DEONTIC LOGIC[1]

Table of Contents[2]

---

[1] At the invitation of the editors of this series, this essay is a minor adaptation of McNamara 2005 (Fall) . Thus it is primarily systematic, but with historical information weaved in throughout, especially in notes on the literature associated with a problem or a development.

[2] Items in brackets are items boxed off in the text. They can be skipped without loss of continuity. Similarly for the four appendices to which the reader is optionally directed at appropriate places in the main essay.

<u>Introduction</u>

Deontic logic[3] is that branch of symbolic logic that has been the most concerned with the contribution that the following notions make to what follows from what:

---

[3] The term "*deontic* logic" appears to have arisen in English as the result of C. D. Broad's suggestion to von Wright (von Wright 1951); Mally used "Deontik" earlier to describe his work (Mally 1926). Both terms derive from the

permissible (permitted)	must
impermissible (forbidden, prohibited)	supererogatory (beyond the call of duty)
obligatory (duty, required)	indifferent / significant
gratuitous (non-obligatory)	the least one can do
optional	better than / best / good / bad
ought	claim / liberty / power / immunity.

To be sure, some of these notions have received more attention in deontic logic than others. However, virtually everyone working in this area would see systems designed to model the logical contributions of these notions as part of deontic logic proper.

As a branch of symbolic logic, deontic logic is of *theoretical* interest for some of the same reasons that modal logic is of theoretical interest. However, despite the fact that we need to be cautious about making too easy a link between deontic logic and practicality, many of the notions listed *are* typically employed in attempting to regulate and coordinate our lives together (but also to evaluate states of affairs). For these reasons, deontic logics often directly involve topics of considerable practical significance such as morality, law, social and business organizations (their norms, as well as their normative constitution), and security systems. To that extent, studying the logic of notions with such *practical* significance perhaps adds some practical significance to deontic logic itself.

<u>On Defining Deontic Logic</u>: Defining a discipline or area within one is often difficult. Deontic logic is no exception. Standard characterizations of deontic logic are arguably either too narrow or too wide. Deontic logic is often glossed as the logic of obligation, permission, and prohibition, but this is too narrow. For example, it would exclude a logic of supererogation as well as any non-reductive logic for legal notions like claims, liberties, powers, and immunities from falling within deontic logic. On the other hand, we might say that deontic logic is that branch of symbolic logic concerned with the logic of normative expressions: a systematic study of the contribution these expressions make to what follows from what. This is better in that it does not appear to be too exclusive, but it is arguably too broad, since deontic logic is not traditionally concerned with the contribution of every sort of normative expression. For example, "credible" and "dubious" are normative expressions, as are "rational" and "prudent" but these two pairs are not normally construed as within the purview of deontic logic (as opposed to say epistemic logic, and rational choice theory, respectively). Nor would it be enough to simply say that the normative notions of deontic logic are always practical, since the operator "it ought to be the case that", perhaps the most studied operator in deontic logic, appears to have no greater intrinsic link to practicality than does "credible" or "dubious". The following seem to be without practical import: "It ought to be the case that early humans did not exterminate Neanderthals."[4] Perhaps a more refined link to practicality is what separates deontic logic from epistemic logic, but this

Greek term, δεον, for 'that which is binding', and ικ, a common Greek adjective-forming suffix for 'after the manner of', 'of the nature of', 'pertaining to', 'of', thus suggesting roughly the idea of a logic of duty. (The intervening "τ" in "δεοντικ" is inserted for phonetic reasons.)

[4] Although this example has no practical significance for us, it is still true that without such capacities for counterfactual *evaluation*, we would have no capacity for such deeply human traits as a sense of tragedy and misfortune, and of course some judgments about what ought to be the case do and should guide our actions, but the link is not simple, and it is not clear that such evaluations of states of affairs are any less a part of deontic logic than evaluations of the future courses of action of agents.

doesn't help distinguish it from rational choice theory, the former being concerned with collective practical issues as well as individual ones. Perhaps there is no non-ad hoc or principled division between deontic logic and distinct formal disciplines focused on the logic of other normative expressions, such as epistemic logic and rational choice theory. These are interesting and largely unstudied meta-philosophical issues that we cannot settle here. Instead we have defined deontic logic contextually and provisionally.

This essay is divided into four main parts. The first provides preliminary background. The next two parts provide an introduction to the most standard monadic systems of deontic logic The fourth, and by far the largest, section is dedicated to various problems and challenges faced by the standard systems. This reflects the fact that the challenges posed to these standard systems are numerous.

1. Informal Preliminaries and Background

Deontic logic has been strongly influenced by ideas in modal logic. Analogies with alethic modal notions and deontic notions were noticed as far back as the fourteenth century, where we might say that the rudiments of modern deontic logic began (Knuuttila 1981). Although informal interest in what can be arguably called aspects of deontic logic continued, the trend toward studying logic using the *symbolic and exact* techniques of mathematics became dominant in the twentieth century, and logic became largely, *symbolic* logic. Work in twentieth century symbolic modal logic provided the explicit impetus for von Wright (von Wright 1951), the central early figure in the emergence of deontic logic as a full-fledged branch of *symbolic* logic in the twentieth century. So we will begin by gently noting a few folk-logical features of alethic modal notions, and give an impressionistic sense of how natural it was for early developments of deontic logic to mimic those of modal logic. We will then turn to a more direct exploration of deontic logic as a branch of symbolic logic.

However, before turning to von Wright, and the launching of deontic logic as an on-going active academic area of study, we need to note that there was a significant earlier episode, Mally 1926 that did not have the influence on symbolic deontic logic that it might have, due at least in part, to serious technical problems. The most notable of these problems was the provable equivalence of what ought to be the case (his main deontic notion) with what is the case, which is plainly self-defeating for a deontic logic. Despite the problems with the system he found, Mally was an impressive pioneer of deontic logic. He was apparently uninfluenced by, and thus did not benefit from, early developments of alethic modal logic. This is quite opposed to the later trend in the 1950s when deontic logic reemerged, this time as a full-fledged discipline, deeply influenced by earlier developments in alethic modal logic. Mally was the first to found deontic logic on the syntax of propositional calculus explicitly, a strategy that others quickly returned to after a deviation from this strategy in the very first work of von Wright. Mally was the first to employ deontic constants in deontic logic (reminiscent of Kanger and Anderson's later use of deontic constants, but without their "reduction"; more below). He was also the first to attempt to provide an integrated account of non-conditional and conditional ought statements, one that provided an *analysis* of conditional ‘ought’s via a monadic deontic operator coupled with a material conditional (reminiscent of similar failed attempts in von Wright 1951 to analyze the dyadic notion of commitment), and that allowed for a form of factual detachment (more below).

All in all, this seems to be a remarkable achievement in retrospect. For more information on Mally's system, including a diagnosis of the source of his main technical problem, and a sketch of one way he might have avoided it, see the easily accessible Lokhorst 2004.

1.1 Some Informal Rudiments of Alethic Modal Logic

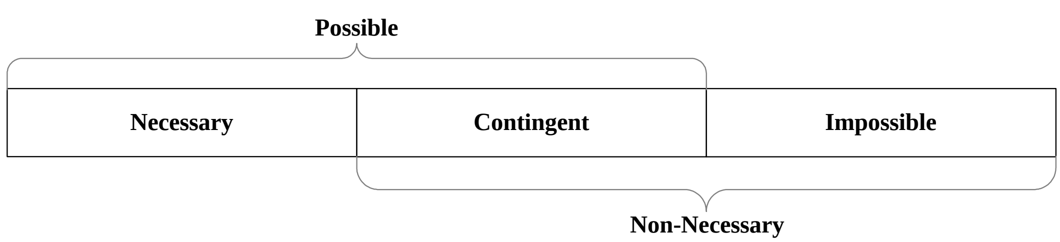
Alethic modal logic is roughly the logic of necessary truth and related notions. Consider five basic alethic modal statuses, expressed as sentential operators--constructions that, when applied to a sentence, yield a sentence (as does "it is not the case that"):

it is necessary (necessarily true) that ($\Box$)
it is possible that ($\Diamond$)
it is impossible that
it is non-necessary that
it is contingent that[5]

Although all of the above operators are generally deemed definable in terms of any one of the first four, the necessity operator is typically taken as basic and the rest defined accordingly:

It is possible that p ($\Diamond p$) $=_{df} \sim\Box\sim p$
It is impossible that $=_{df} \Box\sim p$
It is non-necessary that $=_{df} \sim\Box p$
It is contingent that $=_{df} \sim\Box p \ \& \sim\Box\sim p$

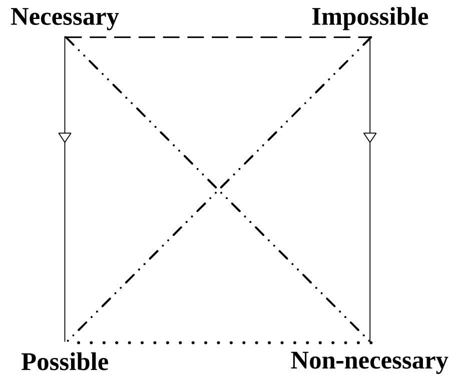
It is routinely assumed that the following threefold partition of propositions holds:



The three rectangular cells are jointly exhaustive and mutually exclusive: every proposition is either necessary, contingent, or impossible, but no proposition is more than one of these. The possible propositions are those that are either necessary or contingent, and the non-necessary propositions are those that are either impossible or contingent.

Another piece of folk logic for these notions is the following modal square of opposition:

[5] In keeping with very wide trends in logic over the past century or so, we will treat both modal notions and deontic notions as sentential (or propositional) operators unless otherwise stated. Although it is controversial whether the most fundamental (if there are such) modal and deontic notions have the logical form of propositional operators, focusing on these forms allowed for essentially seamless integration of these logics with propositional logics.



<u>Arrowed Lines</u>: represent implications
<u>Dashed Line</u>: connects contraries.
<u>Dotted Line</u>: connects sub-contraries.
<u>Dotted-Dashed Lines</u>: connect contradictories. [6]

Furthermore it is generally assumed that the following hold:

> If $\Box$p then p (if it is necessary that p, then p is true).
> If p, then $\Diamond$p (if p is true, then p is possible).

These reflect the idea that we are interested here in alethic (and thus truth-implicating) necessity and its siblings.

We now turn to some of the analogies involved in what is a corresponding bit of deontic folk logic: "The Traditional Scheme" (McNamara 1990, 1996a). This is a minor elaboration of what can be found in von Wright 1953 and Prior 1962 [1955].

<u>1.2 The Traditional Scheme and the Modal Analogies</u>

The five normative statuses of the Traditional Scheme are:[7]

it is obligatory that (**OB)**
it is permissible that (**PE**)

---

[6] This key will be relied on throughout for similar diagrams. Recall that propositions are *contraries* if they can't both be true, *sub-contraries* if they can't both be false, and *contradictories* if they always have opposing truth-values. The square can be easily augmented as a hexagon by including nodes for contingency (McNamara 1996a). Cf. the deontic hexagon below.

[7] Only deontic operators will appear in boldface. These abbreviations are not standard. "O" is routinely used instead of "**OB**", and "O" is often read as "It ought to be the case that". "P" instead of "**PE**", and if used at all, "F" (for "forbidden") instead of "IM" and "I" (for "indifference") instead of "**OP**". Deontic non-necessity, here denoted by "**GR**" is seldom ever named, and even in English it is hard to find a term for this condition. The double letter choices used here are easy mnemonics expressing *all five* basic conditions (which, from a logical standpoint, are on a par), and they will facilitate later discussion involving just what notions to take SDL and kin to be modeling, and how it might be enriched to handle other related normative notions. Both deontic logic and ethical theory is fraught with difficulties when it comes to interchanging allegedly equivalent expressions for one another. Here we choose to read the basic operator as "it is obligatory that" so that all continuity with permissibility, impermissibility, and indifference is not lost, as it would be with the "it ought to be the case that" reading (McNamara 1996c). A choice must be made. "It is obligatory that" may also be read personally, but non-agentially as "it is obligatory for Jones that" (Krogh and Herrestad 1996, McNamara 2004a) We will return to these issues again below.

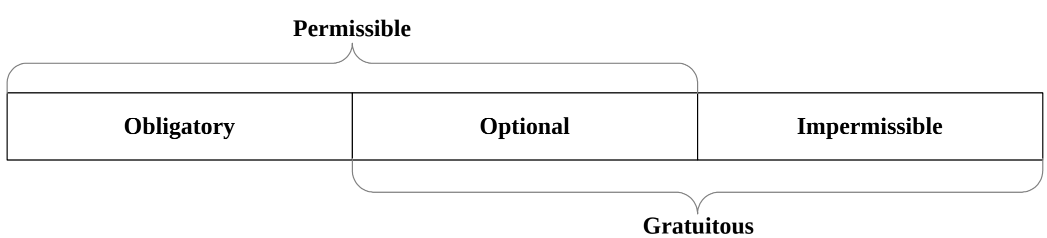
it is impermissible that (**IM**)
it is gratuitous that (**GR**)
it is optional that (**OP**)

The first three are familiar, but the fourth is widely ignored, and the fifth has regularly been conflated with "it is a matter of *indifference* that p" (by being defined in terms of one of the first three), which is not really part of the traditional scheme (more below). Typically, one of the first two is taken as basic, and the others defined in terms of it, but any of the first four can play the same sort of purported defining role. The most prevalent approach is to take the first as basic, and define the rest as follows:

$$\mathbf{PE}p \leftrightarrow {\sim}\mathbf{OB}{\sim}p$$
$$\mathbf{IM}p \leftrightarrow \mathbf{OB}{\sim}p$$
$$\mathbf{GR}p \leftrightarrow {\sim}\mathbf{OB}p$$
$$\mathbf{OP}p \leftrightarrow ({\sim}\mathbf{OB}p \;\&\; {\sim}\mathbf{OB}{\sim}p).$$ [8]

These assert that something is permissible iff (if and only if) its negation is not obligatory, impermissible iff its negation is obligatory, gratuitous iff it is not obligatory, and optional iff neither it nor its negation is obligatory. Call this "The Traditional *Definitional* Scheme (TDS)". If one began with **OB** alone and considered the formulas on the right of the equivalences above, one could easily be led to consider them as at least candidate defining conditions for those on the left. Although not uncontestable, they are natural, and this scheme is still widely employed. Now if the reader looks back at our use of the necessity operator in defining the remaining four alethic modal operators, it will be clear that that definitional scheme is perfectly analogous to the deontic one above. From the formal standpoint, the one is merely a syntactic variant of the other: just replace **OB** with $\Box$, **PE** with $\Diamond$, etc.

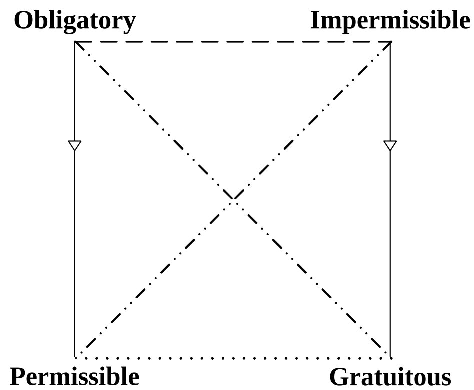
In addition to the TDS, it was traditionally assumed that the following, call it "The Traditional Threefold Classification (TTC)" holds:



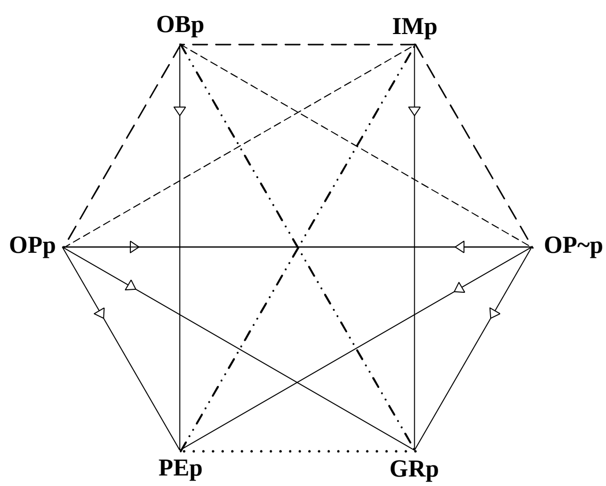
Here too, all propositions are divided into three jointly exhaustive and mutually exclusive deontic classes: every proposition is obligatory, optional, or impermissible, but no proposition falls into more than one of these three categories. Furthermore, the permissible propositions are those that are either obligatory or optional, and the gratuitous propositions are those that are impermissible or optional. The reader can easily confirm that this natural scheme is also perfectly analogous to the threefold classification we gave above for the alethic modal notions.

[8] In this essay we will generally call such equivalences "definitions", sloughing over the distinction between abbreviatory definitions of operators not officially in the formal language, and axiom systems with languages containing these operators, and axioms directly encoding the force of such definitions as equivalences.

Furthermore, "The Deontic Square" (DS)" is part of the Traditional Scheme:



The logical operators at the corners are to be interpreted as in the modal square of opposition. The two squares are plainly perfectly analogous as well. If we weave in nodes for optionality, and shift to formuli, we get a deontic hexagon:



Given these correspondences, it is unsurprising that our basic operator, read here as "it is obligatory that", is often referred to as "deontic *necessity*". However, there are also obvious dis-analogies. Before, we saw that these two principles are part of the traditional conception of alethic modality:

If $\Box$p then p (if it is necessary that p, then p is true).
If p, then $\Diamond$p (if p is true, then it is possible).

But their deontic analogs are:

If **OB**p then p (if it is obligatory that p, then p is true).
If p, then **PE**p (if p is true, then it is permissible).

The latter two are transparently false, for obligations can be violated, and impermissible things do happen.[9] However, as researchers turned to generalizations of alethic modal logic, they began considering wider classes of modal logics, including ones where the necessity operator was not truth-implicating. This too encouraged seeing deontic necessity, and thus deontic logic, as falling within modal logic so-generalized, and in fact recognizing possibilities like this helped to fuel the generalizations of what began with a focus on alethic modal logic (Lemmon 1957, Lemmon and Scott 1977).

1.3 Toward Deontic Logic Proper

It will be convenient at this point to introduce a bit more regimentation. Let's assume that we have a simple propositional language with the usual suspects, an infinite set of propositional variables (say, $P_1,\ldots,P_n,\ldots$) and complete set of truth-functional operators (say, ~ and $\rightarrow$), as well as the one-place deontic operator, **OB**.

Deontic Wffs: Here is a more formal definition. Suppose that we have:

A set of Propositional Variables (PV): P_1, ..., P_i, ... -- where "i" is a numerical subscript; three propositional operators: ~, $\rightarrow$, **OB**; and a pair of parentheses: (,).

The set of D-wffs (deontic well-formed formuli) is then the smallest set satisfying the following conditions (lower case "p" and "q" are metavariables):

FR1. PV is a subset of D-wffs.
FR2. For any p, p is in D-wffs only if ~p and **OB**p are also in D-wffs.
FR3. For any p and q, p and q are in D-wffs only if (p $\rightarrow$ q) is in D-wffs.

We then assume the following abbreviatory definitions:

DF1-3. &, $\vee$, $\rightarrow$ as usual.
DF4. **PE**p $=_{df}$ ~**OB**~p.
DF5. **IM**p $=_{df}$ **OB**~p.
DF6. **GR**p $=_{df}$ ~**OB**p.
DF7. **OP** $=_{df}$ (~**OB**p & ~**OB**~p).

Unless otherwise stated, we will only be interested in deontic logics that contain classical propositional calculus (PC). So let's assume we add that as the first ingredient in specifying any deontic logic, so that, for example, **OB**p $\rightarrow$ ~~**OB**p, can be derived in any system to be considered here.

Above, in identifying the Traditional Definitional Scheme, we noted that we could have taken any of the first four of the five primary normative statuses listed as basic and defined the rest in terms of that one. So we want to be able to generate the corresponding equivalences derivatively

[9] The logic of Mally 1926 was saddled with the T-analog above. Mally reluctantly embraced it since it seemed to follow from premises he could find no fault with. See Lokhorst 2004.

from the scheme we did settle on, where **OB** is basic. But thus far we cannot. For example, it is obviously desirable to have **OB**p → ~**PE**~p as a theorem from the traditional standpoint. After all, this wff merely expresses one half of the equivalence between what would have been definiens and definiendum had we chosen the alternate scheme of definition in which "**PE**" was taken as basic instead of "**OB**". However, **OB**p → ~**PE**~p is not thus far derivable. For **OB**p → ~**PE**~p is definitionally equivalent to **OB**p → ~~**OB**~~p, which reduces by PC to **OB**p → **OB**~~p, but the latter formula is not tautological, so we cannot complete the proof. So far we have deontic wffs and propositional logic, but no deontic logic. For that we need some distinctive principles governing our deontic operator, and in particular, to generate the alternative equivalences that reflect the alternative definitional schemes alluded to above, we need what is perhaps the most fundamental and least controversial rule of inference in deontic logic, and the one characteristic of "classical modal logics" (Chellas 1980):

OB-RE: If p ↔ q is a theorem, then so is **OB**p ↔ **OB**q.

This rule tells us that if two formulas are provably equivalent, then so are the results of prefacing them with our basic operator, **OB**. With its aid (and the Traditional Definitional Scheme's), it now easy to prove the equivalences corresponding to the alternative definitional schemes. For example, since ⊢ p ↔ ~~p, by **OB**-RE, we get ⊢ **OB**p ↔ **OB~**~p, i.e. ⊢ **OB**p ↔ ~~**OB~**~p, which generates ⊢ **OB**p ↔ ~**PE**~p, given our definitional scheme. To the extent that the alternative definitional equivalences are supposed to be derivable, we can see RE as presupposed in the Traditional Scheme.

All systems we consider here will contain RE (whether as basic or derived). They will also contain (unless stated otherwise) one other principle, a thesis asserting that a logical contradiction (conventionally denoted by "⊥") is always gratuitous:

OD: ~**OB**⊥

So, for example, OD implies that it is a logical truth that it is not obligatory that my taxes are paid and not paid. Although OD is not completely uncontestable[10], it is plausible, and like RE, has been pervasively presupposed in work on deontic logic. In this essay, we will focus on systems that endorse both RE and OD.

Before turning to our first full-fledged system of deontic logic, let us note one very important principle that is not contained in all deontic logics, and about which a great deal of controversy in deontic logic and in ethical theory has transpired.

1.4 The Fundamental Presupposition of the Traditional Scheme:

Returning to the Traditional Scheme for a moment, its Threefold Classification, and Deontic Square of Opposition can be expressed formally as follows:

[10] If Romeo solemnly promised Juliet to square the circle did it thereby become obligatory that he do so?

DS: $(\mathbf{OB}p \leftrightarrow \sim\mathbf{GR}p) \ \& \ (\mathbf{IM}p \leftrightarrow \sim\mathbf{PE}p) \ \& \ \sim(\mathbf{OB}p \ \& \ \mathbf{IM}p) \ \&$
$\sim(\sim\mathbf{PE}p \ \& \ \sim\mathbf{GR}p) \ \& \ (\mathbf{OB}p \rightarrow \mathbf{PE}p) \ \& \ (\mathbf{IM}p \rightarrow \mathbf{GR}p).$

TTC: $(\mathbf{OB}p \vee \mathbf{OP}p \vee \mathbf{IM}p) \ \& \ [\sim(\mathbf{OB}p \ \& \ \mathbf{IM}p) \ \& \ \sim(\mathbf{OB}p \ \& \ \mathbf{OP}p) \ \&$
$\sim(\mathbf{OP}p \ \& \ \mathbf{IM}p)].$

Given the Traditional Definitional Scheme, it turns out that DS and TTC are each *tautologically* equivalent to the principle that obligations cannot conflict (and thus to one another):

NC: $\sim(\mathbf{OB}p \ \& \ \mathbf{OB}\sim p)$.[11]

So the Traditional Scheme rests squarely on the soundness of NC (and the traditional definitions of the operators). Indeed, the Traditional Scheme is nothing other than a disguised version of NC, given the definitional component of that scheme.

NC is not to be confused in content with the previously mentioned principle, OD ($\sim\mathbf{OB}\bot$). OD asserts that no single logical contradiction can be obligatory, whereas NC asserts that there can never be two things that are each separately obligatory, where the one obligatory thing is the negation of the other. The presence or absence of NC arguably represents one of the most fundamental divisions among deontic schemes. As until recently in modern normative ethics (see Gowans 1987), early deontic logics presupposed this thesis. Before turning to challenges to NC, we will consider a number of systems that endorse it, beginning with what has come to be routinely called "Standard Deontic Logic", the benchmark system of deontic logic.

2. Standard Deontic Logic

2.1 Standard Syntax

Standard Deontic Logic (SDL) is the most cited and studied system of deontic logic, and one of the first deontic logics axiomatically specified. It builds upon propositional logic, and is in fact essentially just a distinguished member of the most studied class of modal logics, "normal modal logics". It is a monadic deontic logic, since its basic deontic operator is a one-place operator (like ~, and unlike →): syntactically, it applies to a *single* sentence to yield a compound sentence.[12]

Assume again that we have a language of classical propositional logic with an infinite set of propositional variables, the operators ~ and →, and the operator, **OB**. SDL is then often axiomatized as follows:

[11] For DS becomes $(\mathbf{OB}p \leftrightarrow \sim\sim\mathbf{OB}p) \ \& \ (\mathbf{OB}\sim p \leftrightarrow \sim\sim\mathbf{OB}\sim p) \ \& \ \sim(\mathbf{OB}p \ \& \ \mathbf{OB}\sim p) \ \& \ \sim(\sim\sim\mathbf{OB}\sim p \ \& \ \sim\sim\mathbf{OB}p) \ \&$ $(\mathbf{OB}p \rightarrow \sim\mathbf{OB}\sim p) \ \& \ (\mathbf{OB}\sim p \rightarrow \sim\mathbf{OB}p)$, and although the first two conjuncts are tautologies, the remaining four are each tautologically equivalent to NC above. Similarly, TTC becomes $(\mathbf{OB}p \vee (\sim\mathbf{OB}p \ \& \ \sim\mathbf{OB}\sim p) \vee \mathbf{OB}\sim p) \ \&$ $[\sim(\mathbf{OB}p \ \& \ \mathbf{OB}\sim p) \ \& \ \sim(\mathbf{OB}p \ \& \ (\sim\mathbf{OB}p \ \& \ \sim\mathbf{OB}\sim p)) \ \& \ \sim((\sim\mathbf{OB}p \ \& \ \sim\mathbf{OB}\sim p) \ \& \ \mathbf{OB}\sim p)]$, and the exhaustiveness clause is tautological, as are the last two conjuncts of the exclusiveness clause, but the first conjunct of that clause is just NC again. Likewise for the assumptions that *the gratuitous* is the disjunction of *the permissible* and *the obligatory* and that *the permissible* is the disjunction of *the obligatory* and *the optional*. (See McNamara 1996a, 42-46.)

[12] In a monadic system one can easily *define* dyadic deontic operators *of sorts* (Hintikka 1971). For example, we might define "deontic implication" as follows: $p \ {}_{d}\!\!\rightarrow q =_{df} \mathbf{OB}(p \rightarrow q)$. We will consider non-monadic systems later on.

<u>SDL</u>:	A1.	All tautologous wffs of the language	(TAUT)
	A2.	$\mathbf{OB}(p \to q) \to (\mathbf{OB}p \to \mathbf{OB}q)$	(**OB**-K)
	A3.	$\mathbf{OB}p \to \sim\mathbf{OB}\sim p$	(**OB**-D)
	MP.	If $\vdash p$ and $\vdash p \to q$ then $\vdash q$	(MP)
	R2.	If $\vdash p$ then $\vdash \mathbf{OB}p$	(**OB**-NEC)[13]

SDL is just the normal modal logic "D" or "KD", with a suggestive notation expressing the intended interpretation.[14] TAUT is standard for normal modal systems. **OB**-K, which is the K axiom present in all normal modal logics, tells us that if a material conditional is obligatory, and its antecedent is obligatory, then so is its consequent.[15] **OB**-D tells us that p is obligatory only if its negation isn't. It is just "No Conflicts" again, but it is also called "D" (for "Deontic") in normal modal logics. MP is just Modus Ponens, telling us that if a material conditional and its antecedent are theorems, then so is the consequent. TAUT combined with MP gives us the full inferential power of the Propositional Calculus (often referred to, including here, as "PC"). As noted earlier, PC has no distinctive deontic import. **OB**-NEC tells us that if anything is a theorem, then the claim that that thing is obligatory is also a theorem. Note that this guarantees that something is always obligatory (even if only logical truths).[16] Each of the distinctively deontic principles, **OB**-K, **OB**-D, and **OB**-NEC are contestable, and we will consider criticisms of them shortly. However, to avoid immediate confusion for those new to deontic logic, it is perhaps worth noting that **OB**-NEC is generally deemed a convenience that, among other things, assures that SDL is in fact just one of the well-studied normal modal logics with a deontic interpretation. Few have spilled blood to defend its cogency substantively, and these practical compromises can be strategic, especially in early stages of research.

