1937). Zur Unbegründbarkeit der Forderungsätze. *Theoria* 3: 330-342.

Forrester, J.W. (1984). Gentle murder, or the adverbial Samaritan. *Journal of Philosophy* 81: 193-197.

Gabbay, Dov (1996). *Labelled Deductive Systems, vol. 1*. Oxford University Press: Oxford.

Hansson, Bengt (1969) An analysis of some deontic logics. *Nous* 3: 373-398. Reprinted in Hilpinen, R. (ed.) (1971 and 1981) *Deontic Logic: Introductory and Systematic Readings*, 121-147. Reidel: Dordrecht.

Hansson, Sven Ove & David Makinson (1997). Applying normative rules with restraint. In Dalla Chiara, M.L. et al (eds) *Logic and Scientific Methods*, 313-332. Kluwer: Dordrecht.

Horty, John F. (1997). Nonmonotonic foundations for deontic logic. In Nute, D. (ed.) *Defeasible Deontic Logic*, 17-43. Kluwer: Dordrecht.

Jones, Andrew J.I. & Ingmar Pörn (1985). Ideality, sub-ideality and deontic logic. *Synthese* 65: 273-290.

Jørgensen, Jørgen (1937-8). Imperatives and logic. *Erkenntnis* 7: 288-296.

Kalinowski, Georges (1972). *La logique des normes*. Presses Universitaires de France: Paris.

Loewer, B. & M. Belzer (1983). Dyadic deontic detachment. *Synthese* 54: 295-318.

Makinson, David (1993). Five faces of minimality. *Studia Logica* 52: 339-379.

Prakken, Henry & Marek Sergot (1996). Contrary-to-duty obligations. *Studia Logica* 57: 91-115.

Prakken, Henry & Marek Sergot (1997). Dyadic deontic logic and contrary-to-duty obligations. In Nute, D. (ed.) *Defeasible Deontic Logic*, 223-262. Kluwer: Dordrecht.

Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence* 13: 81-132.

Ross, Alf (1941). Imperatives and logic. *Theoria* 7: 53-71. Reprinted in *Philosophy of Science* (1944) 11: 30-46.

Stenius, Erik (1963). Principles of a logic of normative systems. *Acta Philosophica Fennica* 16: 247-260.

van der Torre, Leendert W.N. (1997). *Reasoning about Obligations: Defeasibility in Preference-Based Deontic Logic.* Ph.D. diss., Erasmus University of Rotterdam. Tinbergen Institute Research Series n° 140. Thesis Publishers: Amsterdam.

van der Torre, Leendert W.N. (1998). Phased labeled logics of conditional goals. Manuscript to be presented at The First International Workshop on Labelled Deduction, Freiburg Germany, September 1998.

van der Torre, Leendert W.N. & Yao-Hua Tan (1997). The many faces of defeasibility in defeasible deontic logic. In Nute, D. (ed.) *Defeasible Deontic Logic*, 79-122. Kluwer: Dordrecht.

Van der Torre, Leendert W.N. & Yao-Hua Tan (1998). An update semantics for deontic reasoning. In McNamara, P. & H. Prakken (eds) *Norms, Logics and Information Systems. New Studies on Deontic Logic and Computer Science*. IOS Press: Amsterdam.

von Wright, G.H. (1951). Deontic logic. *Mind* 60: 58-74.

von Wright, G.H. (1964). A new system of deontic logic. *Danish Yearbook of Philosophy* 1: 173-182. Reprinted in Hilpinen, R. (ed.) (1971 and 1981) *Deontic Logic: Introductory and Systematic Readings*, 105-107. Reidel: Dordrecht.

von Wright, G.H. (1968). *An Essay on Deontic Logic and the General Theory of Action*. North-Holland: Amsterdam.