Regarding SDL's expressive powers, advocates typically endorse the Traditional Definitional Scheme noted earlier. Below we list some theorems and two important derived rules of SDL.[17]

$\mathbf{OB}\top$	(**OB**-N)
$\sim\mathbf{OB}\bot$[18]	(**OB**-OD)
$\mathbf{OB}(p \,\&\, q) \to (\mathbf{OB}p \,\&\, \mathbf{OB}q)$	(**OB**-M)
$(\mathbf{OB}p \,\&\, \mathbf{OB}q) \to \mathbf{OB}(p \,\&\, q)$	(**OB**-C / Aggregation)
$\mathbf{OB}p \vee \mathbf{OP}p \vee \mathbf{IM}p$	(**OB**-Exhaustion)
$\mathbf{OB}p \to \sim\mathbf{OB}\sim p$	(**OB**-NC or **OB**-D)
If $\vdash p \to q$ then $\vdash \mathbf{OB}p \to \mathbf{OB}q$	(**OB**-RM)
If $\vdash p \leftrightarrow q$ then $\vdash \mathbf{OB}p \leftrightarrow \mathbf{OB}q$	(**OB**-RE)

[13] "$\vdash$" before a formula indicates it is a theorem of the relevant system.

[14] Note that this axiomatization, and all others here, use "axiom schema": schematic specifications by syntactic pattern of classes of axioms (rather than particular axioms generalized via a substitution rule). We will nonetheless slough over the distinction here.

[15] It is also justifies a version of *Deontic Detachment*, from **OB**p and $\mathbf{OB}(p \to q)$ derive **OB**q, an inference pattern to be discussed later.

[16] Compare the rule that contradictions are not permissible: if $\vdash \sim p$ then $\vdash \sim\mathbf{PE}p$. R2 is often said to be equivalent to "not everything is permissible", and thus to rule out only "normative systems" that have no normative force at all.

[17] We ignore most of the simple definitional equivalences mentioned above, as well as DS and TTC.

[18] Compare **OB**-N and **OB**-D with $\mathbf{OB}(p \vee \sim p)$ and $\sim\mathbf{OB}(p \,\&\, \sim p)$, respectively.

We will be discussing a number of these subsequently. For now, let's briefly show that RM is a derived rule of SDL. We note some simple corollaries as well.

Show: If $\vdash p \rightarrow q$, then $\vdash$ **OB**p $\rightarrow$ **OB**q. (**OB**-RM)
Proof: Suppose $\vdash p \rightarrow q$. Then by **OB**-NEC, $\vdash$ **OB** $(p \rightarrow q)$, and then by K, $\vdash$ **OB**p $\rightarrow$ **OB**q.
Corollary 1: $\vdash$ **OB**p $\rightarrow$ **OB**$(p \vee q)$ (Weakening)
Corollary 2: If $\vdash p \leftrightarrow q$ then $\vdash$ **OB**p $\leftrightarrow$ **OB**q (**OB**-RE)[19]

Although the above axiomatization is standard, alternative axiomatizations do have certain advantages. One such axiomatization is given in Appendix A2 and shown to be equivalent to the one above.

von Wright's 1951 System and SDL: A quick comparison of SDL with the famous system in the seminal piece von Wright 1951 is in order. It is fair to say that von Wright 1951 launched deontic logic as an area of active research. There was a flurry of responses, and not a year has gone by since without published work in this area. von Wright's 1951 system is an important predecessor of SDL, but the variables there ranged over *act types* not propositions. As a result, the deontic operator symbols (e.g. **OB**) were interpreted as applying not to sentences, but to names of act types (cf. "to attend" or "attending") to yield a sentence (e.g. "it is obligatory to attend" or "attending is obligatory"). So iterated deontic sequences (e.g. **OBOB**A) were not well-formed formulas and shouldn't have been on his intended interpretation, since **OB**A (unlike A) is a sentence, not an act description, so not suitable for having **OB** as a preface to it (cf. "it is obligatory it is obligatory to run" or "running is obligatory is obligatory"). However, von Wright does think that there can be negations, disjunctions and conjunctions of act types, and so he uses standard connectives to generate not only complex normative sentences (e.g. **OB**A & **PE**A), but complex act descriptions (e.g. A & ~B), and thus complex normative sentences involving them (e.g. **OB**(A & ~B) $\rightarrow$ **PE**(A & ~B)). The standard connectives of PC are thus used in a systematically ambiguous way in von Wright's initial system with the hope of no confusion, but a more refined approach (as he recognized) would call for the usual truth-functional operators and a second set of act-type-compounding analogues to these.[20] Mixed formulas (e.g. A $\rightarrow$ **OB**A) were not well-formed in his 1951 system and shouldn't have been on his intended interpretation, since if **OB**A is well-formed, then A must be a name of an act type not a sentence, but then it can't suitably be a preface to $\rightarrow$, when the latter is followed by an item of the sentence category (e.g. **OB**A). (Cf. "If to run then it is obligatory to run".) However, this also means that the standard violation condition for an obligation (e.g. **OB**p & ~p) is not expressible in his system. von Wright also rejected NEC, but otherwise accepts analogues to the basic principles of SDL.

Researchers quickly opted for a syntactic approach where the variables and operators are interpreted propositionally as they are in PC (Prior 1962 [1955], Anderson 1956, Kanger 1971 [1957], and Hintikka 1957), and von Wright soon adopted this course himself in his key early

[19] RE is the fundamental rule for "Classical Systems of modal logic", a class that includes normal modal logics as a proper subset. See Chellas 1980.
[20] Cf. the deontic logic in Meyer 1988, where a set of operators for action (drawn from dynamic logic) are used along with a separate set of propositional operators.

revisions of his "old system" (e.g. von Wright 1968, 1971 (originally published in 1964 and 1965). Note that this is essentially a return to the approach in Mally's deontic logic of a few decades before.

SDL can be strengthened in various ways, in particular, we might consider adding axioms where deontic operators are embedded within one another. For example, suppose we added the following formula as an axiom to SDL. Call the result "SDL+" for easy reference here:

A4. $\mathbf{OB}(\mathbf{OB}p \rightarrow p)$

This says (roughly) that it is required that obligations are fulfilled.[21] This is not a theorem of SDL (as we will see in the next section), so SDL+ is a genuine strengthening of SDL. Furthermore, it makes a logically *contingent* proposition (i.e. that $\mathbf{OB}p \rightarrow p$) obligatory *as a matter of deontic logic*. SDL does not have this substantive feature. With this addition to SDL, it is easy to prove $\mathbf{OBOB}p \rightarrow \mathbf{OB}p$, a formula involving an iterated occurrence of our main operator.[22] This formula asserts that if it is obligatory that p be obligatory, then p *is* obligatory. (Cf. "the only things that are required to be obligatory are those that actually are").[23]

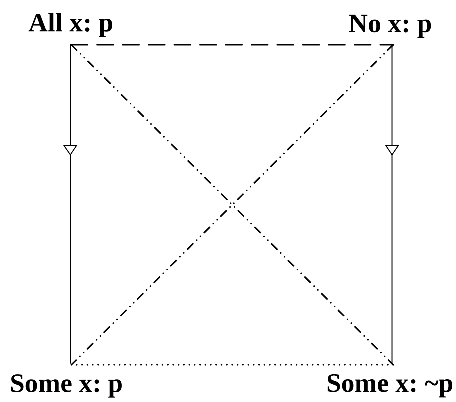
2.2 Standard Semantics

The reader familiar with elementary textbook logic will have perhaps noticed that the deontic square and the modal square both have even better-known analogs for the quantifiers as interpreted in classical predicate logic ("all x: p" is read as all objects x satisfy condition p; similarly for "no x: p" and "some x: p"):

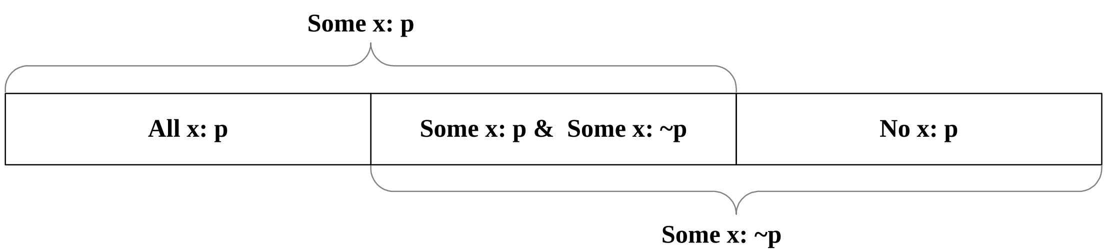
[21] Equivalently, $\mathbf{OB}(p \rightarrow \mathbf{PE}p)$, it is required that only permissible things are true.

[22] For $\mathbf{OB}(\mathbf{OB}p \rightarrow p) \rightarrow (\mathbf{OBOB}p \rightarrow \mathbf{OB}p)$ is just a special instance of **OB**-K. So using A4 above, and MP, we get $\mathbf{OBOB}p \rightarrow \mathbf{OB}p$ directly.

[23] See Chellas 1980, 193-194 for a concise critical discussion of the comparative plausibility of these two formula. (Note that Chellas' rich chapters on deontic logic in this exceptional textbook are gems generally.) However, where Chellas states that if there are any unfulfilled obligations (i.e. **OB**p and ~p both hold), then "ours in one of the worst of all possible worlds", this is misleading, since the semantics does not rank worlds other than to sort them into acceptable and unacceptable ones (relative to a world). The illuminating underlying point is that for any world j whose alternatives are all p-worlds, but where p is false, it follows that not only can't j be an acceptable alternative to itself, but it can't be an acceptable alternative to any other world, i, either. Put simply, A4 implies that any (**OB**p & ~p)-world is *universally* unacceptable. However, though indeed significant, this does not express a *degree or extent* of badness: given some ranking principle allowing for indefinitely better and worse worlds relative to some world i (such as in preference semantics for dyadic versions of SDL and kin—see below), j might be among the absolute best of the i-unacceptable worlds (i.e. ranked second only to those that are simply i-acceptable through and through), for all A4 implies.



Though less widely noted in textbooks, there is also a threefold classification for classical quantifiers:



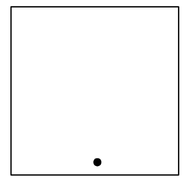
Here all conditions are divided into three jointly exhaustive and mutually exclusive classes: those that hold for all objects, those that hold for none, and those that hold for some and not for others, where no condition falls into more than one of these three categories. These deep quantificational analogies reflect much of the inspiration behind what is most often called "possible worlds semantics" for such logics, to which we now turn.[24] Once the analogies are noticed, this sort of semantics seems all but inevitable.

We now give a standard "Kripke-style" possible world semantics for SDL. Informally, we assume that we have a set of possible worlds, W, and a relation, *A*, relating worlds to worlds, with the intention that *A*ij iff j is a world where everything obligatory in i holds (i.e. no violations of the obligations holding in i occur in j). For brevity, we will call all such worlds so related to i, "i-*A*cceptable" worlds and denote them by A^i.[25] We then add that the acceptability relation is "serial": for every world, i, there is at least one i-acceptable world. Finally, propositions are either true or false at a world, never both, and when a proposition, p, is true at a world, we will often indicate this by referring to that world as a "p-world". The truth-functional operators have their usual behavior at each world. Our focus will be on the contribution deontic operators are taken to make.

The fundamental idea here is that the normative status of a proposition from the standpoint of a world i can be assessed by looking at how that proposition fairs at the i-acceptable worlds.

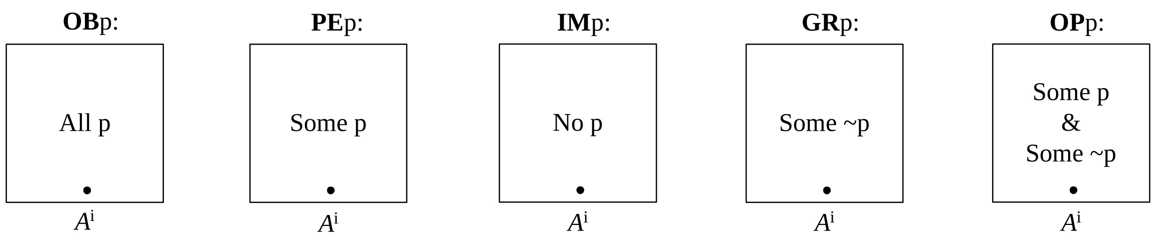
[24] von Wright 1953 and Prior 1962 [1955] (already noted in the 1st ed., 1955).
[25] The worlds related to i by *A* are also often called "ideal worlds". This language is not innocent (McNamara 1996c).

Let’s see how. For any given world, i, we can easily picture the i-accessible worlds as all corralled together in logical space as follows (where seriality is reflected by a small dot representing the presence of at least one world):



A^i

The intended truth-conditions, relative to i, for our five deontic operators can now be pictured as follows:



Thus, p is *obligatory* iff it holds in all the i-acceptable worlds, *permissible* iff it holds in some such world, *impermissible* iff it holds in no such world, *gratuitous* iff its negation holds in some such world, and *optional* iff p holds in some such world, and so does ~p. When a formula must be true at any world in any such model of serially-related worlds, then the formula is *valid*.

Kripke-Style Semantics for SDL: A more formal characterization of this semantic framework follows

We define the frames (structures) for modeling SDL as follows:

F is an Kripke-SDL (or KD) Frame: *F* = <W,A> such that:

1) W is a non-empty set
2) A is a subset of W x W
3) A is serial: $\forall i \exists j Aij$.

A model can be defined in the usual way, allowing us to then define truth at a world in a model for all sentences of SDL (and SDL+):

M is an Kripke-SDL Model: *M* = <*F*,V>, where *F* is an SDL Frame, <W,A>, and V is an assignment on *F*: V is a function from the propositional variables to various subsets of W (the " truth sets’ for the variables—the worlds where the variables are true for this assignment).

Let "$M \models_i$ p" denote p’s truth at a world, i, in a model, *M*.
Basic Truth-Conditions at a world, i, in a Model, *M*:

[PC]: (Standard Clauses for the operators of Propositional Logic.)
[**OB**]: $M \models$i **OB**p: ∀j[if Aij then $M \models$j p]

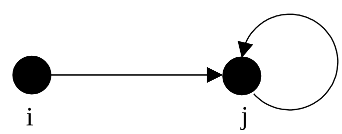
Derivative Truth-Conditions:
[**PE**]: $M \models$i **PE**p: ∃j(Aij & $M \models$j p)
[**IM**]: $M \models$i **IM**p: ~∃j(Aij & $M \models$j p)
[**GR**]: $M \models$i **GR**p: ∃j(Aij & $M \models$j ~p)
[**OP**]: $M \models$i **OP**p: ∃j(Aij & $M \models$j p) & ∃j(Aij & $M \models$j ~p)

p is true in the model, M ($M \models$ p): p is true at every world in M.
p is valid ($\models$ p): p is true in every model.

Metatheorem: SDL is sound and complete for the class of all Kripke-SDL models.[26]

To illustrate the workings of this framework, consider NC (**OB**-D), **OB**p → ~**OB**~p. This is valid in this framework. For suppose that **OB**p holds at any world i in any model. Then each i-accessible world is one where p holds, and by the seriality of accessibility, there must be at least one such world. Call it j. Now we can see that ~**OB**~p must hold at i as well, for otherwise, **OB**~p would hold at i, in which case, ~p would have to hold at all the i-accessible worlds, including j. But then p as well as ~p would hold at j itself, which is impossible (by the semantics for "~"). The other axioms and rules of SDL can be similarly shown to be valid, as can all the principles listed above as derivable in SDL

However, A4, the axiom we added to SDL above to get SDL+, is not valid in the standard serial models. In order to validate A4, **OB**(**OB**p → p), we need the further requirement of "secondary seriality": that any i-acceptable world, j, must be in turn acceptable to itself. We can illustrate such an i and j as follows:

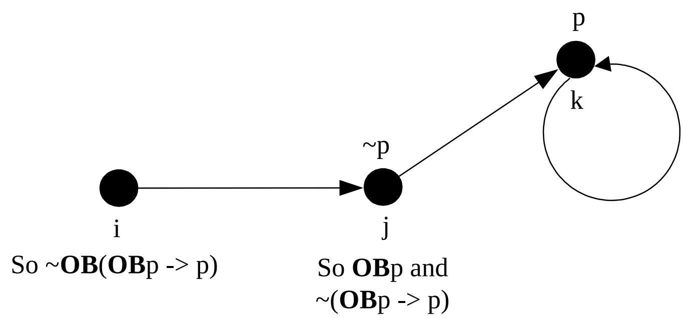


Here we imagine that the arrow connectors indicate relative acceptability, thus here, j (and only j) is acceptable to i, and j (and only j) is acceptable to j. If all worlds that are acceptable to any given world have this property of self-acceptability, then our axiom is valid. For suppose this property holds throughout our models, and that for some arbitrary world i, **OB**(**OB**p → p) is false at i. Then not all i-acceptable worlds are worlds where **OB**p → p is true. So, there must be an i-acceptable world, say j, where **OB**p is true, but p is false. Since **OB**p is true at j, then p must be true at all j-acceptable worlds. But by stipulation, j is acceptable to itself, so p must be true at j, but this contradicts our assumption that p was false at j. Thus **OB**(**OB**p → p) must be true at all worlds, after all.

[26] That is, any theorem of SDL is valid per this semantics (soundness), and any formula valid per this semantics is a theorem of SDL (completeness)

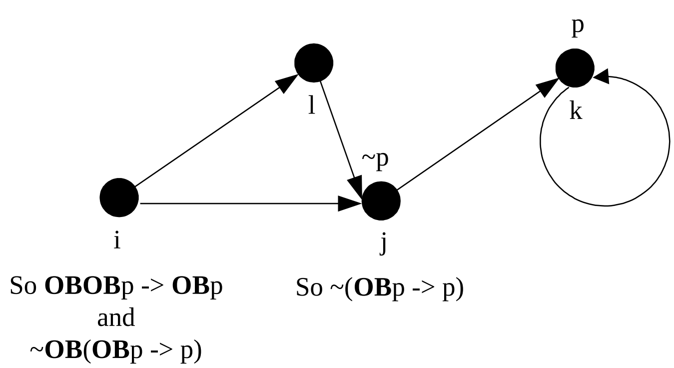
Two Counter-Models Regarding Additions to SDL: Here we show that A4, **OB**(**OB**p → p), is not derivable in SDL and that SDL + **OBOB**p → **OB**p does not imply A4 .

We first provide a counter-model to show that A4 is indeed a genuine (non-derivable) addition to SDL:



Here, seriality holds, since each of the three worlds has at least one world acceptable to it (in fact, exactly one), but *secondary* seriality fails, since although j is acceptable to i, j is not acceptable to itself. Now look at the top annotations regarding the assignment of truth or falsity to p at j and k. The lower deontic formuli derive from this assignment and the accessibility relations. (The value of p at i won't matter.) Since p holds at k, which exhausts the worlds acceptable to j, **OB**p must hold at j, but then, since p itself is false at j, (**OB**p → p) must be false at j. But j is acceptable to i, so not all i-acceptable worlds are ones where (**OB**p → p) holds, so **OB**(**OB**p → p) must be false at i.[27] We have already proven that seriality, which holds in this model, automatically validates **OB**-D. It is easy to show that the remaining ingredients of SDL hold here as well.[28]

We proved above that (**OBOB**p → **OB**p) is derivable from A4. Here is a model that shows that the converse fails. It is left to the reader to verify that given the accessibility relations and indicated assignments to p at j and k, **OBOB**p → **OB**p must be (vacuously) true at i, while **OB**(**OB**p → p) must be false at i.



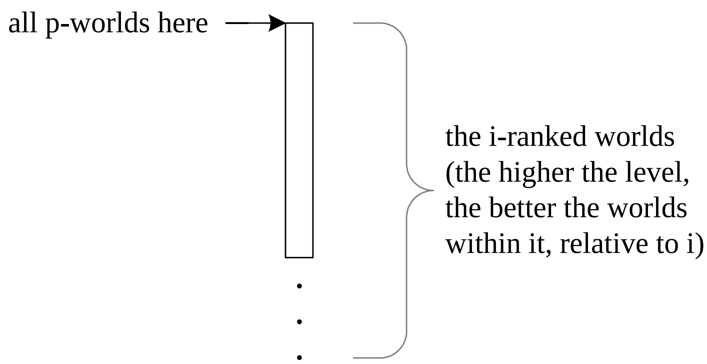
[27] Note that this is in contrast to j itself, where the latter formula does hold, for the reader can easily verify that (**OB**p → p) holds at k in this model, and k is the only world acceptable to j.

[28] The remaining items hold independently of seriality. Completing the proof amounts to both a proof of SDL's soundness with respect to our semantics, and of A4's independence (non-derivability from) SDL.

We should also note that one alternative semantic picture for SDL is where we have a set of world-relative ordering relations, one for each world i in W, where $j \geq_i k$ iff j is as good as k (and perhaps better) relative to i, where not all worlds in W need be in the purview (technically, the field of) of ordering relation associated with i. We then assume that from the standpoint of any world i, a) each world in its purview is as good as itself, b) if one is as good as a second, and the second is as good as a third, then the first is as good as the third, c) and for any two worlds in its purview, either the first is as good as the second or vice versa (i.e. respectively, each such $\geq i$ is reflexive, transitive, and connected in the field of $\geq_i$). **OB**p is then true at a world i iff there is some world k that is first of all as good as itself relative to i, and all worlds ranked as good as k from the standpoint of i are p-worlds. Thus, roughly, **OB**p is true at i iff p is true from somewhere on up in subset of worlds in W ordered relative to i. It is widely recognized that this approach will also determine SDL, but proofs of this are not widely available.[29]

However, if we add "The Limit Assumption", that for each world i, there is always at least one world as good (relative to i) as all worlds in i's purview (i.e. one i-best world), we can easily generate our earlier semantics for SDL derivatively. We need only add the natural analogue to our prior truth-conditions for **OB**: **OB**p is true at a world i iff p is true at all the i-best worlds.

OBp:



Essentially, the ordering relation coupled with the Limit Assumption just gives us a way to *generate* the set of i-acceptable worlds instead of taking them as primitive in the semantics: j is i-acceptable iff j is i-best. Once generated, we look only at what is going on in the i-acceptable worlds to interpret the truth-conditions for the various deontic operators, just as with our simpler Kripke-Style semantics. The analogue to the seriality of our earlier i-acceptability relation is also assured by the Limit Assumption, since it entails that for each world i, there is always some i-acceptable (now i-best) world. Although this ordering semantics approach appears to be a bit of overkill here, it became quite important later on in the endeavor to develop expressively richer deontic logics (ones going beyond the linguistic resources of SDL). We will return to this later.

For now, we turn to the second-most well known approach to monadic deontic logic, one in which SDL will emerge *derivatively*.

[29] But see Goble Forthcoming-b.

3. The Andersonian-Kangerian Reduction

The *Andersonian-Kangerian* reduction is dually-named in acknowledgement of Kanger's and Anderson's independent formulation of it around the same time.[30] As Hilpinen 2001a points out, the approach is adumbrated much earlier in Leibniz. We follow Kanger's development here, noting Anderson's toward the end.

3.1 Standard Syntax

Assume that we have a language of classical modal propositional logic, with a distinguished (deontic) propositional constant:

"*d*" for "all (relevant) normative demands are met".

Now consider the following axiom system, "K*d*":

K*d*:	A1: All Tautologies	(TAUT)
	A2: $\Box(p \to q) \to (\Box p \to \Box q)$	(K)
	A3: $\Diamond d$	($\Diamond d$)
	R1: If $\vdash p$ and $\vdash p \to q$ then $\vdash q$	(MP)
	R2: If $\vdash p$ then $\vdash \Box q$	(NEC)

K*d* is just the normal modal logic K with A3 added.[31] A3 is interpreted as telling us that it is possible that all normative demands are met. In import when added to system K, it is similar to (though stronger than) the "No Conflicts" axiom, A3, of SDL. *All* of the Traditional Scheme's deontic operators are defined operators in K*d*:

$\mathbf{OB}p =_{df} \Box (d \to p)$
$\mathbf{PE}p =_{df} \Diamond(d \,\&\, p)$
$\mathbf{IM}p =_{df} \Box (p \to \neg d)$
$\mathbf{GR}p =_{df} \Diamond(d \,\&\, \neg p)$
$\mathbf{OP}p =_{df} \Diamond(d \,\&\, p) \,\&\, \Diamond(d \,\&\, \neg p)$

So in K*d*, p is *obligatory* iff p is necessitated by all normative demands being met, *permissible* iff p is compatible with all normative demands being met, *impermissible* iff p is incompatible with all normative demands, *gratuitous* iff p's negation is compatible with all normative demands, and *optional* iff p is compatible with all normative demands, and so is ~p. Since none of the operators of the Traditional Scheme are taken as primitive, and the basic logic is a modal logic with

[30] Kanger 1971 [1957] (circulating in 1950 as a typescript) and Anderson 1967 [1956] and Anderson 1958.

[31] K is the basic (weakest) normal modal logic. (See the entry in this volume on modal logics by Rob Goldblatt.) Traditionally, and in keeping with the intended interpretation, the underlying modal logic had T as a theorem, indicating that necessity was truth-implicating. We begin with K instead because T generates a system *stronger* than SDL. We will look at the addition of T shortly. Åqvist 2002 [1984] is an excellent source on the meta-theory of the relationship between SDL-ish deontic logics and corresponding Andersonian-Kangerian modal logics, as well as the main dyadic (primitive conditional operator) versions of these logics. Smiley 1963 is a landmark in the comparative study of such deontic systems. McNamara 1999 gives determination results for various deontic logics that employ three deontic constants allowing for a "reduction" of other common sense normative concepts.

necessity and possibility as the basic modal operators, this is referred to as "a reduction" (of deontic logic to modal logic)

Proofs of SDL-ish wffs are then just K-proofs of the corresponding modal formulae involving "*d*".

Two Simple Proofs in K*d*: First consider the very simple proof of **OB***d*:

By PC, we have $d \rightarrow d$ as a theorem. Then by R2, it follows that $\Box(d \rightarrow d)$, that is, **OB***d*.

Next consider a proof of NC, **OB**p $\rightarrow$ ~**OB~~**p. As usual, in proofs of wffs with deontic operators, we make free use of the rules and theorems that carry over from the normal modal logic K. Here it is more perspicuous to lay the proof out in a numbered-lined stack:

1. Assume ~(**OB**p $\rightarrow$ ~**OB~**p).	(For reductio)
2. That is, assume $\sim(\Box(d\rightarrow p) \rightarrow \sim\Box(d\rightarrow \sim p))$	(Def of "**OB**")
3. So $\Box(d \rightarrow p)$ & $\Box(d \rightarrow \sim p)$.	(2, by PC)
4. So $\Box(d \rightarrow (p \;\&\; \sim p))$.	(3, derived rule of modal logic, K)
5. But $\Diamond d$	(A3)
6. So $\Diamond(p \;\&\; \sim p)$.	(4 and 5, derived rule of modal logic, K)
7. But $\sim\Diamond(p \;\&\; \sim p)$.	(a theorem of modal logic, K)
8. So **OB**p $\rightarrow$ ~**OB~**p	(1-7, PC)

Part of the point of the Andersonian-Kangerian reduction is to find a was to *generate* SDL from non-SDL resources, which can be easily done in K*d* (as the next box shows).

SDL Containment Proof: We give a proof that SDL is indeed contained in K*d.*

Recall SDL:

A1: All tautologous wffs of the language	(TAUT)
A2: **OB**(p $\rightarrow$ q) $\rightarrow$ (**OB**p $\rightarrow$ **OB**q)	(**OB**-K)
A3: **OB**p $\rightarrow$ ~**OB~**p	(**OB**-NC)
R1: If $\vdash$ p and $\vdash$ p $\rightarrow$ q then $\vdash$ q	(MP)
R2: If $\vdash$ p then $\vdash$ **OB**p	(**OB**-NEC)

We have already shown **OB**-NC is derivable in K*d* above, and TAUT and MP are given, since they hold for *all* formulas of K*d*. So we need only derive **OB**-K and **OB**-NEC of SDL, which we will do in reverse order. Note that RM, if $\vdash r \rightarrow s$, then $\vdash \Box r \rightarrow \Box s$), is derivable in K*d*, and so we rely on it in the second proof. [32]

[32] An examination of our earlier proof that RM for **OB** was one of the derived rules within SDL reveals that *for any* system with NEC and K governing a necessity operator, the rule RM is derivable. Here it is again adapted for $\Box$:

Show: If $\vdash p \rightarrow q$, then $\vdash \Box p \rightarrow \Box q$. (RM)
Proof: Suppose $\vdash p \rightarrow q$. Then by NEC, $\vdash \Box(p \rightarrow q)$, and then by K, $\vdash \Box p \rightarrow \Box q$.

Show: If $\vdash$ p then $\vdash$ **OB**p. (**OB**-NEC)
Proof: Assume $\vdash$ p. It follow by PC that $\vdash d \to p$. So by NEC for $\Box$, we get $\vdash \Box(d \to p)$, that is, **OB**p.

Show: $\vdash$ **OB**$(p \to q) \to ($**OB**$p \to$ **OB**$q)$. (K of SDL)
Proof: Assume **OB**$(p \to q)$ and **OB**p. From PC alone, $\vdash (d \to (p \to q)) \to [(d \to p) \to (d \to q)]$. So by RM for $\Box$, we have $\vdash \Box(d \to (p \to q)) \to \Box[(d \to p) \to (d \to q)]$. But the antecedent of this is just, **OB**$(p \to q)$ in disguise, which is our first assumption. So we have $\Box[(d \to p) \to (d \to q)]$ by MP. Applying K for $\Box$ to this, we get $\Box(d \to p) \to \Box(d \to q)$. But the antecedent to this is just our second assumption, **OB**p. So by MP, we get $\Box(d \to q)$, that is, **OB**q.

Metatheorem: SDL is derivable in K*d*.

Note that showing that the pure deontic fragment of K*d* contains *no more than* SDL is a more complex matter. The proof relies on already having semantic metatheorems available. An excellent source for this is Åqvist 2002 [1984].[33]

In addition to containing all theorems of SDL, we note a few theorems specific to K*d* because of the non-overlapping syntactic ingredients, *d*, $\Box$, and $\Diamond$:

$\vdash$ **OB***d*	
$\vdash \Box(p \to q) \to ($**OB**$p \to$ **OB**$q)$	(RM')
$\vdash \Box p \to$ **OB**p	(Nec')
$\vdash$ **OB**$p \to \Diamond p$	(Kant's Law)
$\vdash \sim\Diamond($**OB**p & **OB**$\sim$p$)$	(NC')

These are easily proven.[34]

Although our underlying modal system is just K, adding further *non-deontic* axiom schema (i.e. those neither abbreviate-able via SDL wffs, nor involving *d* specifically) can nonetheless have a deontic impact. To illustrate, suppose we added a fourth axiom, one to the effect that necessity is here truth-implicating, called axiom "T":

T: $\Box p \to p$

Call the system that results from adding this formula to our current system "KT*d*". Axiom T is certainly plausible enough here, since, as mentioned above, this approach to deontic logic is more sensible if necessity is interpreted as truth-implicating, since it takes obligations to be things necessitated by all normative demands being met, but in what sense, if not a truth-implicating sense of necessity?

33 (The "pure deontic fragment" is the set of theorems of K*d* that can be abbreviated using only the truth-functions and the five standard deontic operators.)
34 We proved the first above, and given our definition of **OB**, RM′ and NEC′ follow from standard features of the modal logic K alone, but Kant's Law and NC′ also depend on the distinctive deontic axiom, $\Diamond d$.

Now with T added to Kd, we have gone beyond SDL, since we can now prove things expressible in SDL's language that we have already shown are not theorems of SDL. The addition of T makes derivable our previously mentioned axiom A4 of SDL+, which we have shown is not derivable in SDL itself:

$\vdash \mathbf{OB}(\mathbf{OB}\text{p} \to \text{p})$[35]

So, reflecting on the fact that SDL+ is derivable in KTd, we see that the Andersonian-Kangerian reduction must either rely on a non-truth-implicating conception of necessity in order for its pure deontic fragment to match SDL, or SDL itself is not susceptible to the Andersonian-Kangerian reduction. Put another way, the most plausible version of the Andersonian-Kangerian reduction can't help but view "Standard Deontic Logic" as too weak.

Determinism and Deontic Collapse in the Classic A-K-Framework: A brief exploration of determinisms catastrophic implications for deontic matters in the context of KTd is given here. Note that adding T, $\Box\text{p} \to \text{p}$, allows us to explore a classical issue connected with determinism and deontic notions. Given axiom T, $\Box$ is now naturally taken to encode a truth-implicating notion of necessity in systems containing it. For this reason, we can now easily augment KTd with an axiom expressing determinism:

$\vdash \text{p} \to \Box\text{p}$. (Determinism)

It is obvious on a moments reflection that, along with T, Determinism (as an axiom schemata), yields a collapse of modal distinctions, since $\text{p} \leftrightarrow \Box\text{p}$, and $\text{p} \leftrightarrow \Diamond\text{p}$ would then be provable. However, we can also explore, the classical question of what happens to moral distinctions if determinism holds. This question is also settled from the perspective of KTd, since the following is a derivable rule of that system:

If $\vdash (\text{p} \to \Box\text{p})$, then $\vdash (\text{p} \leftrightarrow \mathbf{OB}\text{p})$.

To prove this, assume Determinism, $\vdash (\text{p} \to \Box\text{p})$.

a) We first show $\vdash \text{p} \to \mathbf{OB}\text{p}$. Assume p. Then by Determinism, $\Box$p. So by NEC′, namely $\Box\text{p} \to \mathbf{OB}\text{p}$, we get **OB**p, and thus $\vdash \text{p} \to \mathbf{OB}\text{p}$.

b) Next, we show $\vdash \mathbf{OB}\text{p} \to \text{p}$. Assume **OB**p and ~p for reductio. By Determinism, we have $\sim\text{p} \to \Box\sim\text{p}$. So $\Box\sim$p. This yields **OB~**p, by NEC′. But then we have **OB**p & **OB**~p, which contradicts a prior demonstrated theorem, NC. So $\vdash \mathbf{OB}\text{p} \to \text{p}$.

So, from the standpoint of the classic Andersonian-Kangerian reduction, where the notion of necessity is truth-implicating (and thus axiom T is intended), the addition of the most natural expression of determinism entails that truth and deontic distinctions collapse. This in turn is easily seen to imply these corollaries:

[35] Proof: By T, $\vdash \Box(d \to \text{p}) \to (d \to \text{p})$. Then by PC, we can get $\vdash d \to [\Box(d \to \text{p}) \to \text{p}]$. From this in turn, by NEC, we have $\vdash \Box(d \to [\Box(d \to \text{p}) \to \text{p}])$, that is, $\mathbf{OB}(\mathbf{OB}\text{p} \to \text{p})$.

If $\vdash (p \to \Box p)$, then $\vdash (p \leftrightarrow \mathbf{PE}p)$
If $\vdash (p \to \Box p)$, then $\vdash (\mathbf{OB}p \leftrightarrow \mathbf{PE}p)$
If $\vdash (p \to \Box p)$, then $\vdash ({\sim}p \leftrightarrow \mathbf{IM}p)$
If $\vdash (p \to \Box p)$, then $\vdash {\sim}\mathbf{OP}p$.

For example, consider the last corollary. By definition, p is optional iff neither p nor ~p is obligatory. But given determinism, this would entail that neither p nor ~p is true, which is not possible. So nothing can be morally optional if determinism is true.

Anderson's approach is practically equivalent to Kanger's. First, consider the fact that we can easily define another constant in K*d*, as follows:

$$s =_{df} {\sim}d,$$

where this new constant would now be derivatively read as follows:

"some (relevant) normative demands has been violated".

Clearly our current axiom, $\Diamond d$, could be replaced with ${\sim}\Box s$, asserting that it is not necessary that some normative demand is violated. We could then define **OB** as:

$$\mathbf{OB}p =_{df} \Box({\sim}p \to s),$$

and similarly for the other four operators.

Essentially, Anderson took this equivalent course with "*s*" being his primitive, and initially interpreted as standing for something like "the sanction has been invoked" or "there is a liability to sanction", and then ${\sim}\Box s$ was the axiom added to some modal system (at least as strong as modal system KT) to generate SDL and kin.

We should also note that Anderson was famous as a founding figure in relevant logic. (See the entry in this volume on relevant logics by Greg Restall.) Instead of using strict implication, $\Box(p \to q)$, he explored the use of a relevant (and thus neither a material nor strict) conditional, $\Rightarrow$, to express the reduction as: $\mathbf{OB}p =_{df} {\sim}p \Rightarrow s$. (A bit more on this can be found in Lokhorst 2004. See references there, but also see Mares 1992.) This alternative reflects the fact that there is an issue in both Kanger's and Anderson's strict necessitation approaches of just what notion of "necessity" can we say is involved in claiming that meetings all normative demands (or avoiding the sanction) *necessitates* p?

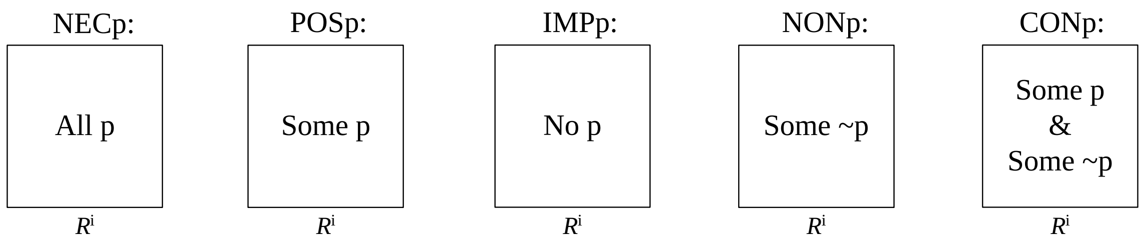
As a substantive matter, how should we think of these "reductions"? For example, should we view them as giving us an *analysis* of what it is for something to be obligatory? Well, taking Kanger's course first, it would seem that *d* must be read as a distinctive deontic ingredient, if we are to get the derivative deontic reading for the "reduced" deontic operators. Also, as our reading suggests, it is not clear that *d* does not, at least by intention, express a complex quantificational notion involving the very concept of obligation (demand) as a proper part, namely that all obligations have been fulfilled, so that the "reduction", *presented as an analysis*, would appear to be circular. If we read *d* instead as "ideal circumstances obtain", the claim of a substantive reduction or analysis appears more promising, until we ask, "Are the circumstances ideal only with respect to meeting normative *demands or obligations*, or are they ideal in other (for example supererogatory) ways that go beyond merely satisfying normative *demands*? Anderson's "liability to sanction" approach may appear more promising, since the idea that something is obligatory if (and only if) and because non-compliance necessitates (in some sense) liability to (or perhaps desert of) punishment does not appear to be circular, (unless the notion of "liability" itself ultimately involves the idea of permissibility of punishment), but is it still controversial (e.g. imperfect obligations are often thought to include obligations where no one has a right to sanction you for violations)? Alternatively, perhaps a norm that is merely an *ideal* cannot be *violated*, in which case perhaps norms that have been *violated* can be distinguished (as a subset) from norms that have not been *complied* with, and then the notion of an obligation as something that must obtain unless some norm is violated will not be obviously circular. The point here is that there is a substantive philosophical question lingering here that the language of a "reduction" brings naturally to the surface. The formal utility of the reduction does not hinge of this, but its philosophical significance does.

3.2 Standard Semantics

The semantic elements here are in large part analogous to those for SDL. We have a binary relation again, but this time instead of a relation interpreted as relating worlds *acceptable* to a given world, here we will have a relation, *R*, relating worlds "accessible" to a given world (e.g. possible relative to the given world). The only novelties are two: 1) we add a simple semantic element to match our syntactic constant "*d*", and 2) we add a slightly more complex analog to seriality, one that links the accessibility relation to the semantic element added in order to model *d*. We introduce the elements in stages.

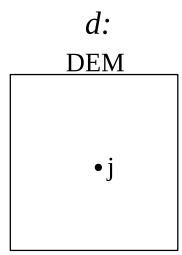
Once again, assume that we have a set of possible worlds, W, and assume that we have a relation, *R*, relating worlds to worlds, with the intention that *R*ij iff j is accessible to i (e.g. j is a world where everything true in j is possible relative to i)[36]. For brevity, we will call all worlds possible relative to i, "i-*accessible* worlds" and denote them by R^i. For the moment, no restrictions are placed on the relation *R*. We can illustrate these truth-conditions for our modal operators with a set of diagrams analogous to those used for giving the truth-conditions for SDL's deontic operators. We use obvious abbreviations for <u>nec</u>essity, <u>pos</u>sibility, <u>imp</u>ossibility, <u>non</u>-necessity, and <u>con</u>tingency:

[36] Note that this means that, for generality, we assume that what is possible may vary from one world to another. This is standard in this sort of semantics for modal logics. For example, if we wanted to model physical possibility and necessity, what is physically possible for our world, may not be so for some other logically possible world with different fundamental physical laws than ours. By adding certain constraints, we can generate a picture where what is possible does not vary at all from one world to another. (See the entry in this volume on modal logics by Rob Goldblatt.)



Here we imagine that for any given world, i, we have corralled all the i-accessible worlds together. We then simply look at the quantificational status of p (and/or ~p) in these i-accessible worlds to determine p’s modal status back at i: at a given world i, p is *necessary* if p holds throughout R^i, *possible* if p holds somewhere in R^i, *impossible* if p holds nowhere in R^i, *non-necessary* if ~p holds somewhere in R^i, and *contingent* if p holds somewhere in R^i, and so does ~p.

The only *deontic* element in the syntax of K*d* is our distinguished constant, *d*, intended to express the fact that all normative demands are met. To model that feature, we simply assume that the worlds are divided into those where all normative demands are met and those that are not. We denote the former subset of worlds by "DEM" in a model. Then *d* is true at a world j iff j belongs to DEM. Here is a picture where *d* is true at an arbitrary world, j:

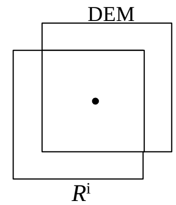


Since j is contained in DEM, that means all normative demands are met at j.[37]

Corresponding to simple seriality for SDL (that there is always an i-acceptable world), we assume what I will call "strong seriality" for K*d*: for every world i, there is an i-accessible world that is among those where all normative demands are met. In other words, for every world i, the intersection of the i-accessible worlds with those where all normative demands are met is non-empty. Given the truth conditions for *d*, strong seriality validates $\Diamond d$, ensuring that for any world i, there is always some i-*accessible* world where *d* is true:

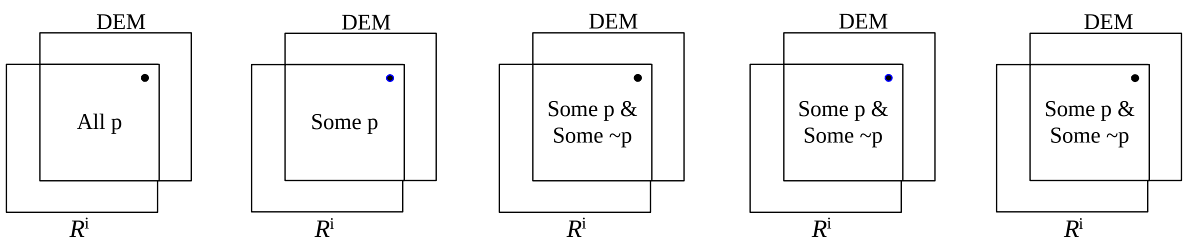
[37] Note that we could add an ordering-relation semantics like that described at the end of our section on SDL in order to generate the DEM component of these models. The main difference would be that instead of a set of world-relative ordering relations, one for each world (e.g. $\geq_i$) there would be just one ordering relation, ≥, whereby all worlds in W (in a given model) would be ranked just once. This relation would be reflexive, transitive and connected, while satisfying the Limit Assumption in W. DEM would then be the set of all the best worlds in W, and then the truth conditions for *d* and the five deontic operators would be cast via DEM so generated.

POS*d*:



Given these semantic elements, if you apply them to the definitions of the five deontic operators of K*d*, you will see that in each case, the normative status of p at i depends on p's relationship to *this intersection* of the i-accessible worlds and the worlds where all normative demands are met:

OBp: **PE**p: **IM**p: **GR**p: **OP**p:



If that inter-section is permeated by p-worlds, p is *obligatory*; if it contains some p-world, p is *permissible*, if it contains no p-world, p is *impermissible*, if it contains some ~p-world, p is *gratuitous*, and if it contains some p-world as well as some ~p-world, then p is *optional*.[38]

Kripke-Style Semantics for K*d*: A more formal characterization of this semantic picture is given here. We define the frames for modeling K*d* as follows:

F is an K*d* Frame: *F* = <W,R,DEM> such that:
1) W is a non-empty set
2) R is a subset of W x W
3) DEM is a subset of W
4) $\forall i \exists j(Rij \ \& \ j \in DEM)$.

A model can be defined in the usual way, allowing us to then define truth at a world in a model for all sentences of K*d* (as well as for KT*d*):

M is an K*d* Model: *M* = <*F*,V>, where *F* is an K*d* Frame, <W,R,DEM>, and V is an assignment on *F*: V is a function from the propositional variables to various subsets of W.

Basic Truth-Conditions at a world, i, in a Model, *M*:
[PC]: (Standard Clauses for the operators of Propositional Logic.)

[38] More explicitly, since **OB**p $=_{df} \Box(d \rightarrow p)$, we need only look at $\Box(d \rightarrow p)$. The latter will be true at a world i, iff $(d \rightarrow p)$ is true at all the i-accessible worlds. But given the truth-conditions for the material conditional "$\rightarrow$", that just amounts to saying p is true at all those i-accessible worlds (if any) where *d* is true, which in turn holds iff p is true at all the i-accessible worlds falling within DEM, i.e. at there intersection (which is non-empty by strong seriality). Similarly for the other four deontic operators.

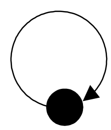
[$\Box$]: $M \models i\ \Box p$ iff $\forall j$(if Rij then $M \models j\ p$).
[d]: $M \models i\ d$ iff i is in DEM.

Derivative Truth-Conditions:
[$\Diamond$]: $M \models i\ \Diamond p$: $\exists j(\text{Rij} \ \&\ M \models j\ p)$
[**OB**]: $M \models i$ **OB**p: $\forall j[\text{if Rij} \ \&\ j \in \text{DEM then } M \models j\ p]$
[**PE**]: $M \models i$ **PE**p: $\exists j(\text{Rij} \ \&\ j \in \text{DEM} \ \&\ M \models j\ p)$
[**IM**]: $M \models i$ **IM**p: $\forall j[\text{if Rij} \ \&\ j \in \text{DEM then } M \models j\ {\sim}p]$
[**GR**]: $M \models i$ **GR**p: $\exists j(\text{Rij} \ \&\ j \in \text{DEM} \ \&\ M \models j\ {\sim}p)$
[**OP**]: $M \models i$ **OP**p: $\exists j(\text{Rij} \ \&\ j \in \text{DEM} \ \&\ M \models j\ p) \ \&\ \exists j(\text{Rij} \ \&\ j \in \text{DEM} \ \&\ M \models j\ {\sim}p)$

(Truth in a model and validity are defined just as for SDL.)

Metatheorem: K*d* is sound and complete for the class of all K*d* models.

If we wish to validate T, $\Box p \to p$ (and derivatively, A4, **OB**(**OB**p $\to$ p)), we need only stipulate that the accessibility relation, *R*, is reflexive: that each world i is i-accessible (possible relative to itself):



i

For then $\Box p \to p$ must be true at any world i, for if $\Box p$ is true at i, then p is true at each i-accessible world, which includes i, which is self-accessible. This will indirectly yield the result that **OB**(**OB**p $\to$ p) is true in all such models as well.

We turn now to a large variety of problems attributed to the preceding closely related systems.

4. Challenges to Standard Deontic Logics

Here we consider some of the "paradoxes" attributed to "Standard Deontic Logics" like those above (SDLs). Although the use of "paradox" is widespread within deontic logic and it does conform to a technical use in philosophical logic, namely the distinction between "paradox" and "antinomy" stemming from Quine's seminal "The Ways of Paradox" (Quine 1976 [1962]), I will also use "puzzle", "problem" and "dilemma" below.

To paraphrase von Wright, the number of outstanding problems in deontic logic is large, and most of these can be framed as problems or limitations attributed to SDLs. In this section we will list and briefly describe most of them, trying to group them where feasible under crucial principles of SDL or more general themes.

4.1 A Puzzle Centering around the Very Idea of a Deontic Logic:

Jorgensen's Dilemma (Jorgensen 1937):

A view still held by many researchers within deontic logic and metaethics, and particularly popular in the first few decades following the emergence of positivism, was that evaluative sentences are not the sort of sentences that can be either true or false. But then how can there be a logic of normative sentences, since logic is the study of what follows from what, and one thing can follow from another only if the things in question are the sort that can be either true or false? So there can be no deontic logic. On the other hand, some normative sentences do seem to follow from others, so deontic logic must be possible. What to do? That's Jorgesson's dilemma.

A widespread distinction is that between a *norm* and a *normative proposition*[39]. The idea is that a normative sentence such as "You may park here for one hour" may be used by an authority to provide permission on the spot or it may be used by a passerby to report on an already existing norm (e.g. a standing municipal regulation). The activity of using a normative sentence as in the first example is sometimes referred to as "norming"—it creates a norm by granting permission by the very use. The second use is often said to be descriptive, since the sentence is then not used to grant permission, but to report that permission to do so is a standing state. It is often maintained that the two uses are mutually exclusive, and only the latter use allows for truth or falsity. Some have challenged the exclusiveness of the division, by blending semantics and speech-act theory (especially regarding performatives), thereby suggesting that it may be that one who is in authority to grant a permission not only grants it in performing a speech act by uttering the relevant sentence (as in the first example), but also thereby makes what it said true (that the person is permitted to park).[40] [41]

4.2 A Problem Centering Around NEC:

The Logical Necessity of Obligations Problem

Consider

1) Nothing is obligatory.

A natural representation of this in the language of SDL would be:

[39] von Wright 1963; Hedenius 1963 [1941]; Alchourron and Bulygin 1971; Alchourron and Bulygin 1981; Makinson 1999. von Wright 1963 attributes the distinction to Ingemar Hedenius (Hedenius 1963 [1941]). See Makinson and van der Torre 2003 for a recent attempt to provide a logic of norms.
[40] See Lemmon 1962a; Kamp 1974, 1979. It is often thought that performative utterances generally work this way (Kempson 1977). For example, if a marriage ceremony conducted by a legitimate authority requires that authority to end the ceremony with the proverbial (but dated) "You are now man and wife" in order to complete an act of marriage, the speech act utilizing this sentence not only marries the couple (in the context), but it appears to also be a true description of their state as of that moment.
[41] Perhaps this is as good a place as any to direct the interested reader to a problem much discussed in metaethics since Hume: the so-called "Is-Ought Problem": Schurz 1997 is an excellent full length study employing the techniques of deontic and modal logic in investigating this problem.

1') $\sim$**OB**q, for all q.

We noted above that **OB**-NEC entails **OB**-N: $\vdash$ **OB**$\top$; but given 1′), we get $\sim$**OB**$\top$, and thus a contradiction. SDL seems to imply that it is a truth of logic that something is always obligatory. But it seems that what 1) expresses, an absence of obligations, is possible. For example, consider a time when no rational agents existed in the universe. Why should we think that any obligations existed then?

von Wright 1951 notes that since the denial of $\sim$**OB**$\top$ is provably equivalent to **PE**$\bot$ (given the traditional definitional scheme and **OB**-RE), and since both **OB**$\top$ and **PE**$\bot$ are odd, we should opt for a "principle of contingency", which says that **OB**$\top$ and $\sim$**PE**$\bot$ are both logically contingent. von Wright 1963 argues that **OB**$\top$ (and **PE**$\top$) do not express real prescriptions (pp.152-4). Føllesdal and Hilpinen 1971 suggests that excluding **OB**-N only excludes "*empty* normative systems" (i.e. normative systems with no obligations), and perhaps not even that, since no one can fail to fulfill **OB**$\top$ anyway, so why worry (p.13)[42] However, since it is dubious that anyone can bring it about that $\top$, it would seem to be equally dubious that anyone can "fulfill" **OB**$\top$, and thus matters are not so simple. al-Hibri 1978 discusses various early takes on this problem, rejects **OB**-N, and later develops a deontic logic without it. Jones and Porn 1985 explicitly rejects **OB**-N for "ought" in the system developed there, where the concern is with what people ought to do. If we are reading **OB** as simply "it ought to be the case that", it is not clear that there is anything counterintuitive about **OB**$\top$ (now read as, essentially, "it ought to be that contradictions are false"), but there is also no longer any obvious connection to what is obligatory or permissible for that reading, or to what people ought to do.

<u>4.3 Puzzles Centering Around RM</u>:

Free Choice Permission Paradox (Ross 1941):

Consider:

1) You may either sleep on the <u>s</u>ofa-bed or sleep on the <u>g</u>uest room bed.[43]
2) You may sleep on the sofa-bed and you may sleep on the guest room bed.

The most straightforward symbolization of these in SDL appears to be:

1′) **PE**(s $\vee$ g)
2′) **PE**s & **PE**g

Now it is also natural to see 2) as following from 1): if you permit me to sleep in either bed, it would seem that I am permitted to sleep in the first, *and* I am permitted to sleep in the second (though perhaps not to sleep in both, straddling the two, as it were). But 2′) does not follow from 1′) and the following is not a theorem of SDL:

---

[42] cf. Prior 1958.
[43] I will underline key letters to serve as cues for symbolization schemes left implicit, but hopefully clear enough.

* $\mathbf{PE}(p \vee q) \rightarrow (\mathbf{PE}p \ \& \ \mathbf{PE}q)$

Furthermore, suppose * were added to a system that contained SDL. Disaster would result. For it follows from **OB**-RM that $\mathbf{PE}p \rightarrow \mathbf{PE}(p \vee q)$.[44] So with * it would follow that $\mathbf{PE}p \rightarrow (\mathbf{PE}p \ \& \ \mathbf{PE}q)$, for any q, so we would get

** $\mathbf{PE}p \rightarrow \mathbf{PE}q$,

that if anything is permissible, then everything is, and thus it would also be a theorem that nothing is obligatory, $\vdash \sim\mathbf{OB}p$.[45]

Some have argued for two senses of "permissibility" here.[46]

<u>The Violability Puzzle</u>:[47] Here is another puzzle centering around RM. It would seem that it is of the very nature of obligations that they are violable in principle, unlike simple assertions, so that the following seems to be a conceptual truth:

1) If it is logically impossible that p is false, then it is logically impossible that p is obligatory.

But in SDL, this would naturally be expressed as a rule of inference:

If $\vdash p$ then $\vdash \sim\mathbf{OB}p$ (Violability)

But since $\top$ is a logical truth, Violability would yield $\sim\mathbf{OB}\top$, which directly contradicts theorem **OB**-N. Thus, SDL seems to make it a logical truth that there are inviolable obligations. But the idea that it is obligatory that it is either raining or not raining, something that couldn't be otherwise on logical grounds, seems counterintuitive. Furthermore, even in a system that lacked the force of **OB**-NEC and **OB**-N, if it has the force of just the rule RM (if $\vdash p \rightarrow q$ then $\vdash \mathbf{OB}p \rightarrow \mathbf{OB}q$), then were we to also countenance the Violability rule in such a system, we would be immediately forced to conclude that nothing is obligatory, $\vdash \sim\mathbf{OB}p$, thus rendering the system inapplicable.[48] von Wright 1963 comes close to endorsing Violability (p. 154), but the context there is more complex and less straightforward than that above. Jones and Porn 1985 provides a

[44] This follows from RM and the definition of **PE**: Suppose $\vdash p \rightarrow q$. Then $\vdash \sim q \rightarrow \sim p$. So by RM, $\vdash \mathbf{OB}\sim q \rightarrow \mathbf{OB}\sim p$, and thus $\vdash \sim\mathbf{OB}\sim p \rightarrow \sim\mathbf{OB}\sim q$, i.e. $\vdash \mathbf{PE}p \rightarrow \mathbf{PE}q$. Now just let q be $(p \vee q)$.

[45] For suppose something was obligatory, say **OB**p. Then by NC, it follows that **PE**p. One instance of ** above is $\mathbf{PE}p \rightarrow \mathbf{PE}\sim p$. So we would then have $\mathbf{PE}\sim p$, which by RE, PC and the definition of **PE** amounts to $\sim\mathbf{OB}p$, contradicting our assumption. Thus nothing could be obligatory.

[46] For example, one sense would be as in SDL (the simple absence of a prohibition), the other being a stronger sense of permission (von Wright 1968) with a distinct logic that would, for example, ratify *, but not **, above. Another approach was to say that this is a pseudo-problem, since the conjunctive use of "or" in the context of a permission word can be expressed as a conjunction of permitting conjuncts, **PE**p & **PE**q (Føllesdal and Hilpinen 1971). Kamp 1974, 1979 contain detailed analyses of these issues, one sensitive to both the semantics and pragmatics of permission.

[47] von Wright 1963, p.154 comes very close to stating this objection.

[48] For suppose **OB**p. Then since by PC, $\vdash p \rightarrow \top$, it follows by **OB**-RM that $\vdash \mathbf{OB}p \rightarrow \mathbf{OB}\top$. But since by PC, $\vdash \top$, by Violability, it follows that $\vdash \sim\mathbf{OB}\top$. So by PC, $\vdash \sim\mathbf{OB}p$, for any p.

system designed explicitly to accommodate violability (among other things) for their analysis of "ought".

*Ross's Paradox (*Ross 1941):

Consider:

1) It is obligatory that the letter is mailed.
2) It is obligatory that the letter is <u>m</u>ailed or the letter is <u>b</u>urned.

In SDLs, these seem naturally expressible as:

1′) **OB**m
2′) **OB**(m ∨ b)

But $\vdash$ **OB**p → **OB**(p ∨ q) follows by RM from $\vdash$ p → (p ∨ q). So 2′) follows from 1′), but it seems rather odd to say that an obligation to mail the letter entails an obligation that can be fulfilled by burning the letter (something presumably forbidden), and one that would appear to be violated by not burning it if I don't mail the letter.

The Good Samaritan Paradox (Prior 1958):[49]

Consider:

1) It ought to be the case that Jones helps Smith who has been robbed.
2) It ought to be the case that Smith has been robbed.

Now it seems that the following must be true:

Jones helps Smith who has been robbed if and only if Jones <u>h</u>elps Smith and Smith has been <u>r</u>obbed.

But then it would appear that a correct way to symbolize 1) and 2) in SDLs is:

1′) **OB**(h & r)
2′) **OB**r

But it is a thesis of PC that (h & r) → r, so by RM, it follows that **OB**(h & r) → **OB**r, and then we can derive 2′) from 1′) by MP. But it hardly seems that if helping the robbed man is obligatory it follows that his being robbed is likewise obligatory.[50]

[49] Prior cast it using this variant of RM: If $\vdash$ p → q then $\vdash$ **IM**q → **IM**p (the impermissibility of Smith being robbed then appears to wrongly imply the impermissibility of helping him who has been robbed). See also Åqvist 1967, which has been very influential.
[50] This paradox can also be cast equivalently with just one agent, and via **IM** as easily as **OB**: "The Victims Paradox" notes that the victim of the crime helps herself only if there was a crime. If it is impermissible that there be a crime, it will follow under similar symbolization that it is impermissible for the victim of the crime to help herself,

The Paradox of Epistemic Obligation (Åqvist 1967): This is a much-discussed variant of the preceding paradox. Consider:

1) The bank is being robbed.
2) It ought to be the case that Jones (the guard) knows that the bank is being robbed.
3) It ought to be the case that the bank is being robbed.

Let us symbolize "Jones knows that the bank is being robbed" by "$K_j r$". Then it would appear that a correct way to symbolize 1) - 3) in SDLs (augmented with a "K" operator) is:

1′) r
2′) $\mathbf{OB}K_j r$
3′) $\mathbf{OB}r$

But it is a logical truth that if one knows that p then p is the case (surely Jones knows that the bank is being robbed only if the bank is in fact being robbed). So $\vdash K_j r \rightarrow r$ would hold in any system augmented with a faithful logic of knowledge. So in such a system, it would follow by RM that $\vdash \mathbf{OB}K_j r \rightarrow \mathbf{OB}r$, but then we can derive 3′) from this conditional and 2′) by MP.[51] But it hardly seems to follow from the fact that it is obligatory that the guard knows that the bank is being robbed, that it is likewise obligatory that the bank is being robbed. It seems that SDL countenances inferences from patently impermissible states of affairs that someone is obliged to know hold when they hold to the conclusion that the same impermissible state of affairs is obligatory.[52]

Some RM-Related Literature: One standard response to Ross's Paradox, the Good Samaritan Paradox (and the Paradox of Epistemic Obligation) is to try to explain them away. For example, Ross paradox is often quickly rejected as elementary confusion (Føllesdal and Hilpinen 1971) or it is rejected on the grounds that the inference is only pragmatically odd in ways that are independently predictable by any adequate theory of the pragmatics of deontic language (Castañeda 1981). Similarly, it has been argued that the Good Samaritan paradox is really a conditional obligation paradox, and so RM is not the real source of the paradox (Castañeda 1981; Tomberlin 1981). However, since these paradoxes all at least *appear* to depend on **OB**-RM, a natural solution to the problems is to undercut the paradoxes by rejecting **OB**-RM itself. Two accessible and closely related examples of approaches to deontic logic that reject **OB**-RM from a principled philosophical perspective are Jackson 1985 and Goble 1990a. Jackson 1985 argues for an approach to "ought to be" that links it to counterfactuals, and he informally explores its

which hardly sounds right. Similarly for "The Robber's (Repenter 's) Paradox", where now we focus on the robber making amends (or repenting) for his crime, and again we seem to get the result that it is impermissible for the robber to make amends for his crime, suggesting a rather convenient argument against all obligations to ever make amends for one's crimes. These early variations were used to show that certain initially proposed solutions to the Good Samaritan paradox didn't really solve the problem. Both versions are found in Nowell Smith and Lemmon 1960.

[51] 1′) is not really essential here, it just helps to clarify that 2) does not express some strange standing obligation but a transient one that emerges as a result of the *de facto* robbery.

[52] Theoretically one could claim that we have a conflict of obligations here, but this seems quite implausible. The banks' being robbed appears to be definitely non-obligatory.

semantics and logic; Goble 1990a makes a similar case for "good" and "bad" (as well as "ought"), formally tying these to logical features of counterfactuals explicitly. (Goble 1990b contains the main technical details.) Interestingly, their approaches also intersect with the issue of "actualism" and "possibilism" as these terms are used in ethical theory. Roughly, *possibilism* is the view that I ought to bring about p if p is part of the best overall outcome I *could* bring about, even if the goodness of this overall outcome, depends on all sorts of other things that I *would* not in fact bring about were I to bring about p. In contrast, *actualism* is the view that I ought to bring about p if doing so *would* in fact be better than not doing so, and this of course can crucially depend on what else I would do (ideal or not) were I to bring about p. (See Jackson and Pargetter 1986; Jackson 1988, and for early discussions of this issue, see Goldman 1976 and Thomason 1981a.) In Hansson 1990, and more elaborately in Hansson 2001, S.O. Hansson develops systems of deontic logic where he analyzes prohibitive and prescriptive deontic notions in terms of abstract properties of various preference orderings (e.g. a normative status is prohibitive whenever anything worse than something that has that status also has it). He also sees **OB**-RM as the main culprit in the paradoxes of standard deontic logic, and thus he methodically explores non-standard frameworks where **OB**-RM is not sound. Hansson 2001 is also important for its extensive and original work on preference logic and preference structures, which, as we have already noted, are used regularly in deontic logic (and elsewhere). A very useful general source that covers some of the issues surrounding **OB**-RM, along with many others, is van der Torre 1997.

4.4 Puzzles Centering Around NC, OD and Analogues:

Sartre's Dilemma and Conflicting Obligations (Lemmon 1962b)[53]:

A conflict of obligations is a situation where there are two obligations and it is not possible for both to be fulfilled.

Consider the following conflict:

1) It is obligatory that I now meet Jones (say, as promised to Jones, my friend).
2) It is obligatory that I now do not meet Jones (say, as promised to Smith, another friend).

Here it would seem that I have a conflict of obligations, in fact a quite direct one, since what I promised one person would happen, I promised another would not happen. People do (e.g. under pressure or distraction) make such conflicting promises, and it appears that they incur conflicting obligations as a result.[54] But consider the natural representation of these in SDLs:

1′) **OB**j
2′) **OB**~j

[53] I change the example to have the conflict be direct and explicit. Sartre's much cited example is of a man obligated to join the resistance (to avenge his brother's death and fight the Nazi occupation) and obligated to stay home and aid his ailing mother (devastated by the loss of the man's brother, her son, and deeply attached to the one son still alive).
[54] Whether or not these obligations are both all-things-considered-obligations is a further issue. For our purposes here, the point is that they appear to be obligations. See the upcoming puzzle, Plato's Dilemma, for further issues.

But since NC, **OB**p → ~**OB**~p, is a theorem of all SDLs, we can quickly derive a contradiction from 1) and 2), which means that 1′) conjoined with 2′) represents a logically inconsistent situation. Yet, the original hardly seems logically incoherent.[55]

<u>*Kant's Law and Unpayable Debts*</u>: [56] Here is a simple puzzle about **OB**p → ◇p. Consider:

1) I'm obligated to pay you back $10 tonight.
2) I can’t pay you back $10 tonight (e.g. I just gambled away my last dime).

Since this puzzle typically involves some notion of possibility, let us represent the above sentences in KT*d*, which includes SDL, but also has a possibility operator:

1′) **OB**p
2′) ~◇p

1) and 2) appear to be consistent. It seems to be a sad fact that often, people are unable to fulfill their financial obligations, just as it seems to be a truism that financial obligations are obligations. But in KT*d*, it is a theorem that **OB**p → ◇p. So we derive a contradiction from this symbolization and the assumption that 1′) and 2′) are true.

A variant example is:

1) I owe you ten dollars, but I can't pay you back.
2) I'm obligated to pay you ten dollars, but I can't.

2) seems to follow from 1), and 1) hardly seems contradictory, since owing money clearly does not entail being able to pay the money owed. Thomason 1981b suggests a distinction between deliberative contexts of evaluation and judgmental contexts, where in the latter context evaluations such as 1) above need not satisfy Kant’s law since, roughly, we go back in time and evaluate the present in terms of where things would now be relative to optimal past options that were accessible but no longer are.

<u>*Conflation of Conflicts with Impossible Obligations*</u>: Here is another puzzle associated with NC and OD, one showing an inability in SDL to distinguish logically distinct situations.

We saw above that Kant’s Law, when represented as **OB**p → ◇p, is a theorem of KT*d*. If we interpret possibility here as practical possibility, then as the indebtedness example above suggests, it is far from evident that it is in fact true. However, a stronger claim than that of Kant’s

[55] von Wright 1968 refers to a conflict of obligations as a "predicament" and illustrates with the much-cited example of Jephthah (from the *Book of Judges*), who promises God to sacrifice the first living being he meets upon returning home from war, if God gives him victory, which wish is granted, but his daughter is the first living being he meets upon his return.
[56] *Kant’s* law is more accurately rendered as involving agency (if Doe is obligated to bring something about then Doe is able to do so), but the label is often used in deontic logic for almost any implication from something’s being obligatory to something’s being possible, roughly whatever formula comes closest to Kant’s in the system.

Law is that something cannot be obligatory unless it is at least *logically* possible. In SDL, this might be expressed by the rule:

If $\vdash$ ~p then $\vdash$ ~**OB**p.

This is derivable in SDL, since if $\vdash$ ~p, then $\vdash$ **OB**~p by **OB**-NEC, and then by **OB**-NC, we get $\vdash$ ~**OB**p. Claiming that Romeo is obligated to square the circle because he solemnly promised Juliet to do so is less convincing as an objection than the earlier financial indebtedness case. So SDL is somewhat better insulated from this sort of objection, and, as we noted earlier, we are confining ourselves here to theories that endorse **OB**-OD (i.e. $\vdash$ ~**OB**$\perp$).[57]

However, this points to another puzzle for SDL. The rule above is equivalent to $\vdash$ **OB**-OD in any system with **OB**-RE, and in fact, in the context of SDL, these are both equivalent to **OB**-NC. That is, we could replace the latter axiom with either of the former rules for a system equivalent to SDL. In particular, in any system with K and RM, (**OB**p & **OB**~p) $\leftrightarrow$ **OB**$\perp$ is a theorem.[58] But it seems odd that there is no distinction between a contradiction being obligatory, and having two distinct conflicting obligations. It seems that one can have a conflict of obligations without it being obligatory that some logically impossible state of affairs obtains. A distinction seems to be lost here. Separating **OB**-NC from **OB**-D is now quite routine in conflict-allowing deontic logics.

Some early discussions and attempted solutions to the last two problems can be found in Chellas 1980 and Schotch and Jennings 1981, both of whom use non-normal modal logics for deontic logic.[59] Brown 1996b uses a similar approach to Chellas' for modeling conflicting obligations, but with the addition of an ordering relation on obligations to model the relative stringency of obligations, thus moving in the direction of a model addressing Plato's Dilemma as well.

Let me note that a long-ignored and challenging further puzzle for conflicting obligations, called "van Fraassen's Puzzle" (van Fraassen 1973) has deservedly received increasing attention of late: Horty 1994, Horty 2003, van der Torre and Tan 2000, McNamara 2004a, Hansen 2004, and Goble Forthcoming-a.

The Limit Assumption Problem: Recall our earlier mention of an ordering semantics approach to SDL, and our mention there of the Limit Assumption:

that for each world i, there is always at least one world as good (relative to i) as all worlds in i's purview (i.e. one i-best world).

[57] However see Da Costa and Carnielli 1986 which develops a deontic logic in the context of paraconsistent logic.
[58] For first suppose **OB**p & **OB**~p holds. Then one instance of K is **OB**(~p $\rightarrow$ $\perp$) $\rightarrow$ (**OB~**p $\rightarrow$ **OB**$\perp$). But be **OB**-RE, **OB**(~p $\rightarrow$ $\perp$) is equivalent to just **OB**p, so we get **OB**p $\rightarrow$ (**OB~**p $\rightarrow$ **OB**$\perp$) by PC. So given **OB**p & **OB**~p, we get **OB**$\perp$ by PC. Second assume **OB**$\perp$. By PC, $\vdash$ $\perp$ $\rightarrow$ p. So by RM, we get **OB**$\perp$ $\rightarrow$ **OB**p, and then **OB**p. We can then generate **OB**~p the same way.
[59] Normal modal logics won't do since K and RE hold in all such logics. Chellas uses minimal models and Schotch and Jennings generalize Kripke models.

Although some in deontic logic have operated as if the Limit Assumption is true, it is a questionable assumption to make, especially as a matter of logic. It seems that there are possible scenarios in which the ordering of worlds in the purview of some world i have no upper bound on their goodness. Blake Barley gave a nice example in an unpublished paper, "The Deontic Dial", circulated at the University of Massachusetts-Amherst in the early 1980's: you have a dial that you can turn anywhere from 0 to 1, where both 0 and 1 yield disaster, but all the numbers in between yield better and better value, increasing with the natural order of the numbers (cf. McMichael 1978). In such a case, there seems to be no real sense to the old maxim: "Do the best you can!". This rules out the most natural simple clause for **OB** per optimizing theories:

> **OB**p is true at i iff p holds in all the i-best worlds,

for plainly in scenarios where there are no i-best worlds, nothing is obligatory and everything becomes permissible by this clause, but this seems wrong: even in the dial case, it seems clearly obligatory to not turn the dial to 1.

Lewis 1973 famously argued that the Limit Assumption (as used here or as used for modeling other counterfactuals) is an unjustified assumption, and that our clauses for deontic (and counterfactual) operators must reflect this fact. Most logicians agreed. This led to more complex clauses such as the one used earlier:

> **OB**p is true at i iff p is true from somewhere on up in the subset of worlds in W ordered relative to i,

and this in turn leads to greater complexity in the metatheory for such logics. The new clause has some odd features, for example, in the case of the dial, for each number between 0 and 1, you are obligated to turn the dial past that number, but in the scenario, the set of such obligatory things together entail that you turn the dial to 1, which is also forbidden. Although no conflict will show up in the system (no formula of the form **OB**p and **OB**~p will be validated), we nonetheless have an infinite set of obligations which cannot be jointly fulfilled, and thus a conflict of sorts, for which you can hardly be faulted. Lewis 1978 argue contra McMichael 1978, that the related problem McMichael there refers to (called "The Confinement Problem" in McNamara 1995) is a problem for utilitarianism, not for deontic logic; but see Fehige 1994, who suggests that there are still choices a logician must make and that "...When the best options are lacking, then so are flawless accounts of the lack" (p.42). Fehige provides a systematic critical discussion of deontic logicians approaches to the Limit Assumption.

Plato's Dilemma and Defeasible Obligations (Lemmon 1962b):[60]

1) I'm obligated to meet you for a light lunch meeting at the restaurant.
2) I'm obligated to rush my choking child to the hospital.

Here we seem to have an indirect conflict of obligations, if we assume that satisfying both obligations is practically impossible. Yet here, unlike in our prior example, where the two

[60] Here too I change the example. Plato's case involves returning weapons as promised to someone who now in a rage intends to unjustly kill someone with the weapon. Lemmon interprets the issues raised by Sartre's dilemma a bit differently than I do here.

promises might naturally have been on a par, we would all agree that the obligation to help my child *overrides* my obligation to meet you for lunch, and that the first obligation is *defeated* by the second obligation, which *takes precedence*. Ordinarily, we would also assume that no other obligation overrides my obligation to rush my son to the hospital, so that this obligation is an *all things considered obligation*, but not so for the obligation to meet you for lunch. Furthermore, we are also prone to say that the situation is one where the general obligation we have to keep our appointments (or to keep our promises, still more generally) has an *exception*--the circumstances are extenuating. Once we acknowledge conflicts of obligation, there is the further issue of representing the logic of reasoning about conflicting obligations where some override others, some are defeated, some are all things considered obligations, some are not, some hold generally, but not unexceptionally, etc. So the issue here is one of conflicting obligations of different weight and the defeasibility of one of two obligations. Clearly, there is no mechanism in SDL for this, since SDL does not allow for conflicts to begin with, yet this is an issue that goes well beyond that of merely having a logic that allows for conflicts. There have been a variety of approaches to this dilemma, and to defeasibility among conflicting obligations.

<u>Some Literature on Defeasible Obligations</u>: von Wright 1968 suggested that minimizing evil is a natural approach to conflict resolution, thereby suggesting that a sort of minimizing (and thus reliance on an ordering) is apt. Alchourron and Makinson 1981 provide an early formal analysis of conflict resolution via partial orderings of regulations and regulation sets. Chisholm 1964 has been very influential conceptually, as witnessed, for example, by Loewer and Belzer 1983. In ethical theory, the informal conceptual landmark is Ross 1939. Horty 1994 is a very influential discussion forging a link between Reiter's default logic developed in AI (see Brewka 1989), and an early influential approach to conflicts of obligation, van Fraassen 1973, which combines a preference ordering with an imperatival approach to deontic logic. Prakken 1996 discusses Horty's approach and an alternative that strictly separates the defeasible component from the deontic component, arguing that handling conflicts should be left to the former component only. See also Makinson 1993 for a sweeping discussion of defeasiblity and the place of deontic conditionals in this context. Other approaches to defeasibility in deontic logic that have affinities to semantic techniques developed in artificial intelligence for modeling defeasible reasoning about defeasible conditionals generally are Asher and Bonevac 1996 and Morreau 1996, both of which attempt to represent W. D. Ross-like notions of prima facia obligation, etc. Also notable are the discussions of defeasibility and conditionality in Alchourron 1993, 1996, where a revision operator (operating on antecedents of conditionals) is relied on in conjunction with a strict implication operator and a strictly monadic deontic operator. Note that Alchourron 1996, Prakken 1996, Asher and Bonevac 1996, Morreau 1996, and Prakken 1996 are all found in *Studia Logica* **57**.1, 1996 (guest edited by A. Jones and M. Sergot). Nute 1997 is dedicated to defeasibility in deontic logic and is the best single source on the topic, with articles by many of the key players, including Nute himself. See Bartha 1999 for an approach to contrary-to-duty conditionals and to defeasible conditionals layered over a branching time framework with an agency operator. Smith 1994 contains an interesting informal discussion of conflicting obligations, defeasibility, violability and contrary-to-duty conditionals. Since it is very much a subject of controversy and doubt as to whether deontic notions contribute anything special to defeasible inference relations (as opposed to defeasible conditionals), we leave this issue aside here, and turn to conditionals, and the problem in deontic logic that has received the most concerted attention.

4.5 Puzzles Centering Around Deontic Conditionals

The Paradox of Derived Obligation/Commitment (Prior 1954):

Consider the following statements:

1a) Bob's promising to meet you commits him to meeting you.
1b) It is obligatory that Bob meets you if he promises to do so.

It was suggested that these might be represented in either of two ways in SDL:

1′) p → **OB**m[61]
1″) **OB**(p → m)[62]

Consider 1′) first. The following are both simply tautologies: ~r → (r → **OB**s) and **OB**s → (r → **OB**s). So if 1′) reflected a proper analysis of 1a/b), anything false would commit us to anything whatsoever (e.g. since I am not now standing on my head, it would follow that my standing on my head commits me to giving you all my money) and everything commits us to anything obligatory (e.g. if I'm obligated to call you, then my standing on my head commits me to doing so). What of 1″) then? The following are theorems of SDL: **OB~**r → **OB**(r → s) and **OB**s → **OB**(r → s). So if 1″) reflected an apt analysis of commitment, it would follow from SDL that anything impermissible commits us to everything, and once again, everything commits us to anything obligatory. So, these seem to be troublesome candidates for symbolizing 1a) or 1b) in SDL. The problems are reminiscent of paradoxes about material implication (reading 1′)), and strict implication (reading 1″)), respectively.[63] So the question arose, are there any *special* problems associated with the interaction of deontic notions and conditionality? The next paradox (Chisholm's), increased the perception that there might very well be. Many consider it to be the most challenging and distinctive puzzle of deontic logic.

Contrary-to-Duty (or Chisholm's) Paradox (Chisholm 1963a):

Consider the following:

1) It ought to be that Jones goes to the assistance of his neighbors.
2) It ought to be that if Jones does go then he tells them he is coming.
3) If Jones doesn't go, then he ought not tell them he is coming.
4) Jones doesn't go.

This certainly appears to describe a possible situation. It is widely thought that 1)-4) constitute a *mutually consistent* and *logically independent* set of sentences. We treat these two conditions as

[61] In the 1st edition of Prior 1962 [1955].
[62] In von Wright 1951.
[63] In the case of symbolization 1″), since (r → s) is logically equivalent to (~r ∨ s), and the two troublesome formulas associated with this symbolization reduce to **OB~**r → **OB**(~r ∨ s) and **OB**s → **OB**(~r ∨ s), these are also instances of Ross' Paradox given this SDL interpretation of the sentences.

desiderata. Note that 1) is a *primary obligation*, saying what Jones ought to do unconditionally.[64] 2) is a *compatible-with-duty obligation*, appearing to say (in the context of 1)) what else Jones ought to do on the condition that Jones fulfills his primary obligation. In contrast, 3) is a *contrary-to-duty obligation or "imperative" (a "CTD")* appearing to say (in the context of 1)) what Jones ought to do conditional on his violating his primary obligation. 4) is a factual claim, which conjoined with 1), implies that Jones *violates* his primary obligation. Thus this puzzle also places not only deontic *conditional* constructions, but the *violability* of obligations, at center stage. It raises the challenging question: what constitutes proper reasoning about what to do in the face of violations of obligations?

How might we represent this quartet in SDL? The most straightforward symbolization is;

1′) $\mathbf{OB}g$.
2′) $\mathbf{OB}(g \rightarrow t)$.
3′) $\sim g \rightarrow \mathbf{OB}{\sim}t$.
4′) $\sim g$.

But Chisholm points out that from 2′) by principle **OB**-K we get $\mathbf{OB}g \rightarrow \mathbf{OB}t$, and then from 1′) by MP, we get $\mathbf{OB}t$; but by MP alone we get $\mathbf{OB}{\sim}t$ from 3′) and 4′). From these two conclusions, by PC, we get $\sim(\mathbf{OB}t \rightarrow \mathbf{OB}{\sim}t)$, contradicting NC of SDL. Thus 1′) - 4′) leads to inconsistency per SDL. But 1)-4) do not seem inconsistent at all, so the representation cannot be a faithful one. Various less plausible representations in SDL are similarly unfaithful. For example, we might try reading the second and third premises uniformly, either on the model of 2′) or on the model of 3′). Suppose that instead of 3′) above, we use 3″) $\mathbf{OB}(\sim g \rightarrow \sim t)$. The trouble with this is 3″) is derivable from 1′) in SDL, but there is no reason to think 3) in fact follows from 1), so we have an unfaithful representation again. Alternatively, suppose that instead of 2′) above, we use 2″) $g \rightarrow \mathbf{OB}{\sim}t$. This is derivable from 4′) in PC (and thus in SDL). But there is no reason to think 2) follows from 4). So again, we have an unfaithful representation.

The following displays the difficulties trying in obvious ways to interpret the quartet in SDL in tabular form:[65]

First Try :	Second Try:	Third Try:
1′) $\mathbf{OB}g$	1′) $\mathbf{OB}g$	1′) $\mathbf{OB}g$
2′) $\mathbf{OB}(g \rightarrow t)$	2′) $\mathbf{OB}(g \rightarrow t)$	2″) $g \rightarrow \mathbf{OB}t$
3′) $\sim g \rightarrow \mathbf{OB}{\sim}t$	3″) $\mathbf{OB}(\sim g \rightarrow \sim t)$	3′) $\sim g \rightarrow \mathbf{OB}{\sim}t$
4′) $\sim g$	4′) $\sim g$	4′) $\sim g$
----------	----------	----------
From 1′), 2′), $\mathbf{OB}t$.	3″) follows from 1′).	2″) follows from 4′).
From 3′), 4′), $\mathbf{OB}{\sim}t$.	So independence is lost.	So independence is lost.
By NC, $\mathbf{OB}t \rightarrow \sim\mathbf{OB}{\sim}t$.		
So $\perp$; consistency is lost.		

[64] Here we follow tradition (albeit self-consciously) in sloughing over the differences between what ought to be, what one ought to do, and what is obligatory.

[65] The remaining truly strained combination would replace 2′) with 2″) *and* 3′) with 3″), but that just doubles the trouble with the second and third readings, so it is routinely ignored.

Each reading of the original violates one of our desiderata: mutual consistency or joint independence.

If von Wright launched deontic logic as an academic specialization, Chisholm's Paradox was the booster rocket that provided the escape velocity deontic logic needed from subsumption under normal modal logics, thus solidifying deontic logic's status as a distinct specialization. It is now virtually universally acknowledged that Chisholm was right: the sort of conditional deontic claim expressed in 3) can't be faithfully represented in SDL, nor more generally by a composite of some sort of unary deontic operator and a material conditional. This is one of the few areas where there is nearly universal agreement in deontic logic. Whether or not this is because some special *primitive dyadic deontic conditional* is operating or because it is just that some *non-material* conditional is essential to understanding important deontic reasoning is still a hotly contested issue.

<u>Some Literature on Contrary-to-Duty Obligations</u>: von Wright 1956, 1971 take the now-classic non-componential dyadic operator approach to the syntax of CTDs. Danielsson 1968; Hansson 1969, Lewis 1973, 1974, and Feldman 1986 provide samples of a "next best thing" approach: the interpretation of conditional obligations via a primitive non-componential dyadic operator, in turn interpreted via a preference ordering of the possible worlds where the (perhaps obligation-violating) antecedent holds; see also Åqvist 2002 [1984] for an extensive systematic presentation of this sort of approach (among other things), and al-Hibri 1978 for an early widely-read systematic discussion of a number of approaches to CTDs (among other things). van Fraassen 1972, Loewer and Belzer 1983, and Jones and Porn 1985 also offer influential discussions of CTDs and propose distinct formal solutions, each also employing orderings of outcomes, but offering some twists on the former more standard pictures. An important forthcoming source on the metatheory of classical and near-classical logics via classic and near-classic ordering structures for the dyadic operator is Goble Forthcoming-b. Mott 1973; Goble Forthcoming-b and Chellas 1974 (and Chellas 1980) offer influential analyses of the puzzle by combining a *non-material* conditional and a *unary* deontic operator to form a genuine componential compound, p $\Rightarrow$ **OB**q, for representing conditionals like 3) above); DeCew 1981 is an important early critical response to this sort of approach. Tomberlin 1983 contains a very influential informal discussion of various approaches. Bonevac 1998 is a recent argument against taking conditional obligation to be a primitive non-componential operator, suggesting roughly that techniques like those developed in AI (see Brewka 1989) for defeasible reasoning suffice for handling woes with CTDs. Smith 1993, 1994 contain important discussions stressing the difference between violability and defeasibility, and the relevance of the former rather than the latter to CTDs. Åqvist and Hoepelman 1981, and van Eck 1982 (and again, Loewer and Belzer 1983) are classic representatives of attempts to solve the puzzle by incorporating temporal notions into deontic logic. Jones 1990 contains an influential argument against any temporal-based general solution to the puzzle. Castañeda 1981 argued that by carefully distinguishing between (roughly) propositions and actions in the scope of deontic operators, Chisholm's puzzle, as well as most puzzles for deontic logic, can be resolved; Meyer 1988 offers a version of this general approach using dynamic logic. Prakken and Sergot 1996 contains an influential argument against any such action-based general solution to the puzzle. For recent work on CTDs in the context of a branching time framework with agency represented a la Horty-Belnap, see Horty 1996, 2001, Bartha 1999, and Bartha's contribution (chapter 11) to Belnap 2001. A recent source that reviews a good deal of the literature on CTDs and proposes its own solution is Carmo and Jones 2002;

but see also material on this problem in Nute 1997 (especially van der Torre and Tan 1997, and Prakken and Sergot 1997).

Appendix A2 contains additional discussion of this very important paradox. One newer puzzle often discussed in either the context of **OB**-RM or in the context of discussing conditional obligations is the following.

The Paradox of the Gentle Murderer (Forrester 1984)[66]

Consider:

1) It is obligatory that John Doe does not kill his mother.
2) If Doe does k̲ill his mother, then it is obligatory that Doe kills her g̲ently.
3) Doe does kill his mother (say for an inheritance).

Then it would appear that a correct way to symbolize 1) and 2) in SDLs is:

1′) **OB**~k
2′) k → **OB**g
3′) k

First, from 2') and 3′), it follows that **OB**g by MP. But now add the following unexceptionable claim:

Doe kills his mother gently only if Doe kills his mother.

Assuming this, symbolized as g → k, is a logical truth in an expanded system, by **OB**-RM it follows that **OB**g → **OB**k, and so by MP again we get **OB**k. This seems bad enough, for it hardly seems that from the fact that if I kill my mother then I must kill her gently and that I will kill her (scoundrel that I am), we can conclude that I am actually obligated to kill my mother. Add to this that from **OB**k in turn, we get ~**OB**~k by NC of SDL, and thus we have a contradiction as well. So we must either construe 2) so that is does not satisfy *modus ponens* or we must reject **OB**-RM.[67]

[66] Also called "Forrester's Paradox".

[67] Some have suggested this is a problem stemming from scope difficulties, others have argued that the problem is that **OB**-RM is in fact invalid, and rejecting it solves the problem. (Sinnott-Armstrong 1985 argues for a scope solution; Goble 1991 criticizes the scope solution approach, and argues instead for rejecting **OB**-RM.) We have listed this puzzle here rather than under the Good Samaritan Puzzle (in turn under puzzles associated with **OB**-RM) since, unlike the Standard Good Samaritan, this puzzle seems to crucially involve a contrary-to-duty conditional, and so it is often assumed that a solution to the Chisholm Paradox should be a solution to this puzzle as well (and vice versa). Alternatively, one might see the puzzle as one where we end up obligated to kill our mother gently because of our decision to kill her (via factual detachment), and then by **OB**-RM, we would appear obligated to kill her, which has no plausibility by anyone’s lights, and thus calls for rejecting **OB**-RM. However, this would still include a stance on contrary-to-duty conditionals and detachment.

4.6 Problems Surrounding (Normative) Expressive Inadequacies of SDL:

Here we look at some monadic normative notions that appear to be inexpressible in SDL.

The Normative Gaps Puzzle (von Wright 1968)[68]:

In some normative systems, permissions, prohibitions and obligations are explicitly given. So it would seem to be possible for there to be normative systems with gaps: where something is neither obligatory, impermissible, nor permissible. Yet **OB**p ∨ (**PE**p & **PE**~p) ∨ **IM**p is a thesis ("exhaustion") of SDL (given the Traditional Definitional Scheme), which makes any such gaps impossible.

Urmson's Puzzle—Indifference versus Optionality (Urmson 1958):

Consider:

1) It is optional that you attend the meeting, but not a matter of indifference that you do so.

This seems to describe something quite familiar: optional matters that are nonetheless not matters of indifference. But when deontic logicians and ethicists gave an operator label for the condition (~**OB**p & ~**OB**~p), it was almost invariably "It is indifferent that p", "**IN**p". But then it would seem to follow from the theorem **OB**p ∨ (~**OB**p & ~**OB**~p) ∨ **IM~**p, that (~**OB**p & ~**IM~**p) → **IN**p, that is, everything that is neither obligatory nor prohibited is a matter of indifference. But many actions, including some heroic actions, are neither obligatory nor prohibited, yet they are hardly matters of indifference. We might put this by saying that SDL can represent optionality, but not indifference, despite the fact that the latter concept has been a purported target for representation since nearly its beginning (See also Chisholm 1963b and McNamara 1996a).

The Supererogation Problem (Urmson 1958):

Some things are beyond the call of duty or supererogatory (e.g. volunteering for a costly or risky good endeavor where others are equally qualified and no one person is obligated). SDL has no capacity to represent this complex concept.[69]

The Must versus Ought Dilemma (McNamara 1990, 1996c):

Consider:

1) Although you can skip the meeting, you ought to attend.[70]

[68] See also Alchourron and Bulygin 1971.
[69] For some attempts to accommodate supererogation in deontic logic, see Chisholm 1963b, Chisholm and Sosa 1966, Humberstone 1974, Forrester 1975, Sajama 1985; Hrushka and Joerden 1987; McNamara 1990, 1996a, 1996b, 1999.
[70] Cf. Chisholm 1963b.

This seems perfectly possible, even in a situation where no conflicting obligations are present, as we will suppose here. 1) appears to imply that it is optional that you attend—that you *can* attend and that you *can* fail to attend. It seems clear that the latter two uses of "can" express permissibility. Yet "ought" is routinely the reading given for deontic necessity in deontic logic (and in ethical theory), and then "permissibility" is routinely presented as its dual. But then if we symbolize 1) above accordingly, we get,

1′) **PE**~p & **OB**p

which is just ~**OB**p & **OB**p in disguise (given **OB**-RE and the Traditional Definitional Scheme). So 1′), given **OB**-NC, yields a contradiction. Another way to put this is that the "can" of permissibility is much more plausibly construed as the dual of "must" than as the dual of "ought". This yields a dilemma for standard deontic logic (really for most work in deontic logic):

> Either deontic necessity represents "ought", in which case, its dual does not represent permissibility (and neither does any other construction in SDL), or permissibility is represented in SDL, but "ought" is inexpressible in it despite the ubiquitous assumption otherwise.

That "ought" is the dual of permissibility is really a largely overlooked pervasive bipartisan presupposition in both ethical theory and deontic logic.[71]

The Least You Can Do Problem (McNamara 1990):

1) You should have come home on time; the least you could have done was called, and you didn't do even that.

The expression in the second clause has been completely ignored in the literature on deontic logic and ethical theory both. 1) appears to express the idea that there is some minimal but acceptable alternative (and the criticism suggested in the emphatic third clause is that not even that minimal acceptable option was taken, much less the preferable option identified in the first clause using "ought"). This notion of what is minimally acceptable among the permissible options is not expressible in SDL.

Regarding the last four problems, McNamara 1990, 1996a, 1996c, 1996b, and Mares and McNamara 1997 attempt to devise logics distinguishing "must" from "ought", indifference from optionality, as well as distinctly representing "the least you can do" idiom, and using this unstudied idiom to analyze one central sense of "supererogation". Appendix A3 contains a sketch of this framework for common sense morality.

---

[71] Jones and Porn 1986 gives an early attempt to distinguish the two, although "must" ends up looking more like *practical* necessity in their framework (that which holds in all scenarios—permissible or not) than *deontic* necessity (that which holds in all permissible scenarios). McNamara 1990, 1996c provide cumulative case arguments that "must" not "ought" is the dual of permissibility, and thus that it is this almost universally ignored term "must", not "ought", that tracks the traditional concern in ethical theory and deontic logic with permissibility. Forrester 1975 is an early attempt to sketch an operator scheme distinguishing "ought" from "obligatory".

4.7 Agency in Deontic Contexts

We routinely talk about both what *ought to be* and what people *ought to do*. These hardly look like the same things (for example, the latter notion calls for an agent, the former does not). This issue, and the general issue of representing agency in deontic logic has been much discussed, and continues to be an area of active concern.

The Jurisdictional Problem and the Need for Agency[72]

Consider the following:

1) Jane Doe is obligated to not bring it about that *your child is disciplined*.
2) Jane Doe is obligated to not bring it about *that your child is not disciplined*.

Suppose you have a child. For almost any Jane Doe, 1) is then true: she is obligated to not bring it about that *your* child is disciplined, since that is none of her business. Similarly, for 2): she also is obligated to not bring it about that *your* child is not disciplined, since bringing that about is also none of her business. How might we represent these in SDL? Suppose we try to read the **OB** of SDL as "Jane Doe is obligated to bring it about that __"[73]; then how do we express 1) and 2)? The closest we appear to be able to come is:

1′) ~**OB**p.
2′) ~**OB**~p.

But these won't do. Collectively, 1′) and 2′) amount to saying that two obligations are absent, that it is neither obligatory that Jane Doe brings it about that your child is disciplined nor obligatory that she brings it about that your child is not disciplined. But this is compatible with its being the case that both 1) and 2) above are false. After all, suppose now that you are Jane Doe, the single parent of your child. Then in a given situation, it may be that you, the child's sole parent and guardian, are both permitted to bring it about that the child is disciplined and permitted to bring it about that the child is not disciplined, in which case both 1) and 2) are false. (These permissions in fact appear to be equivalent to the negations of 1) and 2).) But the falsity of 1) (and the first permission) implies the truth of 2′) on the current reading, and the falsity of 2) (and the second permission) implies the truth of 1′) on the current reading. So clearly 1) and 1′) are not equivalent, nor are 2) and 2′).

Alternatively, on the proposed reading of **OB,** shifting the outer negation signs to the right of the operators in 1′) and 2′) will just get us this conflicting pair:

---

[72] The following formulation of the problem has the status of reconstructed deontic folklore in the form of an argument or problem explicitly showing the inability of SDL to be taken to represent agential obligations. The need to eventually represent agency in order to represent agential obligations was so widely recognized early on that arguments for it are hard to find. The earliest reference I have found that comes close to formulating the problem in just the following way is Lindahl 1977, p.94, which explicitly uses the "none of your business" terminology. However, it was surely known to Kanger, and fair to say it was presupposed by him in his attempted analysis of rights-related notions as far back as his seminal paper, Kanger 1971 [1957]. Cf. also von Wright 1968.

[73] Cf. "Jane Doe is obligated to see to it that __", von Wright 1971.

1″) **OB**~p
2″) **OB**~~p

which are hopeless candidates for symbolizing 1) and 2), which do not conflict with one another.

Also, consider this traditional equivalence:

IMp ↔ **OB**~p.

If we are going to read "**OB**" as having agency built into it, presumably we want to do the same for the other operators, and so **IM**p above will be read as "it is impermissible for Jane Doe to bring it about that p". However, this renders the left to right implication in the equivalence above unsound, for it may be true that it is impermissible for me to discipline your child, but false that it is obligatory for me to see to it that your child is positively not disciplined. The matter must be left up to you.

On the face of it, the "not"s in 1) and 2) are not external to the deontic operators, as it were, nor are they directly operating on p; rather they pertain to Jane Doe's agency with respect to p. They come "between" the deontic element and the agential element, so reading **OB** as an amalgamation of a deontic and agential operator does not allow for the "insertion" of any such negation. So, unsurprisingly, it looks like we simply must have some explicit representation of agency if we are to represent agential obligations like those in 1) and 2) above.

A Simple Kangerian Agency Framework

So let us introduce a standard operator for this missing element,

BA: Jane Doe brings it about that __.[74]

Then clearly the following relations expressing an agent's simple position with respect to a proposition, p, are to be distinguished:

BAp: Jane Doe brings it about that p
BA~p: Jane Doe brings it about that ~p
~BAp: Jane Doe does not bring it about that p
~BA~p: Jane Doe does not bring it about that ~p

Plainly, if neither of the first two hold, then the conjunction of the last two holds. In such a case we might say that Jane Doe is *passive* with respect to p, or more adequately, passive with respect to herself bringing about p or its negation.[75] Let's introduce such an operator:

---

[74] With two or more agents, we would need to represent agents explicitly: BA_sp, $BA_{s'}$p, etc. "E" is often used for this operator.

[75] This "passivity" terminology, although used elsewhere, is perhaps not ideal and can't seriously be viewed as an analysis of "passivity" per se, since one might bring about neither a proposition nor its negation, and yet be quite influential regarding it (e.g. intentionally and actively increasing its probability without making it happen), thus the

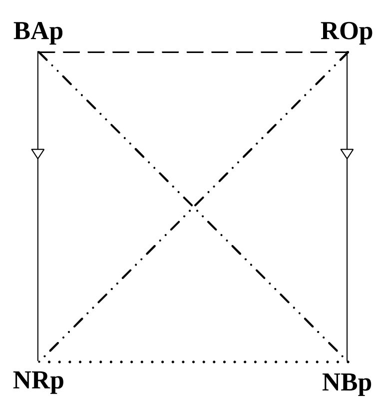
$PVp =_{df} \sim BAp \ \& \sim BA{\sim}p$

Clearly we have here another potential set of modal operators, and we can introduce rough analogues to our traditional definitional schemes for alethic modal operators and deontic operators as follows:

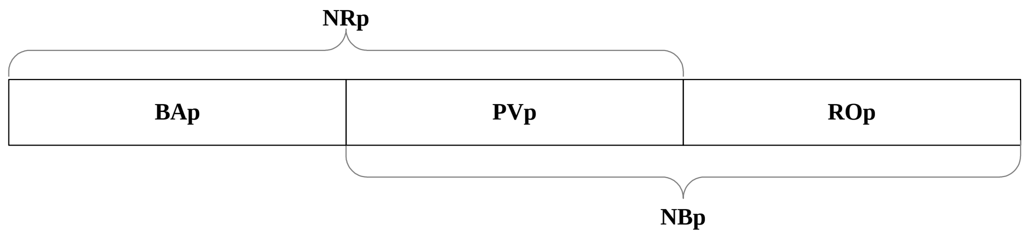
$ROp =_{df} BA{\sim}p$
$NRp =_{df} \sim BA{\sim}p$
$NBp =_{df} \sim BAp$
$PVp =_{df} \sim BAp \ \& \sim BA{\sim}p$

The first says that *it is ruled out by what our agent does that p* if and only if our agent brings it about that ~p. Note that this notion does not apply to all things that are ruled out per se, but only to those that are specifically ruled out *by* our agent's exercise of her agency. So contradictions, the negations of laws of nature and of past events are not ruled out *by* what our agent now does. The second says *it is not ruled out by anything our agent does that p* if and only if our agent does not bring it about that ~p. Laws of logic (which are necessarily ruled in) as well as contradictions (which are necessarily ruled out) are not things that are ruled out *by* our agent. The third says *our agent does not bring it about that p* (p is not ruled *in* by anything our agent does) if and only if it is false that our agent brings it about that p. This is of course compatible with p's holding for some other reason, such as that it is a law of logic or nature, or because it holds due to another person’s exercise of her own agency. The fourth says *our agent is passive regarding p* (does nothing herself that determines the status of p) if and only if our agent neither brings about p nor rules p out by what she does do (if anything). Again, it does not follow from the fact that our agent leaves something open that it is open per se. PVp is consistent with its being fixed that p and consistent with its being fixed that ~p, as long as it is not fixed *by* anything our agent has done. These notions are all intended to have a strong agential reading.

It is quite plausible to think that the first five agential operators satisfy the conditions of the traditional square and the traditional threefold classification scheme:



longer and more cumbersome expression.



For example, in the latter case, for every agent Jane Doe, and any proposition, p, either Doe brings about p, or Doe brings about ~p, or Doe brings about neither, and furthermore, no more than one of these three can hold (i.e. the three are mutually exclusive and jointly exhaustive). We will come back to this in a moment.

Virtually all accounts of this operator take it to satisfy the rule,

If $p \leftrightarrow q$ is a theorem, so is $BAp \leftrightarrow BAq$ (BA-RE),

as well as the scheme,

$BAp \rightarrow p$ (BA-T)

(if an agent brings about p, then p holds—"success" clause), and the schema,

$(BAp \;\&\; BAq) \rightarrow BA(p \;\&\; q)$ (BA-C)[76]

It is also the majority opinion that this operator satisfies this scheme:

$\sim BA\top$. (BA-NO)

Consider again:

1) BAp 1') $\sim BAp$
2) $BA{\sim}p$, 2') $\sim BA{\sim}p$,

and consider pairing these with one another. Pruning because of the commutativity of conjunction, we get six combinations:

a) $BAp \;\&\; BA{\sim}p$.	(Contradiction given BA-T axiom)
b) $BAp \;\&\; {\sim}BAp$.	(PC contradiction)
c) $BAp \;\&\; {\sim}BA{\sim}p$.	(The second clause is implied by the first)
d) $BA{\sim}p \;\&\; {\sim}BAp$.	(The second clause is implied by the first)
e) $BA{\sim}p \;\&\; {\sim}BA{\sim}p$.	(PC contradiction)
f) ${\sim}BAp \;\&\; {\sim}BA{\sim}p$.	(i.e. PVp)

[76] Where here we read the antecedent as implying that **BA**p and **BA**q both *now* hold.

Recall that because of the BA-T axiom, 1) above implies 2'), and 2) implies 1'). So the following three pruned down statuses for a proposition, p, and an agent, s, are the only pairs that remain of the six above (redundancies are also eliminated):

BAp,
BA~p,
PVp.

For the reasons alluded to already, it is easy to prove using the above principles that these three statuses (regarding an agent) and a proposition, p, are indeed mutually exclusive and jointly exhaustive, as anticipated.

Inaction versus Refraining/Forebearing: Another operator of considerable pre-theoretic interest is briefly discussed here. It can be defined via a condition involving embedding of "BA":

$$\text{RFp} =_{\text{df}} \text{BA}{\sim}\text{BAp}.$$

This expresses a widely endorsed analysis of refraining (or "forbearing")[77]. In quasi-English, *it is a case of Refraining by our agent that p* if and only if our agent brings it about that she does not bring it about that p. The importance of this in agency theory is based on the assumption that refraining from doing something is distinct from simply not doing something. In the current agential framework, the importance of the above is reflected in the denial of this claim:

$$*\text{:}\ \ {\sim}\text{BAp} \rightarrow \text{RFp}.$$

No agent brings about logical truths, but neither does an agent bring it about by what she does do that she doesn't bring about such logical truths. It has nothing to do with what she does. That * can't hold is easily proven given any consistent system with BA-RE and BA-NO.[78] So refraining from p is not equivalent to merely not bringing about p. Whether or not it is of great importance in deontic logic itself is a more controversial matter. It would hinge on matters like whether or not there is a difference between being obligated to not bring it about that p and being obligated to bring it about that you don't bring it about that p. For example, if it is true that the only things it can be obligatory for me to not bring about are things I can only not bring about by what I do bring about instead, then it would seem that I am obligated to not bring about p iff I am obligated to bring it about by what I do do that I do not bring it about that p. In this case, I would be obligated too not bring p about iff I am also obligated to bring it about that I don't bring it about that p.

An alternative account sometimes given of refraining is that of inaction coupled with ability: to refrain from bringing it about that p is to be *able* to bring it about that p and to not bring it about that p (von Wright 1963, on "forbearance"). This might be expressed as follows:

[77] It has been most utilized by Belnap and coworkers. See Belnap 2001, and its references to prior papers.

[78] Suppose S is any consistent system with BA-NO, BA-RE and PC: For *reductio* assume $\vdash {\sim}\text{BAp} \rightarrow \text{RFp}$. By BA-NO, $\vdash {\sim}\text{BA}\top$. So by our assumption, $\vdash \text{RF}\top$. Now by PC, $\vdash {\sim}\text{BA}\top \leftrightarrow \top$. So by BA-RE, $\vdash {\sim}\text{BA}{\sim}\text{BA}\top$. So by definition of RF, we have $\vdash {\sim}\text{RF}\top$, and hence an inconsistent set of theorems.

$RFp =_{df} \sim BAp \,\&\, ABp$,

where "AB" is interpreted as an agential *ability* operator, perhaps a compound operator of the broad form "◇BAp", with "◇" suitably constrained (e.g. as what is now still possible or still possible relative to our agent). In some frameworks, the two proposed analysans of RF are provably equivalent (e.g. Horty 2001)[79]. Informally one might argue that if I am able to bring it about that p and don't, then I don't bring it about that p by whatever it is that I do bring about, and so I refrain per the first analysis; and if I truly bring it about by what I do that I don't bring it about that p, then I must have been able to bring it about that p even though I didn't, so I refrain per the second analysis.

It is beyond the scope of this essay to delve non-superficially into the logic of agency[80], and here we can only barely touch on the more complex interaction of such logics with deontic logics by keeping the agency component exceedingly simple. Appendix A4 contains a brief sketch of a less abstract and more detailed influential framework for agency, STIT theory.

The Meinong-Chisholm Reduction for Agential Obligations (Chisholm 1964)[81]:

Let us set aside the jurisdictional problem as having established the need to go beyond SDL in order to represent agential obligations. Returning to deontic matters, the question arises: how do we represent not just agency, but agential obligation? With an agency operator in hand, we might now invoke the famous "Meinong-Chisholm Reduction": the idea that Jane Doe's obligation *to do* some thing is equivalent to what it is obligatory that Jane do (cf. what Jane ought to do is what it ought to be that Jane does). If we regiment this a bit using our operator for agency, we get the following versions of the "reduction":

Meinong-Chisholm Reduction: Jane Doe is obligated to bring it about that p iff it is obligatory that Jane Doe brings it about that p.

This is sometimes taken to be a reduction of personal obligation to impersonal obligation and agency (or it is sometimes rephrased as a reduction of the personal "ought to do" to the impersonal "ought to be" and agency).[82] Although not uncontested (e.g. see Horty 2001), by relying on this analysis we appear to have a way to represent the troublesome sentences, 1) and 2) of the jurisdictional problem:

1″) $\mathbf{OB}{\sim}BAp$,
2″) $\mathbf{OB}{\sim}BA{\sim}p$.

These might be taken to assert that Jane Doe is positively obligated to *not* bring it about that p

[79] But not so for the "achievement" agency operator in Belnap 2001.
[80] For example, see the following sources, and the references therein: Segerberg 1982; Elgesem 1993, 1997; Hilpinen 1997a, 1997b; Segerberg 1997; Belnap 2001.
[81] Meinong 1972 [1917]. Chisholm 1964 attributes the idea's endorsement to Nicolai Hartmann as well.
[82] More generally, it can be seen as a reduction of an agential deontic operator to a non-agential deontic operator (but not necessarily an impersonal one) and a non-deontic agency operator (Krogh and Herrestad 1996 and McNamara 2004a).

and that she is also positively obligated to *not* bring it about that ~p. Here we can properly express the fact that she is positively obligated to be non-agential with respect to the status of both p and ~p. These are easily distinguished from the claims that Jane Doe is *not* obligated to bring about p (i.e. ~**OB**BAp) and that she is *not* obligated to bring about ~p (i.e. ~**OB**BA~p). Similar remarks hold for our earlier equivalence **IM**p ↔ **OB**~p.

Generally, if we substitute "BAp" for p in the traditional definitional scheme's equivalences, we get:

IMBAp ↔ **OB**~BAp
PEBAp ↔ ~**OB**~BAp
GRBAp ↔ ~**OB**BAp
OPBAp ↔ ~**OB**BAp & ~**OB**~BAp

If we now read each deontic operator as "it is ____ for Jane Doe that", so that it is impersonal but not agential[83], the earlier problem with **IM**p ↔ **OB**~p, coupled with trying to read the agency into the deontic operators, disappears. For the deontic-agential compound above gets things right: it is impermissible that Jane Doe brings it about that your child is disciplined iff it is obligatory that she does not bring it about that your child is disciplined. We can now clearly and distinctly express the idea that something is simply out of Jane Doe's jurisdiction.

This general approach to obligations to do things has been very widely employed in deontic logic.[84] Recently, Krogh and Herrestad 1996 and McNamara 2004a reinterpret the analysis so that the deontic operator is *personal*, yet not agential. This is arguably a more plausible way to preserve a componential analysis of agential obligation. McNamara 2004a also makes the case that a person's being obligated to be such that a certain condition holds (e.g. being obligated to be at home at noon, as promised) is the more basic idiom, and being obligated to bring about something is just being obligated to be such that you do bring it about.

Appendix A4 contains a brief discussion of the Meinong–Chisholm Analysis in the context of STIT theory.

A Glimpse at the Theory of Normative Positions (Kanger 1971 [1957]):

One way in which this analysis has been fruitfully employed is in the study of what are called

[83] McNamara 2004a.

[84] For example, see Porn 1970; Kanger 1971 [1957]; Lindahl 1977; Porn 1977, 1989; Horty 1996; Jones and Sergot 1996; Santos and Carmo 1996; Belnap 2001, . As indicated earlier, Horty 1996 and Horty 2001 is of interest for (among other things) its argument against the Meinong-Chisholm reduction (see Appendix A4), and for providing an alternative non-componential analysis of agential obligation in the context of a branching-time analysis of agency. McNamara 2004b provides a critical exposition of the basic framework. This is in contrast to the branching-time approach to deontic contexts in Belnap 2001 (with Paul Bartha), where agential obligation is a componential compound of an agency operator and an obligation operator (one in turn analyzed via an Andersonian-Kangerian reduction). Another alternative to the major trend above, one that would unfortunately also take us too far afield, is the adaptation of modal logics for representing computer programs (e.g. dynamic logic) to represent actions in deontic logic. A classic sources here is Meyer 1988 which combines a dynamic logic approach to action with an adaptation of the Andersonian-Kangerian reduction to generate deontic notions. See also Segerberg 1982.

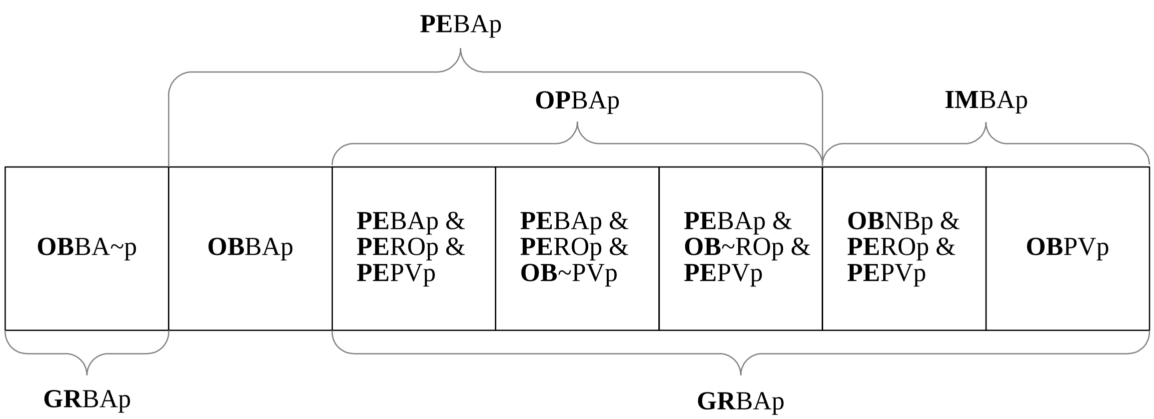
"normative positions". A set of normative positions is intended to describe the set of all possible mutually exclusive and jointly exhaustive positions that a person or set of persons may be in regarding a proposition and with respect to a set of selected primitive normative statuses and a set of agency operators. For a given propositions, p, recall the partition regarding how Jane Doe may be positioned agentially with respect to p:

$$(\mathrm{BA}p \vee \mathrm{RO}p \vee \mathrm{PV}p) \mathbin{\&} {\sim}(\mathrm{BA}p \mathbin{\&} \mathrm{RO}p) \mathbin{\&} {\sim}(\mathrm{BA}p \mathbin{\&} \mathrm{PV}p) \mathbin{\&} {\sim}(\mathrm{RO}p \mathbin{\&} \mathrm{PV}p)$$

Now also recall our partition with respect to obligations:

$$(\mathbf{OB}p \vee \mathbf{IM}p \vee \mathbf{OP}p) \mathbin{\&} {\sim}(\mathbf{OB}p \mathbin{\&} \mathbf{IM}p) \mathbin{\&} {\sim}(\mathbf{OB}p \mathbin{\&} \mathbf{OP}p) \mathbin{\&} {\sim}(\mathbf{IM}p \mathbin{\&} \mathbf{OP}p)$$

We might consider "merging" these two partitions, as it were, and try to get a representation of the possible ways Jane Doe may be positioned normatively with respect to her agency regarding p. Given certain choices of logic for BA and for **OB**, we might get a set of mutually exclusive and exhaustive "normative positions" for Jane Doe regarding p, the basic normative status, **OB**, and the basic agency operator, BA, such as that pictured below:



As usual, the partition above is intended to assert that the following seven classes are mutually exclusive and jointly exhaustive:

OBBA~p	
OBBAp	
PEBAp & **PE**ROp & **PE**PVp	(~**OB**~BAp & ~**OB**~BA~p & ~**OB**~PVp)
PEBAp & **PE**ROp & **OB**~PVp	(~**OB**~BAp & ~**OB**~BA~p & **OB**~PVp)
PEBAp & **OB**~ROp & **PE**PVp	(~**OB**~BAp & **OB**~BA~p & ~**OB**~PVp)
OBNBp & **PE**ROp & **PE**PVp	(**OB**~BAp & ~**OB**~BA~p & ~**OB**~PVp)
OBPVp	(i.e. **OB**NBp & **OB**NRp, given **OB**-C & **OB**-M)

The respective cases where it is permissible, impermissible, optional or gratuitous to bring about p are indicated as well.

Some Literature on the Theory of Normative Positions: The theory of normative positions has been a dynamic area of research that we have barely touched on here. It has been an important and active sub-area since its inception in Stig Kanger's seminal work (Kanger 1971 [1957], 1972), developed in a book-length study in Lindahl 1977, and thus sometimes referred to as "the Kanger-Lindahl theory". It has been used in attempts to analyze legal relations, like those made famous by Hohfeld 1919, among other things. The Kanger-Lindahl theory has been further developed by Jones and Sergot 1993, Sergot 1999, Herrestad and Krogh 1995 and Lindahl 2001. See also Allen 1996 for a somewhat different approach to Hohfeldian legal relations, and Porn 1970 and 1977 for a framework employed to analyze various normatively laden social positions and relations. Lindahl 2001 provides an excellent overview and orientation on Kanger's work in this area, and various problems informing subsequent research. (Other stunning contributions of Kanger to deontic logic are discussed in Hilpinen 2001b in the same volume.) Sergot 1999 takes the formal work of normative positions to a new level of abstraction and precision, and the later work mentioned above by Lindahl, and Herrestad and Krogh continue the exploration of refinements of the earlier Kanger-Lindahl conceptual framework to adapt it better to the analysis of legal notions.

Deontic Compliments: A further issue regarding the Meinong-Chisholm Reduction can be found here. One current issue in dispute is whether or not deontic operators call for agential complements or not. We outline the issue loosely here. Consider:

> Libertarian Deontic Compliment Thesis (LDCT): Any of the fundamental five deontic operators followed by *any* sentential compliment is well-formed.

Let an *LDCT system* be any classical sentential modal logic containing any of the above deontic operators (but at least **OB**) that satisfies LDCT. In contrast, consider the

> Strict Deontic Compliment Thesis (SDCT): Each fundamental deontic status must be followed immediately by an operator ascribing agency to an agent (here, by "BA") to be well-formed.

A *strict omission* is now a wff of the form RFp (i.e. BA~BAp). "~BAp" is just a *non-action. Strict deontic omissions* are deontic operators immediately followed by strict omissions.

Recall that if we substitute "BAp" for "p" in the equivalences associated with the Traditional Definitional Scheme's, we get:

IMBAp ↔ **OB**~BAp
PEBAp ↔ ~**OB**~BAp
GRBAp ↔ ~**OB**BAp
OPBAp ↔ ~**OB**BAp & ~**OB**~BAp

The instances above are all consistent with LDCT, but not SDCT. Essentially, non-action statements would have to be replaced by strict omissions. The needed replacements are given below with underlining stressing the trouble spots from the perspective of SDCT:

IMBAp ↔ **OB**~BAp	**IM**BAp ↔ **OB**BA~BAp
PEBAp ↔ ~**OB**~BAp	**PE**BAp ↔ ~**OB**BA~BAp
GRBAp ↔ ~**OB**BAp	(original is fine per SDCT)
OPBAp ↔ ~**OB**BAp & ~**OB**~BAp	**OP**BAp ↔ ~**OB**BAp & ~**OB**BA~BAp

Belnap 2001 provisionally endorses SDCT. McNamara 2004a raises doubts about SDCT. He notes that we are sometimes obligated to be a certain way (e.g. to be in our office), and furthermore, it is plausible to think that agential obligations reduce to this form—to obligations to be the agents of states of affairs, so that obligations to be a certain way are analytically prior to agential obligations.

An Obligation Fulfillment Dilemma (McNamara 2004a)[85]

Obligations can be fulfilled and violated. These are among the most characteristic features of obligations. It is often thought that fulfillment and violation conditions for what is obligatory are easily represented in SDL as follows:

OBp & p (fulfillment)
OBp & ~p (violation).

Call this the "Standard Analysis". Now consider cases where p is itself some agential sentence, say BAq, where we continue to read this as saying that Jane Doe brings it about that q. The Standard Analysis then implies:

OBBAq & BAq (fulfillment?)
OBBAq & ~BAq (violation?).

These suggest that Doe's obligation to bring it about that q is fulfilled iff she brings it about that q and is violated iff she doesn't. But if this is the proper analysis of obligation fulfillment, then it is hard to see how someone else could ever fulfill our obligations when we don't fulfill ours, for then our obligation would be unfulfilled and violated according to the Standard Analysis. Yet surely people can fulfill other people's obligations, and when they do so, it certainly seems to follow that our obligation is fulfilled. So the question then becomes, just what is obligatory? It would seem that it can't be that what is obligatory is that *Jane Doe* brings it about that p, for it is incoherent to say that someone else does that unless we mean that someone else gets Jane Doe to bring it about that p; but that is hardly the usual way in which we fulfill other's obligations. I might bring your book back to the library for you, thereby fulfilling one of your obligations without getting you to return the book yourself, at gunpoint say. So we face a dilemma:

> Since others can sometimes discharge our obligations, either *our* obligations are not always obligations for *us* to do things, and thus personal obligations need not be agential or obligation fulfillment is more complex than has been previously realized, and perhaps both (McNamara 2004a).

[85] This puzzle/dilemma is made explicit as such in McNamara 2004a, and one solution is there explored. However, the issue derives from Krogh and Herrestad 1996, who note that obligations can be yours yet fulfilled by someone else, and they use this distinction to offer a solution to the Leakage Problem below.

The Leakage Problem (Krogh and Herrestad 1996):[86] This is a problem closely related to the preceding one. As noted previously, when discussing two or more agents, subscripts are usually introduced to identify and distinguish the agents, for example BA_ip & BA_jq would indicate that i brings it about that p and j brings it about that q. Now let's assume that one agent can sometimes bring it about by what she does that another agent brings some thing about. For example, let's suppose that a parent can sometimes bring it about that a child brings it about that the child's room is cleaned (however rare this may in fact be). Carmo notes the following problem for the Meinong Chisholm analysis. Consider:

1) $BA_iBA_jp \rightarrow BA_jp$
2) $\mathbf{OB}BA_iBA_jp \rightarrow \mathbf{OB}BA_jp$

1) follows from BA-T, the virtually universally endorsed "success" condition for the intended agency operator. 1) is a logical truth. But then, in any system containing **OB**-RM, 2) will be derivable from 1), and so if 1) is a theorem in that system, 2) will be as well. But given the Meinong-Chisholm analysis, this will imply that if *I* am obligated to bring it about that some else does some thing, then she is obligated to do that thing as well. However, this is surely false. If I am obligated to get my very young child to feed herself, it does not follow that she is herself, at her young age, obligated to feed herself, even if she is just becoming capable of doing so. So it appears that the natural augmentation of SDL with an agency operator allows my obligation to implausibly "leak" beyond its proper domain and generate an obligation for her.

4.8 Challenges regarding Obligation, Change and Time

Although we have seen that obligations can be obligations to be (i.e. to satisfy a condition) as well as obligations to do, and that the former may be a special case of the latter, nonetheless, it is plausible to think that one is obligated *to do* something only if that thing is in the future. Thus even if attempts to solve Chisholm's contrary to duty paradox by invoking time do not look very plausible, this does not mean that there is no interesting work needed to forge relationships between time and obligations. For example, consider the system K*d*. If we read *d* atemporally as all obligations past, present, and future are met, then the only relevant worlds are those so ideal that in them there has never been a single violation of a mandatory norm. But as a parent, I may be obligated to lock the front door at night even though this would not be a norm unless there had been past violations of other norms (e.g. against theft and murder). People also acquire obligations over time, create them for themselves and for others by their actions, discharge them, etc.[87]

[86] Krogh and Herrestad 1996 attributes the identification of this problem to Jose Carmo. They offer a solution there by distinguishing between personal and agential obligations.

[87] Two classics on time and deontic logic are Thomason 1981b and Thomason 1981a, where temporal and deontic interactions are discussed, including an often invoked distinction between *deliberative 'ought's* (future-oriented/decision-oriented 'ought's) versus judgmental 'ought's (past, present or future oriented 'ought's from a purely evaluative, rather than action-oriented perspective). (Cf. the notion of "cues" for action in van Eck 1982.). Some other important earlier entries are Loewer and Belzer 1983, van Eck 1982, Åqvist and Hoepelman 1981, and Chellas 1969. For a sample of some recent work, see Bailhache 1998 and her references to her earlier work and that of others, as well as Brown 1996a for an attempt to develop a diachronic logic of obligations, representing obligations coming to be, and being discharged over time, where, for example, someone can now have an obligation

Conclusion

Plainly, there are a number of outstanding problems for deontic logic. Some see this as a serious defect; others see it merely as a serious challenge, even an attractive one. There is some antecedent reason to expect that the challenges will be great in this area. Normativity is challenging generally, not just in deontic logic. Normative notions appear to have strong semantic and pragmatic features. Normative notions must combine with notions for agency and with temporal notions to be of maximal interest--which introduces considerable logical complexity. There is also reason to think that there are hidden complexities in the interaction of normative notions and conditionals. Finally, there appears to be a wide array of normative notions with interesting interactions, some easily conflated with others (by ethicists as much as deontic logicians). Clearly, there is a lot of work to be done.

Appendices

A1. Alternative Axiomatization of SDL

The following alternative axiom system, which is provably equivalent to SDL, "breaks up" SDL into a larger number of "weaker parts" (SDL *a la carte*, as it were). This has the advantage of facilitating comparisons with other systems that reject one or more of SDL's theses in response to one or more of the problems discussed above.[88]

SDL':	A1.	All tautologous wffs of the language	(TAUT)
	A2′.	$\mathbf{OB}(p \,\&\, q) \rightarrow (\mathbf{OB}p \,\&\, \mathbf{OB}q)$	(**OB**-M)
	A3′.	$(\mathbf{OB}p \,\&\, \mathbf{OB}q) \rightarrow \mathbf{OB}(p \,\&\, q)$	(**OB**-C)
	A4.′	$\sim\mathbf{OB}\bot$	(**OB**-OD)
	A5′.	$\mathbf{OB}\top$	(**OB**-N)
	R1.	If $\vdash p$ and $\vdash p \rightarrow q$, then $\vdash q$	(MP)
	R2′.	If $\vdash p \leftrightarrow q$, then $\mathbf{OB}p \leftrightarrow \mathbf{OB}q$	(**OB**-RE)

We recall SDL for easy comparison:

SDL:	A1.	All tautologous wffs of the language	(TAUT)
	A2.	$\mathbf{OB}(p \rightarrow q) \rightarrow (\mathbf{OB}p \rightarrow \mathbf{OB}q)$	(**OB**-K)
	A3.	$\mathbf{OB}p \rightarrow \sim\mathbf{OB}\sim p$	(**OB**-D)
	MP.	If $\vdash p$ and $\vdash p \rightarrow q$ then $\vdash q$	(MP)
	R2.	If $\vdash p$ then $\vdash \mathbf{OB}p$	(**OB**-NEC)

Below is a proof that these two system are "equipollent": any formula derivable in the one is derivable in the other.

to bring about p only if p is (now) false.

[88] The interrelationships between the rules and axioms which constitute the equivalence between these systems is taken for granted in work on deontic logic, and is thus useful to know.

I. First, we need to prove that *each axiom (scheme) and rule of SDL′ can be derived in SDL*. A1 and R1 are common to both systems, so we need only show that A2′-A5′ and R2′ are derivable. Recall that **OB**-RM, and **OB**-RE (i.e. R2′) are derivable in SDL:

Show: If $\vdash$ p $\rightarrow$ q, then $\vdash$ **OB**p $\rightarrow$ **OB**q. (**OB**-RM)
Proof: Assume $\vdash$ p $\rightarrow$ q. By **OB**-NEC, $\vdash$ **OB** (p $\rightarrow$ q), and then by **OB**-K, $\vdash$ **OB**p $\rightarrow$ **OB**q.
Corollary: If $\vdash$ p $\leftrightarrow$ q then $\vdash$ **OB**p $\leftrightarrow$ **OB**q (R2′ or **OB**-RE)

So it remains to show A2′-A5' are derivable in SDL, and to do so we make free use of our already derived rules, **OB**-RM and **OB**-RE.

Show: $\vdash$ **OB**(p & q) $\rightarrow$ (**OB**p & **OB**q) (A2′ or **OB**-M)
Proof: By PC, $\vdash$ (p & q) $\rightarrow$ p. So by **OB**-RM $\vdash$ **OB**(p & q) $\rightarrow$ **OB**p. In the same manner, we can derive $\vdash$ **OB**(p & q) $\rightarrow$ **OB**q. From these two, by PC, we then get **OB**(p & q) $\rightarrow$ (**OB**p & **OB**q).

Show: $\vdash$ (**OB**p & **OB**q) $\rightarrow$ **OB**(p & q) (A3′ or **OB**-C)
Proof: By PC, $\vdash$ p $\rightarrow$ (q $\rightarrow$ (p & q)). So by **OB**-RM $\vdash$ **OB**p $\rightarrow$ **OB**(q $\rightarrow$ (p & q)). But by **OB**-K, we have $\vdash$ **OB**(q $\rightarrow$ (p & q)) $\rightarrow$ (**OB**q $\rightarrow$ **OB**(p & q)). So from these two, by PC,
$\vdash$ **OB**p $\rightarrow$ (**OB**q $\rightarrow$ **OB**(p & q)), which is equivalent by PC to $\vdash$ (**OB**p & **OB**q) $\rightarrow$ **OB**(p & q).

Show: $\vdash$ ~**OB**$\bot$ (A4′ or **OB**-OD)
Proof: (By reductio) Assume **OB**$\bot$. Since by PC, $\vdash$ $\bot$ $\leftrightarrow$ (p & ~p), by **OB**-RE, we get $\vdash$ **OB**$\bot$ $\leftrightarrow$ **OB**(p & ~p). So from this and our assumption, we get **OB**(p & ~p). Given **OB**-M, this yields **OB**p & **OB**~p, and then from A3 of SDL, we get **OB**p & ~**OB**~p, a contradiction. So $\vdash$ ~**OB**$\bot$.

Show: $\vdash$ **OB**$\top$ (A5′ or **OB**-N)
Proof: By PC, $\vdash$ $\top$. So By **OB**-NEC, we have $\vdash$ **OB**$\top$.

II. It remains for us to show that *each axiom (scheme) and rule of SDL can be derived in SDL′.* Again, A1 and R1 are common to both systems, so we need only show that A2, A3 and R2 are of SDL are in SDL′. It will be useful (but not necessary) to first show that **OB**-RM is derivable in SDL′, and then show the remaining items.

Show: If $\vdash$ p $\rightarrow$ q, then $\vdash$ **OB**p $\rightarrow$ **OB**q. (**OB**-RM)
Proof: Assume $\vdash$ p $\rightarrow$ q. By PC, it follows that $\vdash$ p $\leftrightarrow$ (p & q). So by R2′, $\vdash$ **OB**p $\leftrightarrow$ **OB**(p & q), and so by PC, $\vdash$ **OB**p $\rightarrow$ **OB**(p & q). But by A2′, $\vdash$ **OB**(p & q) $\rightarrow$ (**OB**p & **OB**q). So from the last two results, by PC, $\vdash$ **OB**p $\rightarrow$ (**OB**p & **OB**q), and thus $\vdash$ **OB**p $\rightarrow$ **OB**q.

Corollary: **OB**-RM is inter-derivable with **OB**-RE + **OB**-M.
This follows from the preceding proof and the first proof in the showing that SDL contains SDL′.

Show: $\vdash$ **OB**(p $\rightarrow$ q) $\rightarrow$ (**OB**p $\rightarrow$ **OB**q) (A2 or **OB**-K)
Proof: By PC, $\vdash$ ((p $\rightarrow$ q) & p) $\rightarrow$ q. So by **OB**-RM, $\vdash$ **OB**((p $\rightarrow$ q) & p) $\rightarrow$ **OB**q. But by A2′ conjoined with A3′, we get $\vdash$ **OB**((p $\rightarrow$ q) & p) $\leftrightarrow$ (**OB**(p $\rightarrow$ q) & **OB**p). So from the last two

results, by PC, we get $\vdash$ (**OB**(p $\rightarrow$ q) & **OB**p) $\rightarrow$ **OB**q, and thus $\vdash$ (**OB**(p $\rightarrow$ q) $\rightarrow$ (**OB**p $\rightarrow$ **OB**q).[89]

<u>Show</u>: $\vdash$ **OB**p $\rightarrow$ ~**OB~**p (A3 or **OB**-D)
<u>Proof</u>: (Reductio) Assume ~(**OB**p $\rightarrow$ ~**OB~**p). By PC, (**OB**p & **OB~**p). So by A3′, **OB**(p & ~p), which is equivalent, by R2′ to **OB**$\bot$, which contradicts A4′.

<u>Show</u>: If $\vdash$ p then $\vdash$ **OB**p (R2 or **OB**-NEC)
<u>Proof</u>: Assume $\vdash$ p. From this by PC, it follows that $\vdash$ p $\leftrightarrow$ $\top$. So by R2′, $\vdash$ **OB**p $\leftrightarrow$ **OB**$\top$. But then from A5′, we get $\vdash$ **OB**p. So if $\vdash$ p then $\vdash$ **OB**p.

<u>A2. A Bit More on Chisholm's Paradox</u>

Recall the quartet and its most natural symbolization in SDL:

1) It ought to be that Jones <u>g</u>oes to assist his neighbors.	1′) **OB**g.
2) It ought to be that if Jones goes, then he <u>t</u>ells them he is coming.	2′) **OB**(g $\rightarrow$ t).
3) If Jones doesn't go, then he ought not tell them he is coming.	3′) ~g $\rightarrow$ **OB**~t.
4) Jones doesn't go.	4′) ~g.

There is a general point to be made regarding the key inferences that generate the paradox per the above symbolization. There is a sense in which the inference from 1′) and 2′) to **OB**t and the inference from 3′) & 4′) to **OB**~t involve "detachment" of an obligation from a pair of premises, one of which involves a deontic conditional in some way. Let us introduce a bit of regimentation. Let

"**OB**(q/p)"

represent a shorthand for a conditional obligation or ought statement like that in the natural language sentence, 3), above.[90] So we will read **OB**(q/p) as "if p, then it ought to be (or it is obligatory) that q", in the manner of 3) above. Suppose we also assume, as almost all have[91], that monadic obligations are disguised dyadic obligations, per the following analysis:

OBp $=_{df}$ **OB**(p/$\top$).

---

[89] A direct proof of A2 without first proving RM is:

<u>Show</u>: $\vdash$ **OB**(p $\rightarrow$ q) $\rightarrow$ (**OB**p $\rightarrow$ **OB**q) (A2 or **OB**-K)
<u>Proof</u>: By PC, $\vdash$ ((p $\rightarrow$ q) & p) $\leftrightarrow$ (p & q). So by R2′, $\vdash$ **OB**((p $\rightarrow$ q) & p) $\leftrightarrow$ **OB**(p & q). But by A2′ conjoined with A3′, we get $\vdash$ **OB**((p $\rightarrow$ q) & p) $\leftrightarrow$ (**OB**(p $\rightarrow$ q) & **OB**p). So from the last two results, by PC, we get $\vdash$ (**OB**(p $\rightarrow$ q) & **OB**p) $\leftrightarrow$ **OB**(p & q), and thus $\vdash$ (**OB**(p $\rightarrow$ q) & **OB**p) $\rightarrow$ **OB**(p & q). But by A2′, we have $\vdash$ **OB**(p & q) $\rightarrow$ (**OB**p & **OB**q). So from the last two results, by PC, we get $\vdash$ (**OB**(p $\rightarrow$ q) & **OB**p) $\rightarrow$ (**OB**p & **OB**q), and thus $\vdash$ (**OB**(p $\rightarrow$ q) $\rightarrow$ (**OB**p $\rightarrow$ **OB**q).

[90] We continue to ignore the differences between "obligation" and "ought" for simplicity.

[91] Alchourron 1993 is a salient exception.

With this in mind we distinguish between two relevant types of "detachment principles"[92] that we might ascribe to these iffy-'ought's:

Factual Detachment (FD): p & **OB**(q/p) .→ **OB**q
Deontic Detachment (DD): **OB**p & **OB**(q/p) .→ **OB**q

Factual detachment tells us that from the fact that p, and the deontic conditional to the effect that if p then it ought to be that q, we can conclude that it ought to be that q. *Deontic Detachment* in contrast tells us that from the fact that it ought to be that p and that if p, then it ought to be that q, we can conclude that it ought to be that q. If we interpret a deontic conditional as a material conditional with an obligatory consequent (as in 3′) above), FD, but not DD is supported. Conversely, if we interpret deontic conditionals as obligatory material conditionals (as in 2′) above), DD, but not FD is supported.[93] Although we have shown earlier that neither of these interpretations is acceptable, the contrast reveals a general problem. *Carte blanche e*ndorsement of both types of detachment (without some restriction) is not tenable, since it leads implausibly to the conclusion that we are both obligated to tell (the neighbor we are coming) and obligated to not tell. Thus researchers tended to divide up over which principle of the two they endorse (Loewer and Belzer 1983). The Factual Detachment camp typically endorses the view that the conditional in 3) in the Chisholm Quartet needs to be interpreted as a non-material conditional, but otherwise things are as they seem in 3): we have a conditional obligation that is a simple composite of a non-deontic conditional and a pure unary deontic operator in the consequent:

OB(q/p) $=_{df}$ p ⇒ **OB**q, for some independent conditional.[94]

---

[92] Greenspan 1975.
[93] As already noted, some reject both analyses and think deontic conditionals are *sui generis*. Note also that 2) above has the conditional explicitly in the scope of the English "ought to be" operator, and this is not explicitly a deontic conditional as just characterized unless we add that it should be read as at least necessarily equivalent to "if Jones does go, then he ought to tell them he is coming". There is no uniform agreement about this, although often the Chisholm Paradox is characterized so that both 2) and 3) above would have the same superficial form ("if …, then it ought to be that…"), with the deontic term appearing in the second clause. We have instead followed Chisholm's original formulation. In either event, the inference from 1) and 2) to "it ought to be that Jones tells" is also called "deontic detachment" as is that from their formal analogues in SDL, where **OB**-K validates the inference from 1′) and 2′) to **OB**t.
[94] Smith 1994 notes that adding factual detachment to SDL with **OB**(q/p) interpreted as **OB**(p → q), yields Mally's problem: ⊢ **OB**p ↔ p. That SDL yields the first half, p → **OB**p, given factual detachment, is easily seen. Just substitute p for q in FD to yield ⊢ p & **OB**(p → p) .→ **OB**p. Then, since ⊢ **OB**(p → p) by **OB-**N, it can fall out and we get ⊢ p → **OB**p. Note that the proof depends crucially on the highly controversial rule of necessitation. However, Smith, crediting Andrew Jones, pointed out that even a very minimal deontic logic entails the second half of the equivalence in question, **OB**p → p, which is still enough to make Voltaire grin.

Thm: For any system with PL, **OB**-D & **OB**-RE, FD yields ⊢ **OB**p → p.
Proof: Assume PL, **OB**-D, **OB**-RE and FD. By substitution of ⊥ for q in FD, we get ⊢ p & **OB**(p → ⊥) → **OB**⊥. So from that and **OB**-D, we get ⊢ ~(p & **OB**(p → ⊥)), that is ⊢ ~p ∨ ~**OB**(p → ⊥). From the latter by **OB**-RE we get ⊢ ~p ∨ ~**OB**~p, that is ⊢ **OB**~p → ~p, which by substitution of ~p for p, along with **OB**-RE, yields ⊢ **OB**p → p.

Given how minimal **OB**-RE and **OB**-D are, the friend of factual detachment with conditionals so interpreted cannot shrug this off. Factual attachment with a material conditional would license the following: Since I steal, and it ought

Typically, the conditional was a non-classical conditional of the sort made famous by Stalnaker and Lewis.[95] It is then generally maintained that deontic detachment is flawed, since the conditional obligations like those in 2) tell us only what to do in ideal circumstances, but they do not necessarily provide "cues"[96] for action *in the actual world*, where things are often typically quite sub-ideal, as 4) combined with 1) indicate. Thus from the fact that Jones ought to go and he ought to tell if he goes, it doesn't follow that what he ought to *actually* do is tell—that would be so only if it was also a fact that he goes to their aid. At best, we can only say that he ought *ideally* to go.

This suggestion seems a bit more difficult when we change the conditional to something like "If Doe does kill his mother, then it is obligatory that Doe kills her gently". The idea that my obligation to not kill my mother gently (say for an inheritance) merely expresses an "ideal" obligation, but not an actual obligation, given that I will kill her, seems hard to swallow. So this case makes matters a bit harder for those favoring a factual detachment approach for generating actual obligations. Similarly, it would seem that if it is impermissible for me to kill my mother, then it is impermissible for me to do so gently, or to do so while dancing.[97] So *carte blanche* factual detachment seems to allow the mere fact that I will take an action in the future (killing my mother) that is horribly wrong and completely avoidable now to render obligatory another horrible (but slightly less horrible) action in the future (killing my mother gently). The latter action must be completely avoidable if the former is, and the latter action is one that I would *seem* to be equally obligated to not make intuitively.

The main alternative camp represented conditional obligations via dyadic non-composite obligation operators modeled syntactically on conditional probability. They rejected the idea that **OB**(q/p) $=_{df}$ p $\Rightarrow$ **OB**q, for some independent conditional. In a sense, on this view, deontic conditionals are viewed as idioms: the meaning of the compound is not a straightforward function of the meaning of the parts. The underlying intuition regarding the Chisholm example is that even if it might be true that we will violate some obligation, that doesn't get us off the hook from obligations that derive from the original one that we will violate. If I must go help and I must inform my neighbors that I'm coming, if I do go help, then I must inform them, and the fact that I will in fact violate the primary obligation does not block the derivative obligation anymore than it does the primary one itself.

One early semantic picture for the latter camp was that a sentence of the form **OB**(q/p) is true at a world i iff the i-best p-worlds are all q-worlds. **OB**q is then true iff **OB**(q/$\top$), and so iff all the unqualifiedly i-best worlds are q-worlds (Hansson 1969). Note that this weds preference-based semantic orderings with dyadic conditional obligations.[98] This reflects a widespread trend.

---

to be that either I don't steal or the world blows up, then it ought to be that the world blows up. Clearly a non-material conditional is essential here.

[95] See Mott 1973, Chellas 1974, 1980 for examples, and DeCew 1981 for an influential critical evaluation, arguing that although such conditionals are indeed important, there is still a special conditional they overlook at the heart of the Chisholm puzzle.

[96] van Eck 1982.

[97] In Chisholm's example it is easier to accept that telling is merely *ideal*, but not required, since it is easy to interpret Chisholm's example as one where giving advanced notice is what the agent perhaps *ought* to do, but not something the agent *must* do (even assuming the neighborly help is itself a must).

[98] As Makinson 1993 notes, it was also a forerunner of semantics for defeasible conditionals generally (cf. "if p,

Factual detachment does not work in this case, since even if our world is an I-don't-go-help-world, and the best among the I-don't-go-help-worlds are I-don't-call-worlds, it does not follow that the *unqualifiedly* best worlds are I-don't-call-worlds. In fact, in this example, these folks would maintain, the unqualifiedly best worlds are both I-go worlds and I-call worlds, and the fact that I won't do what I'm supposed to do won't change that.

But one is compelled to ask those in the Deontic Detachment Camp: what then is the point of such apparent conditionals if we can't ever detach them from their apparent antecedents, and how are these conditionals related to regular ones? This seems to be the central challenge for this camp. Thus they often endorse a restricted form of factual detachment, of which the following is a representative instance:

Restricted Factual Detachment: $\Box$p & **OB**(q/p) .→ **OB**q.

Here $\Box$p might mean various things, for example that p is physically unalterable or necessary as of this moment in history.[99] Only if p is settled true in some sense, can we conclude from **OB**(q/p) that **OB**q. This certainly helps, but it still leaves us with a bit of a puzzle about why this apparent composite of a conditional and a deontic operator is actually some sort of primitive idiom.

So it seems like we are left with a dilemma: either 1) you allow factual detachment and get the consequences earlier noted to the effect that simply because someone will act like a louse, he is obligated to do slightly mitigating louse-like things, or 2) instead you claim that "if p, then ought q" is really an idiom, and the meaning of the whole is not a function of the meaning of its conditional and deontic parts. Each seems to be a conclusion one would otherwise prefer to avoid.

There have been many attempts to try to solve Chisholm's problem by carefully distinguishing the times of the obligations.[100] This was fueled in part by shifts in the examples, in particular to examples where the candidate "derived" obligations were clearly things to be done after the primary obligation was either fulfilled or violated (called "forward" versions of CTDs). This made the ploy of differentiating the times and doing careful bookkeeping about just which things were obligatory at which times promising. However, Chisholm's own example is most plausibly interpreted as either a case where the obligation to go help and the perhaps-derivable obligation to tell are *simultaneous* (called "parallel" versions), or where telling is even something to be done *before* you go (called "backward" versions).[101] It is easy to imagine that the way to tell the neighbors that you will help might be to phone, and that would typically take place before you left to actually help. (For younger readers: there were no cell phones back in 1964, and phones were attached to boxes in houses by

---

normally q").

[99] The idea is perhaps implicit in Hansson 1969; it is argued for explicitly in Greenspan 1975, and adopted by many since.

[100] Thomason 1981b, 1981a are classics arguing for the general importance of layering deontic logic on top of temporal logic. Åqvist and Hoepelman 1981; Thomason 1981b; van Eck 1982; Loewer and Belzer 1983; Feldman 1986 argue that attention to time is crucial (or at least helpful) in handling the Chisholm puzzle, among other puzzles.

[101] DeCew 1981. Smith 1994 contains an illuminating discussion of the three different versions (backward, parallel, and forward) in evaluating different approaches to solving the Chisholm paradox.

yard-length coiled chords.) Concerns to coordinate aid, or to assure those stressed that aid is coming, often favor giving *advanced* notice.

Alternatively, it was suggested that carefully attending to the action or agential components of the example and distinguishing those from the circumstances or propositional components would dissolve the puzzle.[102] However, the phenomena invoked in the Chisholm example appear to be too general for that. Consider the following non-agential minor variant of an example (say of possible norms for a residential neighborhood) introduced in Prakken and Sergot 1996:

1) It ought to be the case that there are no dogs.
2) It ought to be the case that if there are no dogs, then there are no warning signs.
3) If there are dogs, then it ought to be the case that there are warning signs.
4) There are dogs.

Here we seem to have the same essentially puzzling phenomena present in Chisholm's original example, yet there is no apparent reference to actions above at all; instead the reference seems to be to states of affairs only. (Notice also that there is no issue of different times either for the presence/absence of dogs and the presence/absence of signs.)

Thus, it looks like tinkering with the temporal or action aspects of the Chisholm-style examples (however much time and action are important elsewhere to deontic logic) merely postpones the inevitable. So far, this problem appears to be not easy to convincingly solve.

A formal sketch of a sample system favoring factual detachment can be easily found in chapter 10 of Chellas 1980, which is widely available. A system that favors deontic detachment over factual detachment is quickly sketched in the following box (see Goble Forthcoming-b).

Here, we assume a classical propositional language now extended with a dyadic construction, **OB**(/), taken as primitive. A monadic **OB** operator is then defined in the manner mentioned above:

$\mathbf{OB}p =_{df} \mathbf{OB}(p/\top)$.

We can define an ordering relation between propositions as follows:

$p \geq q =_{df} {\sim}\mathbf{OB}({\sim}p/p \vee q)$.

This says that p is ranked as at least as high as q iff it is not obligatory that ~p on the condition that either p or q.[103]

An axiom system that is a natural dyadic correlate to SDL follows:

A1: All instances of PC tautologies (TAUT)

[102] Castañeda 1981; Meyer 1988.
[103] Cf. p is permissible given $p \vee q$, where $\mathbf{PE}(p/q) =_{df} {\sim}\mathbf{OB}({\sim}p/q)$.

A2: $\mathbf{OB}(p \to q/r) \to (\mathbf{OB}(p/r) \to \mathbf{OB}(q/r)$ (**OB**-CK)
A3: $\mathbf{OB}(p/q) \to \sim\mathbf{OB}(\sim p/q)$ (**OB**-CNC)
A4: $\mathbf{OB}(\top/\top)$ (**OB**-CN)
A5: $\mathbf{OB}(q/p) \to \mathbf{OB}(q \,\&\, p/p)$ (**OB**-CO&)
A6: $(p \geq q) \,\&\, q \geq r) \to p \geq r$ (Trans)
R1: If $\vdash p$ and $\vdash p \to q$ then $\vdash q$ (MP)
R2: If $\vdash p \leftrightarrow q$ then $\vdash \mathbf{OB}(r/p) \leftrightarrow \mathbf{OB}(r/q)$ (**OB**-CRE)
R3: If $\vdash p \to q$ then $\vdash \mathbf{OB}(p/r) \to \mathbf{OB}(q/r)$ (**OB**-CRM)

A1-A4 and R1-R3 are conditional analogues of formulas or rules we have seen before in discussing axiomatizations of SDL itself. A5 and A6 are needed to generate a complete system relative to an ordering semantics of the following sort (merely sketched here).

Assume we have a set of worlds and a set of ordering relations, P_i, for each world, i, where jP_ik is to be interpreted as saying that relative to i's normative standards, j is at least as good as k. Assume also that all of the ordering relations are non-empty: for each world i, there is a world k and a world m such that either kP_im or mP_ik. Call this structure a "preference frame". For any preference relation in a preference frame, let $F(P_i)$ represent the field of that relation: the set of all worlds that appear in some ordered pair constituting the relation, P_i. As usual, a model on a frame is an assignment to each propositional variable of a set of worlds (those where it will be deemed true). We then define the basic dyadic operators truth-condition as follows:

$M \vDash_i \mathbf{OB}(q/p)$ iff there is a j in $F(P_i)$ such that $M \vDash_j p \,\&\, q$ and for each k such that kP_ij, if $M \vDash_k p$, then $M \vDash_k q$.

That is, at i, it is obligatory that q given p iff there is some world j in the field of the i's preference relation where both p and q are true, and for every world ranked at least as high j, if p is true at that world, then so is q.

Call a preference frame standard iff all the preference relations in it are connected (and thus reflexive) and transitive relative to their fields:

For each i-relative preference relation, P_i,

1) if j and k are in $F(P_i)$, then either jP_ik or kP_ij (connectedness).
2) if j, k, and m are in $F(P_i)$, then if jP_ik and kP_im, then jP_im (transitivity).

Goble Forthcoming-b shows that the axiom system for dyadic obligation above is sound and complete for the set of standard preference frames. It is also easy to derive SDL using the above dyadic axiom system and the definition given for the monadic obligation operator. Goble's paper contains a number of other such results, for both monadic and dyadic systems, including generalizations that allow for conflicting obligations.

A3. Doing Well Enough (DWE)[104]

A3.1 DWE Syntax

Assume that we have a language of classical propositional logic with these additional (personal but non-agential) *primitive* unary operators:

OBp: It is *Obligatory* (for S) that p
MAp: The *Maximum* (for S) involves p
MIp: The *Minimum* (for S) involves p
INp: It is *Indifferent* (for S) that p

We might then tentatively analyze some other agential deontic notions as follows:

"S *must* bring it about that p": **OB**BA**p**
"S *ought* to bring it about that p": **MA**BAp
"The *least* S can do involves bringing it about that p": **MI**BAp
"It is a matter of indifference for S to bring it about that p ": **IN**BAp.

Suppose that I am obligated to contact you to conduct some business, and that I can do so by emailing you, calling you, or stopping by. Add that these are the only ways to conduct the business.[105] Now imagine that the morally relevant value of these actions matches the extent to which the response is personal. Assuming you would not let me conduct our business twice, the three alternatives are exclusive. Then it is *obligatory* for me that I contact you in one of the three ways, but no one in particular, since any one of the three will discharge my obligation to contact you. Now if I choose to discharge my obligation in the minimally acceptable way, I will do so by email rather than by telephone or in person. So *doing the minimum involves* emailing you. On the other hand, if I conduct the business in person, I will have discharged my obligation in the optimal way. *Doing the maximum (what morality recommends) involves* stopping by your place. Finally, we can easily imagine that nothing of moral worth hinges on whether I wear my black socks when I contact you. So wearing them is *a matter of moral indifference*. This illustrates one application of the four primitive operators.

We introduce some defined operators, and their intended readings:

PEp $=_{df}$ ~**OB**~p.	(It is *Permissible* for S that p.)
IMp $=_{df}$ **OB**~p.	(It is *Impermissible* for S that p.)
GRp $=_{df}$ ~**OB**p.	(It is *Gratuitous* for S that p.)
OPp $=_{df}$ ~**OB**p & ~**OB**~p.	(It is *Optional* for S that p.)
SIp $=_{df}$ ~**IN**p.	(It is *Significant* for S that p.)
SUp $=_{df}$ **PE**p & **MI**~p.	(It is *Supererogatory* for S that p.)
PSp $=_{df}$ **PE**p & **MA**~p.	(It is *Permissibly Suboptimal* for S that p.)

[104] We presuppose a simple no-conflicts atmosphere.
[105] To minimize complications, we will assume no one else can do these things for me.

Continuing with our example, note that although the three alternatives, conducting the business by email, phone, or in person, are not on a par morally speaking, each is still *morally optional*. For each, the agent is permitted to do it or to refrain from doing it. Now we saw that doing the minimum involves e-mailing you. But suppose that rather than e-mailing you, I either call or stop by. Both of the latter alternatives are *supererogatory*. In each case, I will have done *more than I had to do*--more good than I would have if I had done the minimum permitted.[106] On the other hand, if I do not stop by, I will have done something sup-optimal, but, since emailing you and calling you are each nonetheless permissible, each is *permissibly suboptimal*. Finally, although each of the three ways of contacting you is optional, none is without *moral significance*. For whatever option I take of the three, I will have done something supererogatory or I will have done only the minimum; in either case, I will have done something with *moral significance*.

Where "*****" ranges over **OB**, **MA**, **MI**, the associated **DWE Logic** is:

A0. All tautologous DWE-wffs;
A1. $*(p \to q) \to (*p \to *q)$
A2. $\mathbf{OB}p \to (\mathbf{MI}p \ \& \ \mathbf{MA}p)$
A3. $(\mathbf{MI}p \vee \mathbf{MA}p) \to \mathbf{PE}p$
A4. $\mathbf{IN}p \to \mathbf{IN}{\sim}p$
A5. $\mathbf{IN}p \to ({\sim}\mathbf{MI}p \ \& \ {\sim}\mathbf{MA}p)$
A6. $(\mathbf{OB}(p \to q) \ \& \ \mathbf{OB}(q \to r) \ \& \ \mathbf{IN}p \ \& \ \mathbf{IN}r) \to \mathbf{IN}q$

R1: If $\vdash p$ and $\vdash p \to q$ then $\vdash q$
R2: If $\vdash p$, then $\vdash \mathbf{OB}p$.

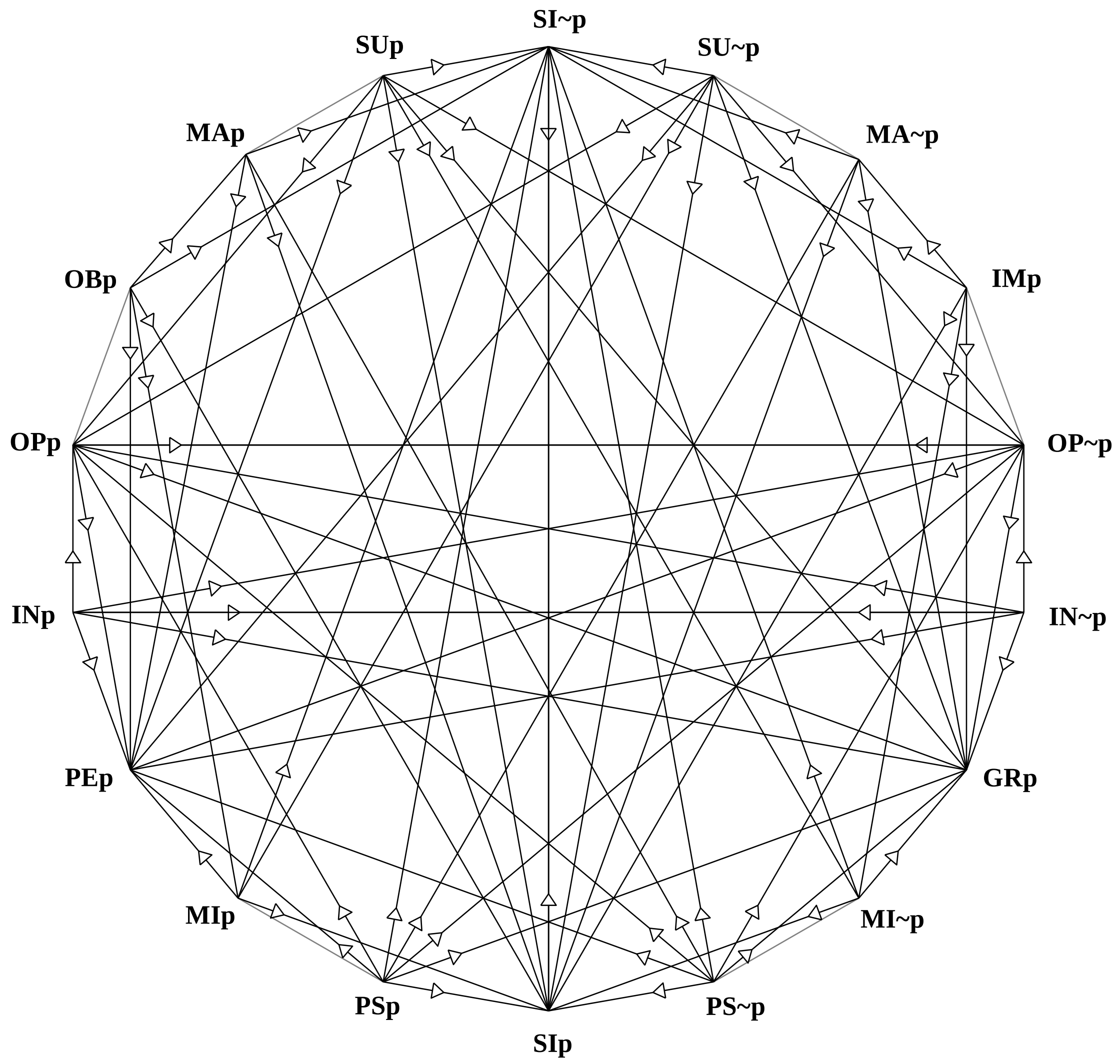
It is easily shown that SDL logics for **OB**, **MA**, and **MI** are derivable from DWE (McNamara 1996c).

The increased complexity brought on by the enriched expressive power is graphically reflected in analogues to SDL's deontic hexagon and threefold partitions. [107]

[106] I slough over subtleties here about different sense of the philosopher's term "supererogatory".
[107] Recall our prior scheme for diagrams:

<u>Arrowed Lines</u> represent implications	<u>Dotted Line</u>: connects sub-contraries
<u>Dashed Line</u>: connects contraries	<u>Dotted-Dashed Lines</u>: connect contradictories

The Deontic Octodecagon--Part I

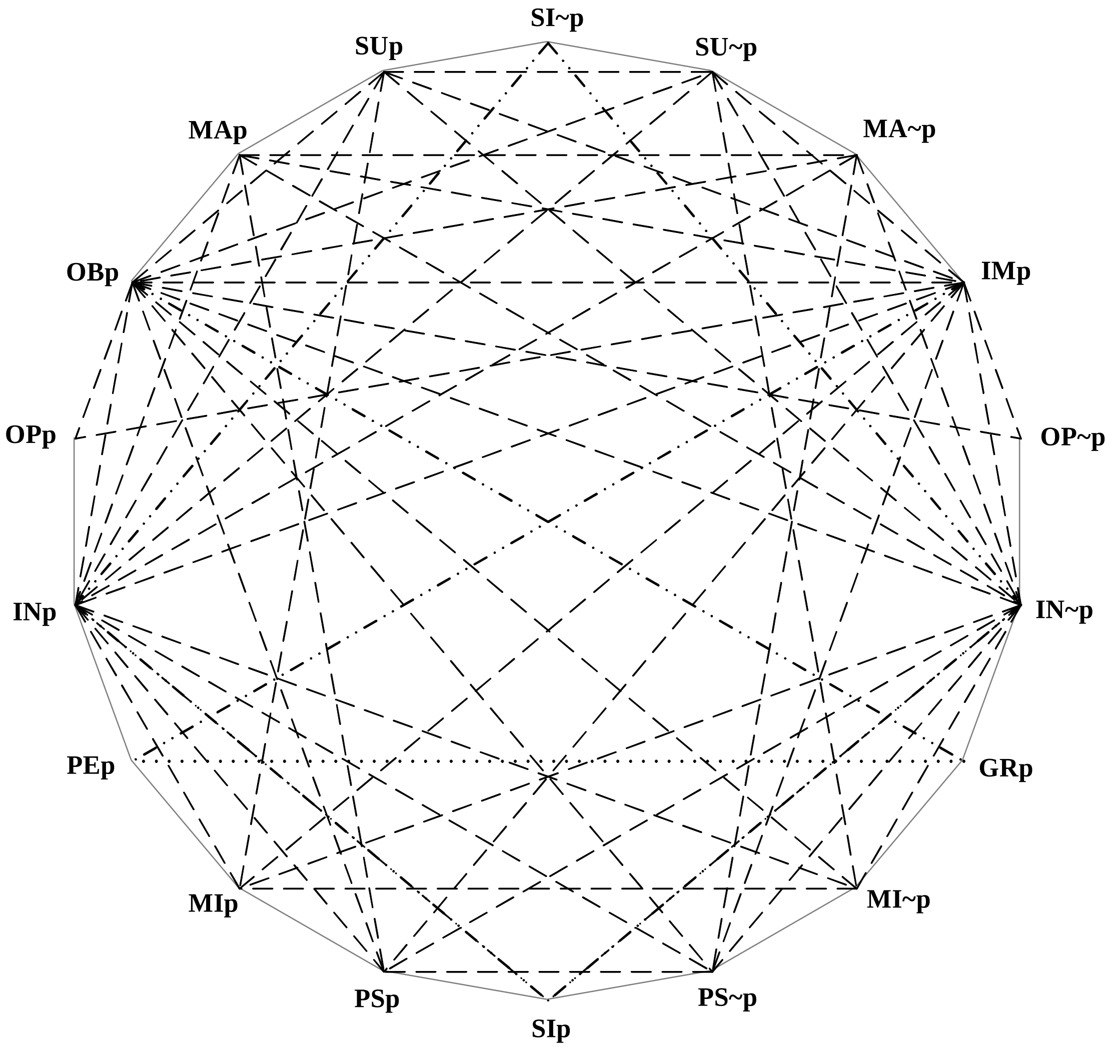


<u>Grayed Plain Border Lines</u>: added for purely aesthetic reasons.

<u>Operator Key</u>:

OBp: it is obligatory that p (cf. "must").
PEp: it is permissible that p (cf. "can").
IMp: it is impermissible that p (cf. "can’t"),
GRp: it is gratuitous that p.
OPp: it is optional that p.
MAp: doing the maximum involves p (cf. "ought").
MIp: doing the minimum involves p (cf. "the least one can do involves")
SUp: it is supererogatory that p (cf. "exceeding the minimum).
PSp: it is permissibly suboptimal that p (cf. "you can, but ought not").
INp: it is indifferent that p.
SIp: it is significant that p.

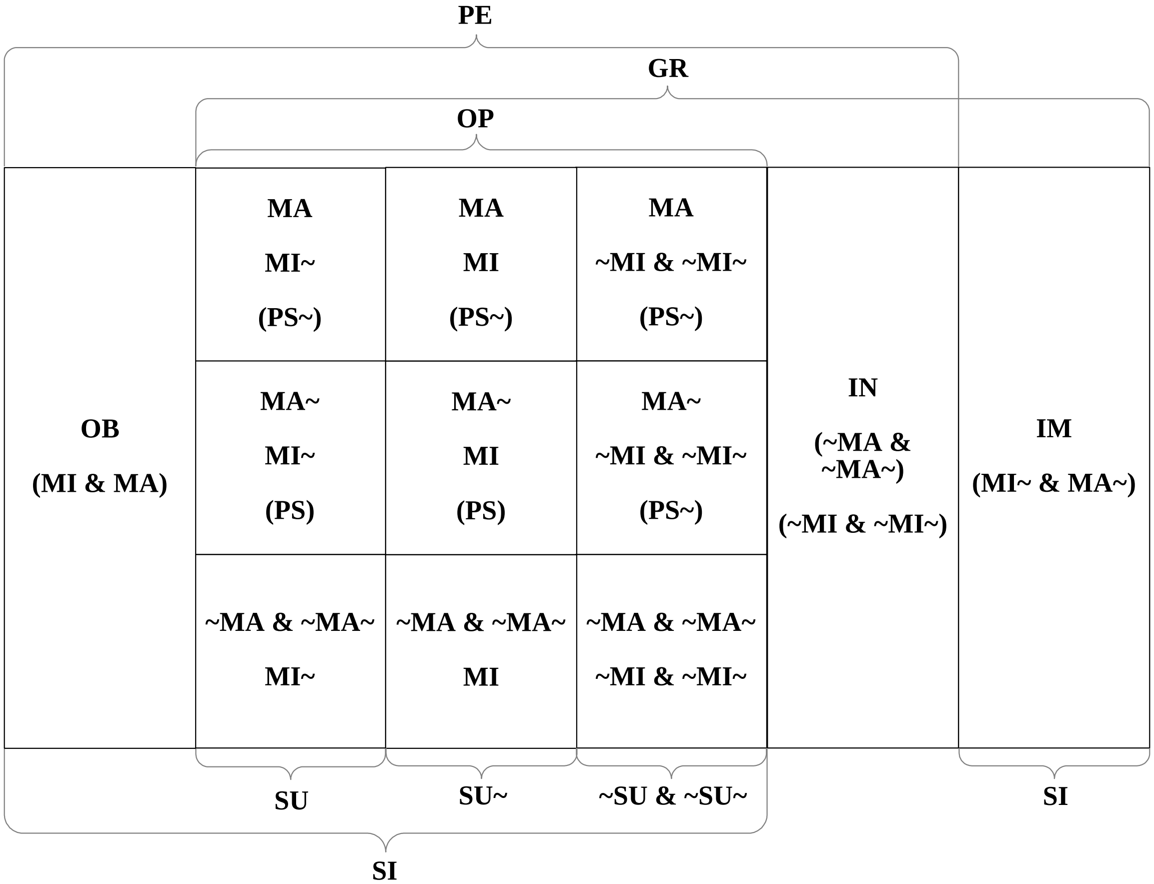
The Deontic Octodecagon--Part II



The Deontic Octodecagon is the result of the superimposition of Part II on Part I.[108]

[108] Roderick Chisholm brought my attention to the similarity between these diagrams, and those in Hrushka and Joerden 1987. I began creating a series of diagrams expanding on the deontic square in the early 1980s prompted by remarks from Fred Feldman in an ethical theory class with him at the University of Massachusetts.

The Twelvefold Partition



The partition is drawn with the black lines. As with the Traditional Threefold Classification, the twelve cells are mutually exclusive and jointly exhaustive. Parenthetical operators, as well as those tagged to grayed curly brackets outside the partition, highlight the location of various non-finest classes within the partition. Below, the twelve classes are defined via schemata, using only primitives, without redundancies.

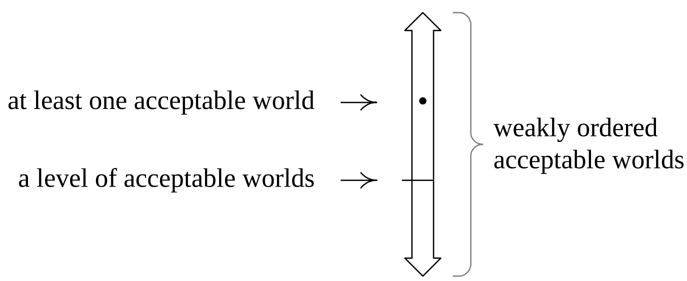
The Twelve Finest Classes Expressed via Schemata

OB	**MA**~ & ~**MI** & ~**MI**~
MA & **MI**~	~**MA** & ~**MA**~ & **MI**~
MA & **MI** & ~**OB**	~**MA** & ~**MA**~ & **MI**
MA & ~**MI** & ~**MI**~	~**MA** & ~**MA**~ & ~**MI** & ~**MI**~ & ~**IN**
MA~ & **MI**~ & ~**OB~**	**IN**
MA~ & **MI**	**OB~**

We turn now to one semantic framework for this logic.

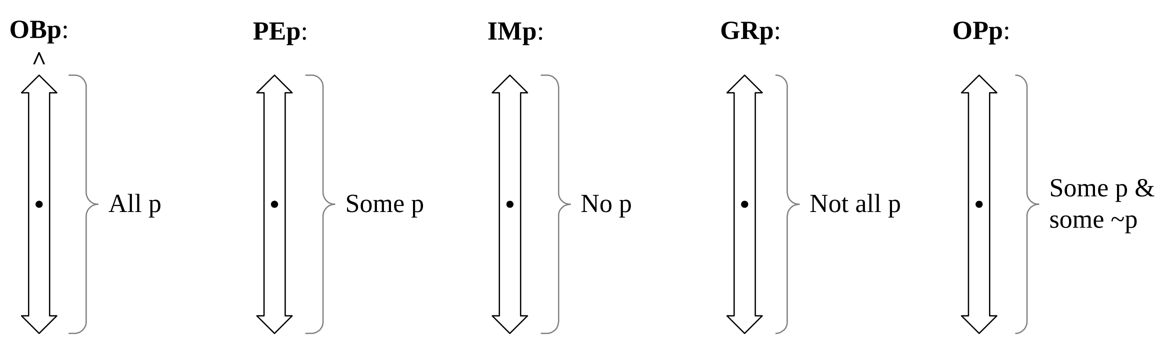
A3.2 DWE Semantics

To get the semantic structures we need, we simply combine and interpret two familiar ingredients in a convenient way: an accessibility relation and an ordering relation. We imagine that we have a set of worlds, and an accessibility relation--interpreted here as relating worlds to their morally *acceptable* alternatives. We assume that seriality holds: for each world, there is a morally acceptable alternative. Note that we do not think of these acceptable worlds as morally *ideal* or *optimal* alternatives. Rather, we assume that for any world i, there is a morally relevant i-relative weak ordering of the i-acceptable worlds (i.e. the i-relative ordering relation is reflexive, connected, and transitive with respect to the i-acceptable worlds). Thus, although all the acceptable alternatives to a given world are just that--acceptable, they needn't be on a par morally speaking. Some may be ranked higher than others, some may be ranked highest/lowest among the acceptable worlds, and there may be ties throughout (and thus there may be genuine *levels* of acceptable worlds). We can represent the i-acceptable worlds and their i-relative ordering as follows:

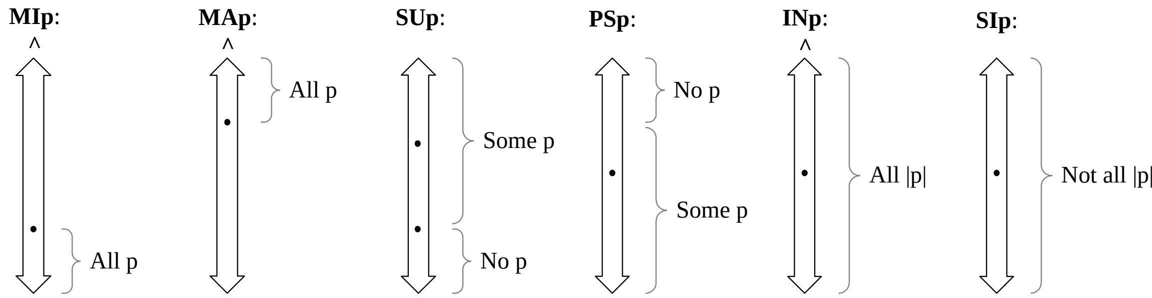


The vertical arrowed bar represents the weakly ordered i-acceptable worlds. The horizontal line through the bar is a reminder that there can be levels of i-acceptable worlds (each an equivalence class with respect to equi-rank), as in the fact that we choose vertical figures with width. The dot indicates there is always at least one i-acceptable world in these structures.

We can informally represent the truth-conditions (relative to a world i) for the traditional SDL operators as follows (where a "^" under an operator indicates that it is primitive in DWE).



For these operators, the interpretation does not depend on the ordering and matches that for SDL: p is *obligatory* (for agent S) iff p occurs in all of the i-acceptable alternatives; p is *permissible* iff it occurs in some, etc. However, the interpretation of the remaining operators depends crucially on the ordering of the i-acceptable worlds:



The *minimum* (for S) involves p iff p holds in all the lowest ranked acceptable alternatives; whereas *the maximum (optimum)* involves p iff it holds in all the highest ranked acceptable alternatives.[109] Thus as cast here, the minimum and the maximum are linked to the respective poles of the ranked acceptable alternatives and are mirror images of one another, which effects various symmetries in the logic (McNamara 1996a). p will be *supererogatory* if it holds in some acceptable alternative, but fails to hold in any of the lowest ranked acceptable alternatives. Similarly, p will be *permissibly suboptimal* if it holds in some acceptable alternative, but fails to hold in any of the highest ranked acceptable alternatives. Regarding moral indifference and moral significance, since we allow for ties, the ranked acceptable alternatives can be divided into "levels" (equivalence classes with respect to equal rank). An "all |p|" indicates that both p-worlds and ~p-worlds occur at each of the associated levels. p will then be a matter of *moral indifference* if at every such level its performance and its non-performance occurs somewhere therein. Conversely, p will be *morally significant* if there is some level of value that uniformly includes it or uniformly excludes it.

DWE Formal Semantics: The following formal semantics is generalized in Mares and McNamara 1997.

Frames are defined as follows:

F = < W, A, ≤□ > is a DWE-Frame:

(1) W is non-empty
(2) A is a subset of W^2 and A is serial: (Aij: j is an i-acceptable world)
(3) □≤ is a subset of W^3: (a) ($k \leq_i j$ or $j \leq_i k$) iff (Aij & Aik), for any i,j,k in W
(b) if $j \leq_i k$ and $k \leq_i l$ then $j \leq_i l$, for any i,j,k,l in W.

The notions of an assignment and a model are then easily defined:

P is an Assignment on *F*: *F* = <W, A, ≤□ > is a DWE-Frame and *P* is a function, P, from PV to Power(W), defined on PV (Propositional Variables).

[109] These are informally cast assuming lower and upper limit assumptions hold. The informal glosses can be easily adapted to discharge these assumptions, and the formal clauses below do not depend on any such boundedness.

$M = <F\ P>$ is a DWE-Model: $F = <$ W, A, $\leq\square$ $>$ is a DWE-frame and P is an assignment on F.

Truth at an Index in a Model: Let $M = <F, P>$ be a DWE-model, where $F = <$ W, A, $\leq$ $>$ and $j =_i k =_{df} j \leq_i k \;\&\; k \leq_i j$; then truth at a world in a model ($M \vDash_i$), truth in a model, and validity are easily defined:

Basic Truth-Conditions at a world, i, in a Model, M:

0) (Conditions for variables and truth functional connectives)
1) $M \vDash_i$ **OB**p: $\forall$j (if Aij then $M \vDash_j$ p).
2) $M \vDash_i$ **MA**p: $\exists$j(Aij & $\forall$k(if $j \leq\square_i k$ then $M \vDash_k$ p)).
3) $M \vDash_i$ **MI**p: $\exists$j(Aij & $\forall$k(if $k\ \square\leq_i j$ then $M \vDash_k$ p)).
4) $M \vDash_i$ **IN**p: $\forall$j[if Aij then $\exists$k($k =_i j$ & $M \vDash_k$ p) & $\exists$k($k =_i j$ & $M \vDash_k$ ~p)]

Derivative Truth Conditions:

5) $M\square \vDash_i$ **PE**p: $\exists$j(Aij & $M\square \vDash_j$ p).
6) $M\ \square\vDash_i$ **IM**p: $\forall$j(if Aij then $M \vDash_j$ ~p).
7) $M\square \vDash_i$ **GR**p: $\exists$j(Aij & $M\square \vDash_j$ ~p).
8) $M\square \vDash_i$ **OP**p: $\exists$j(Aij & $M\square \vDash_j$ p) and $\exists$j(Aij & $M\square \vDash_j$ ~p).
9) $M\square \vDash_i$ **SI**p: $\exists$j[Aij & either $\forall$k(if $k =_i j$ then $M\ \square\vDash_k$ p) or $\forall$k(if $k =_i j$ then $M\ \square\vDash_k$ ~p)].
10) $M\square \vDash_i$ **SU**p: $\exists$j(Aij & $M\square \vDash_j$ p) & $\exists$j[Aij & (k)(if $k \leq\square_i j$ then $M\square \vDash_k$ ~p)].
11) $M\square \vDash_i$ **PS**p: $\exists$j(Aij & $M\square \vDash_j$ p) & $\exists$j[Aij & (k)(if $j\ \square\leq_i k$ then $M\square \vDash_k$ ~p)].

Truth in a DWE-Model: $M \vDash$ p iff $M \vDash_i$ p, for every i in W of M.

Validity: $\square \vDash$ p iff $M\ \square\vDash$p, for all M.

In Mares and McNamara 1997, the metatheorem below is proven as a special case:

Metatheorem: The DWE-logic is determined by the class of DWE-models.

A4. A Glimpse at STIT Theory and Deontic Logic

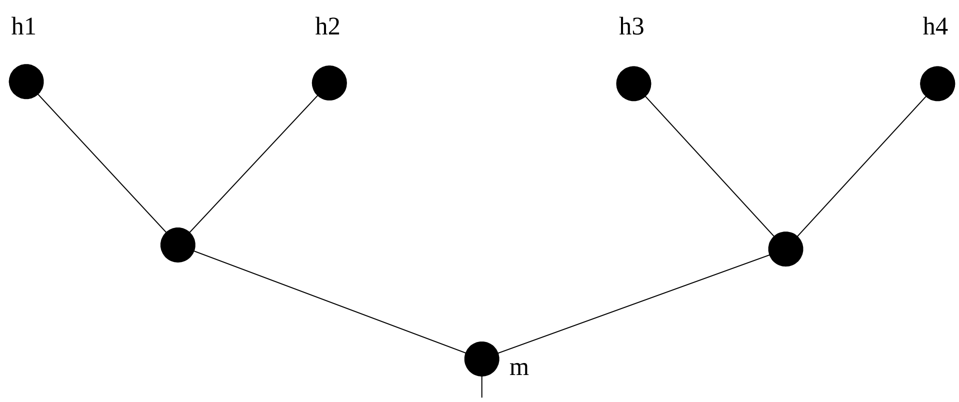
"STIT Theory" is so-called because it is a particular approach to constructions like "Jones sees to it that __". The following exposition draws from Horty 2001 and McNamara 2004b. STIT theory builds on a formal indeterministic "branching time" framework initiated by A. Prior, championed by R. Thomason, and now the basis of a robust research program anchored by N. Belnap, and summarized in Belnap 2001.[110] Here I concentrate on only a few elementary aspects of this sort of account of agency and provide just a glimpse of its employment in deontic logic. The reader is encouraged to consult the above two works, which are rich in details we can hardly touch on here. See also the works mentioned in the main essay under agency, especially

[110] This is the quintessential tome on STIT theory *per se*, and itself contains chapters on deontic logic in the context of STIT theory.

those by Hilpinen, for further critical exposition of this approach to agency.

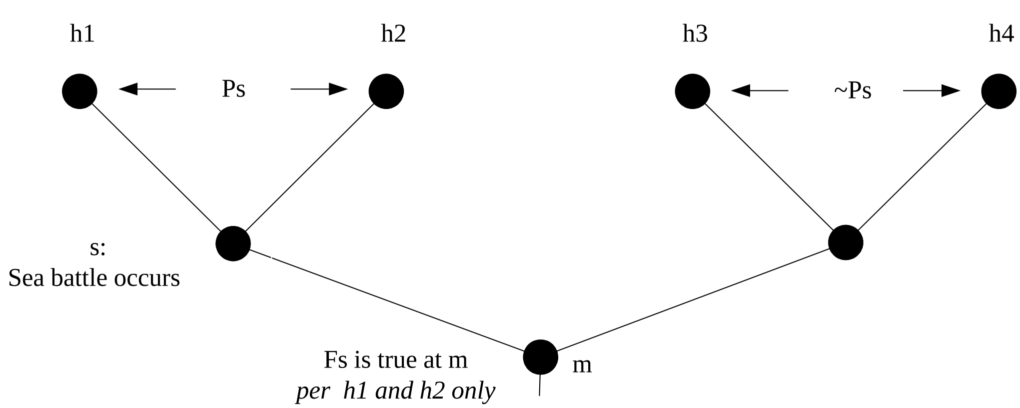
A4.1 The Indeterministic Framework

The basic primitives are a set of ‘moments’, *Tree*, and a two-place ordering relation, *after*, defined on *Tree*. A *moment* (represented as a node below) is thought of as momentary world state (cf. instantaneous possible world slice). Moments are not to be confused with seconds or instants. One moment is (possibly) *after* a second iff the first is some still possible future moment of the second. Moments can branch forward (upward in the diagrams), toward the future, but not backward. (There can be more than one possible future, but only one past, at a moment.)

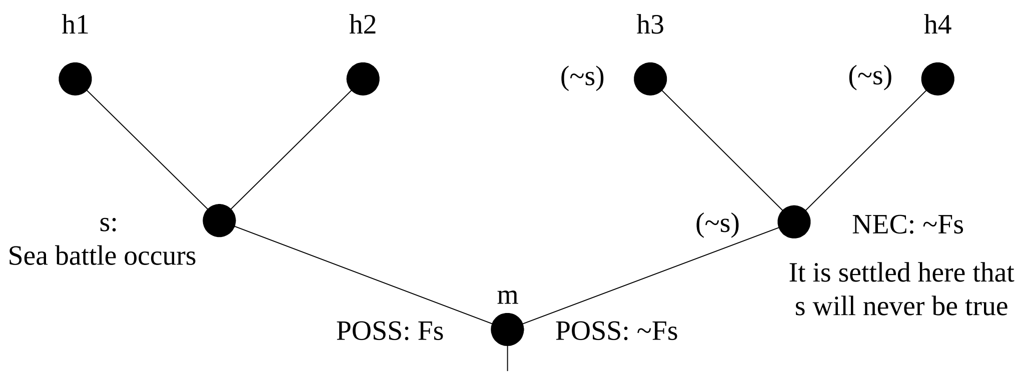


Upper moments are ones that can occur after line-connected lower moments. *A history (cf. possible world)* is construed as a maximal path or branch on a tree (e.g. each of the three-noded paths tracing from h1-h4 back to m above). In models with two or more histories, some moment (e.g. each top moment above) is not comparable to another (neither is a possible future moment of the other), and some moment is common to distinct histories (e.g. all but the top moments above). *A history passes through a moment* when that moment is part of that history. *The past at a moment* is the ordered sequence of moments before the moment in question. *The future at a moment in a history*: the ordered moments in the history after the moment in question. (There is no actual future at a moment per se, since there are many possible such futures, unless determinism is true.)

Since the future is open, contingent future tensed statements, and thus all statements for uniformity, are assigned truth values at a moment-history *pair, m-h.* (I will sometimes ignore histories in formulations where uniformity is the only reason to mention them.) Here is a simple illustration, where "P" is the past tense operator "it was the case that", and "F" is the future tense operator "it will be the case that":

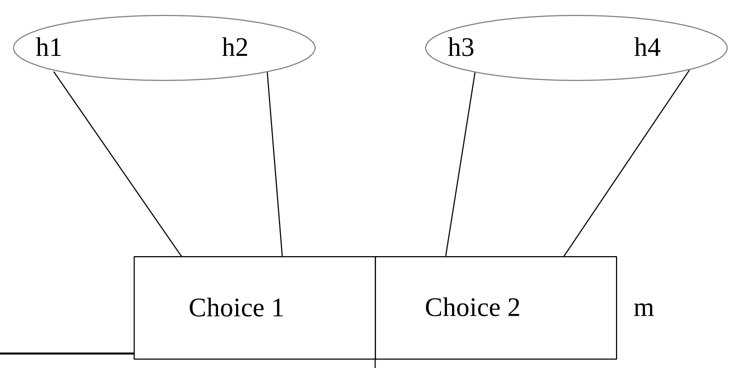


It will be the case that s at a moment *in a history* iff at some later moment *in that history*, s is true; and it was the case that s at a moment in a history iff that moment is after one where s is true in that history. Because the past is closed, simple past truths are true at a moment per se. The only case where we can say at a moment *simpliciter* that a statement will be true is where its future truth is historically necessary. More generally, possibility and necessity are handled as follows: it is *(still) possible (POSS) that* s is true at a moment iff there is a history passing through that moment where s is true. It is *(now) necessary/settled* (NEC) that s at a moment iff s is true at every history passing through that moment. We can illustrate via the sea battle again, which is completely open as of m, but settled false in h3 and h4 just after m.



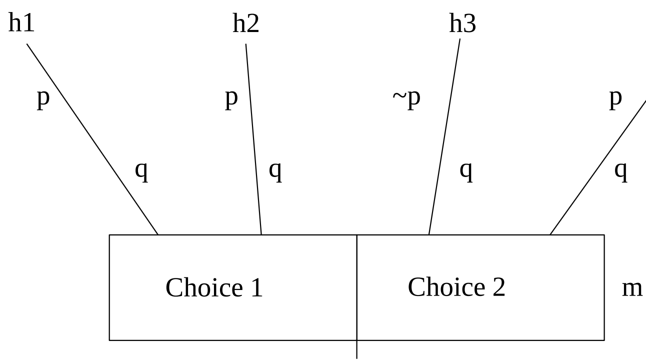
A4.2 Agency

Next, a set of *Agents*, and a *Choice* function are introduced. The *Choice* function partitions the histories passing through a moment relative to each agent. Thus the agent's possible choices or basic actions (the cells) constitute a mutually exclusive and exhaustive division of the histories passing through that moment. Choices at a moment place instantaneous constraints on the possible futures. Intentions are not represented.[111] Where there is more than one cell, no particular basic action is determined at that moment. If a history is part of a choice cell, then that is *the choice the agent makes at that moment in that history*. Below Choice 1 is the set containing just h1 and h2; it rules out all that depends on either h3 or h4 unfolding. Choice 1 is the basic action the agent takes in h1 and in h2 at m.



[111] Thus "basic action" is perhaps better than "choice".

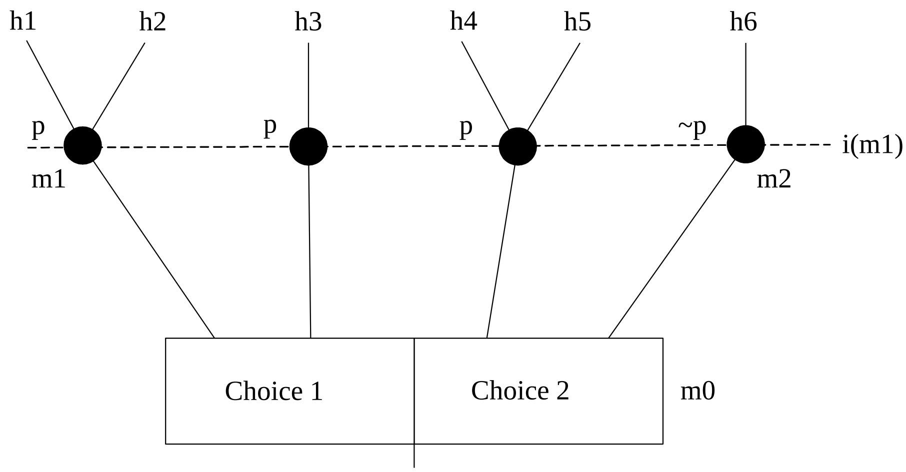
We can now easily distinguish two simple accounts of agency in terms of these basic actions or choices. The first is close to one Chellas gave in his seminal Chellas 1969. Jane Doe *c-sees to it that (c-stit) p* at a moment-history pair iff p is guaranteed by the choice Jane takes at that moment in that history (i.e. that choice cell contains only *p*-histories). Jane Doe is *able to c-see to it that p* at a moment-history pair iff it is possible that she c-stit *p* at that pair. In the illustration below, Jane c-stit p at m in h1 and h2. However, she does not c-stit p at m in h3 and h4 (since ~p's truth value varies independently of choice 2), nor is she able to c-stit ~p (since no history passing through m involves a choice at m that guarantees ~p).



An obvious rub with c-stit is illustrated above: Jane c-stit q at m in h1-h4. The upshot is that agents see to everything that is historically necessary (e.g. that the sun will rise and that 2 + 2 = 4). Enter: d-stit, which just adds the exclusion of necessary things for agency. Jane Doe *d-sees to it that (d-stit) p* at a moment-history pair iff she c-stit *p* at that pair and it is not necessary that p (i.e. ~p is consistent with some other choice open to her at m). This is called the 'deliberative stit' because the second condition, is meant to assure a real choice. Then, Jane Doe is *able to d-see to it that p* at a moment-history pair iff it is possible that she d-stit *p* at that pair. In the illustration above Jane d-sees to it that p at m only in h1-2, but, Jane does not d-see to it that q at m in any of h1-4. Jane is able to d-see to it that p at m (for at a history passing through m she does see to it that p), but Jane is not able to d-see to it that ~p or that q at m.

<u>Belnap's Achievement stit</u>:

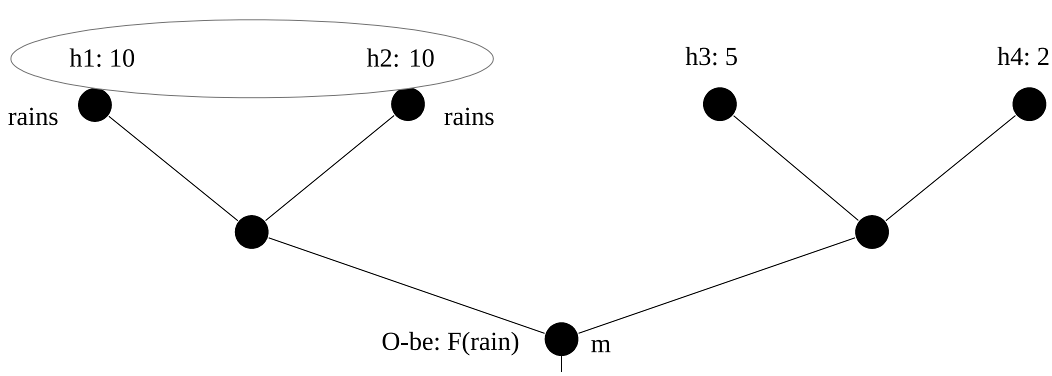
Belnap has a more complex alternative formal account of agency, which we can only briefly allude to here. The basic idea is that Jane sees to it that p now iff p now holds and was guaranteed by a *prior* choice of Jane's. Above, with c-stit and d-stit, one sees to it that something is the case at the moment of the choice or basic action; in Belnap's alternative the focus is on something's now holding as a result of a past action, so that the result and the initial instrumentality on the part of the agent that triggers the result are separated in time in this account of agency. Roughly, Belnap introduces the notion of instants as equivalence classes of contemporaneous moments. Intuitively, on a tree-display, moments on the same level are contemporaneous with one another, and the time or instant is taken to be the set of these moments. Then, *Jane Doe a-stit that p at moment m1 in a history, h* iff 1) there is a moment in h, m0, that is earlier than m1, and p holds at the instant of m1 in all histories consistent with the choice Jane makes at m0 in h, but 2) there is also a moment that is after m0 (and thus was still possible then), that is contemporaneous with m1, and at which p is false.



Above, in h1-3, Jane a-stit p at m1 (but not at m0 where the choice resulting in p is made), but not in h4-h6 at any moment in i(m1), since at m2, ~p holds.

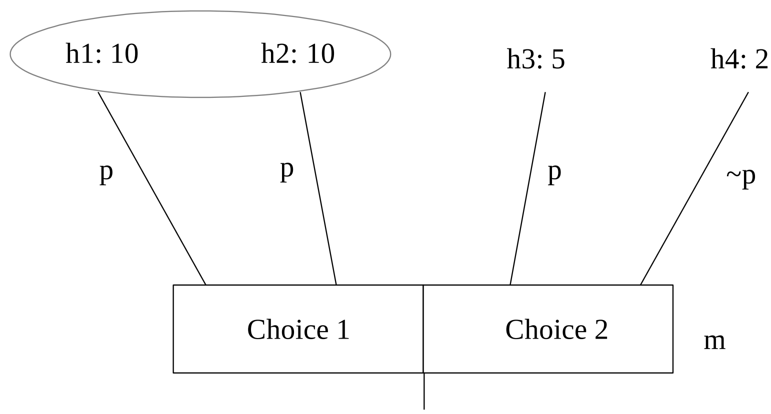
A4.3 Two Deontic Operators

Let's assume that histories have a rank-reflecting numerical value that does not vary from moment to moment (so histories are linearly ordered and thus mutually comparable). For simplicity, I will assume we always have best histories. Impersonal 'ought's may then be analyzed as follows: *it ought to be that p* holds at moment-history pair iff the best histories passing through m are histories where p holds. In the model below, since h1 and h2 are the highest ranked histories still possible as of m, and it rains at each of these at the last moments listed in those histories, it follows that at m, it ought to be the case that it will rain.

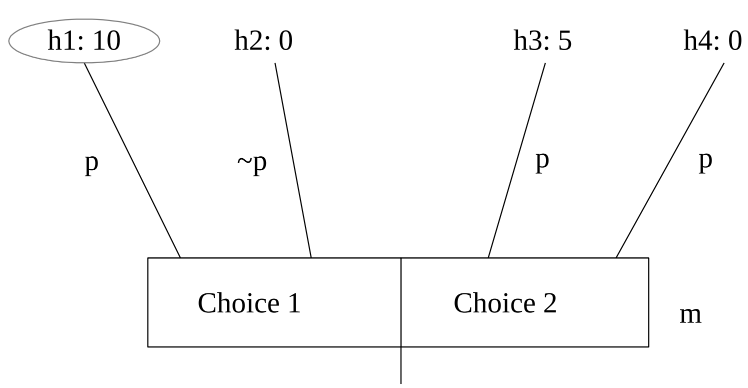


All the principles of Standard Deontic Logic (SDL), including no ought-conflicts, follow. A non-agential version of Kant's Law (it ought to be that *p* only if *p* is historically possible) also follows (Horty 2001).

If we endorse the ‘Meinong-Chisholm reduction’, then recast in the c-stit framework (for simplicity), this becomes *an agent ought to see to p* iff it ought to be the case that the agent c-sees to it that *p*. Similarly for recasting via the d-stit account. Given the previously proposed semantics, an agent ought to c-see it that *p* holds at a moment-history pair iff that agent chooses a p-guaranteeing action at the best histories passing through that moment. Similarly for recasting via the d-stit account. In the diagram below, Jane ought to see to it that p at m (in h1-h4) whether we recast the Meinong-Chisholm reduction via c-stit or d-stit, since Choice 1 guarantees p, and Choice 2 is consistent with ~p, and the best worlds are ones where Jane makes choice 1 and thus sees to it that p.



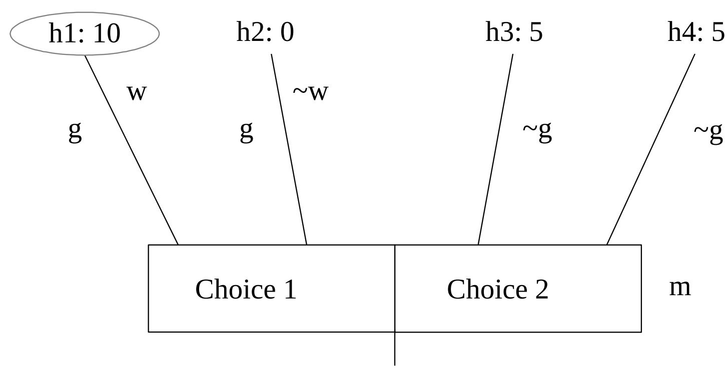
Relativized to c-stit (but not to d-stit) this analysis of agential 'ought's yields a normal modal operator satisfying the principles of SDL. An agential version of Kant’s Law follows: an agent ought to see to *p* only if she is able to. It also follows that what an agent ought to do, ought to be, but that the converse does not hold is illustrated by the next diagram. Here, although it ought to be that Jane makes Choice 1 and that p comes about, since both these things hold throughout the best histories (namely h1), it is not true that she ought to bring it about that p, since it is not true that it ought to be the case that she does. In the best world, h1, Jane does not make a choice that guarantees p's occurrence.



<u>Refraining Again</u>: Following a view championed by Belnap, if we analyze Jane's *refraining* from

seeing to it that p as her d-seeing to it that she does not d-see to it that p (Jane d-stit ~(Jane d-stit p)), and distinguish this from Jane's *omitting* p (~Jane d-stit p) then we can in fact say that it ought to be that Jane refrains from seeing to it that p above. She does d-see to it that p at m in h3 and h4, but these are suboptimal histories. So by making Choice 1 instead of Choice 2, she d-sees to it that she does not d-see to it that p, and this is what happens in the best history. Roughly, the best history is one where p occurs by luck or by some other agency than Jane's. Recall von Wright's alternative analysis of refraining: Jane refrains from seeing to it that p iff Jane is *able* to see to it that p, but she doesn't. Recast via d-stit this becomes Jane refrains from seeing to it that p iff ~Jane d-stit p and it is possible that Jane d-stit p. It can be shown about Belnap's and Von Wrights glosses, when recast via d-stit as indicated, Jane Belnap-refrains from p iff S von-Wright refrains from p, and that Jane refrains from p iff Jane refrains from refraining from p. For c-stit, omitting and refraining are indistinguishable. (See Horty 2001.)

Horty considers some previous objections to the Meinong-Chisholm reduction, and argues that from the standpoint of his framework, these arguments are unsound. He then introduces his own objection to the analysis, via the 'Gambling Problem'.[112] Suppose I have two options available to me, gamble $5 (g) or not. Now suppose that if I gamble and win, I get $10; and if I gamble and lose, I get $0. Suppose the only values at stake are the dollar values, and thus the value of not gambling is $5 saved. Ignore probabilities. To illustrate:



Since I cannot determine whether or not I win (this happens only in h1, which I can't guarantee), it is not true in fact that what I ought to do is gamble (or not gamble for that matter). But in the best histories (h1), I win, and my gambling is entailed by my winning, so it ought to be that I see to it that I gamble, and hence the Meinong-Chisholm reduction implies that I ought to gamble after all. Horty takes this to decisively defeat the M-C reduction, arguing that we need an independent analysis, one where we can rank actions, not just whole histories. Horty goes on to develop an alternative analysis of agential 'ought's, one in which, among other things, he uses the ranking of histories to generate a decision-theoretic dominance ordering account of agential 'ought's. Here we must pass over this fascinating work, and simple raise a few quick questions about a few elementary matters.

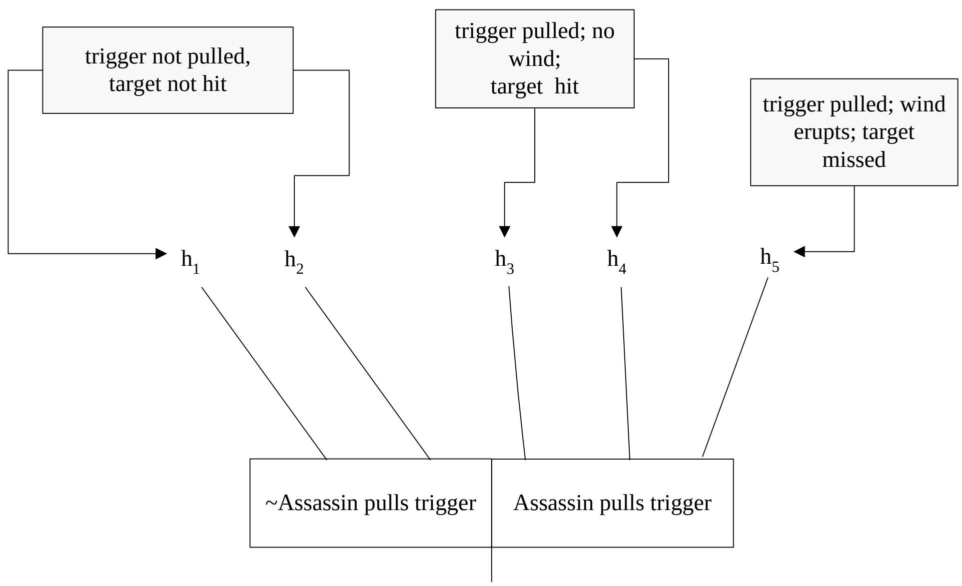
<u>A4.4 Some Challenges</u>

We saw that in the case of c-stit, an agent sees to all necessary truths. Few who work on agency

[112] Cf. Feldman 1986 pp.194-5.

accept this. D-stit is intended to get around this, but here seeing to it that p requires that it still might be that ~p, thus making agency depend ***logically*** on the falsity of compatibilism (Elgesem 1997). Thus nothing inevitable can be the result of my agency. But compatibilism is a widely endorsed live option in philosophy. It does not seem that a logic for agency ought to presuppose the falsity of this widely endorsed philosophical view.

Furthermore, the stit framework seems to make it too easy to undermine genuine agency by making the conditions for agent causation too strong. Consider the following "Windy Day Assassin" scenario:



Suppose I pull the trigger of a gun aiming at you intending to kill you, and you are hit by the bullet and die as a result of being hit by that bullet, just as planned. Now add that when I pulled the trigger a random gust of wind *could* have occurred and knocked the bullet off target, though it didn't occur. On the current analyses, it follows that I did not see to it that you were hit, because no choice I made *guaranteed* that you were hit. The mere fact that the wind *could have* interfered with the course of the bullet is enough to undermine the claim that I was the agent of your being hit. Even if I aim, pull the trigger, and the wind doesn't blow, as in h3 and h4, I still don't see to it that the target is hit on stit theory. This smacks of getting away with murder. (This problem applies to c-stit (the Chellas-inspired stit operator), d-stit (the Horty-Belnap deliberative stit operator), and a-stit (Belnap's achievement stit operator).

Finally, there is a general problem with this general sort of analysis of impersonal 'ought's in terms of what is still possible. It is now settled that some children will die of starvation tomorrow, but that ought to not now be the case. It seems false that everything that ought to be the case still could be the case. Talk of what would be ideal is not constrained by what is still possible, so why should talk of what ought to be the case, which involves after all, just the evaluation of states of affairs, not of agent's actions, be any different?

Bibliography

Alchourron, Carlos E. (1993). "Philosophical Foundations of Deontic Logic and the Logic of Defeasible Conditionals." *Deontic Logic in Computer Science*. Meyer, John-Jules Ch. and Roel J. Wieringa. Chichester, John Wiley and Sons**:** 43-84.

--- (1996). "Detachment and Defeasibility in Deontic Logic." *Studia Logica* **57**(1): 5-18.

Alchourron, Carlos E. and Eugenio Bulygin (1971). *Normative Systems*. Wien/New York, Springer-Verlag.

Alchourron, Carlos E. and Eugenio Bulygin (1981). "The Expressive Conception of Norms." *New Studies in Deontic Logic*. Hilpinen, Risto. Dordrecht, Reidel**:** 95-124.

Alchourron, Carlos E. and David Makinson (1981). "Hierarchies of Regulations and Their Logic." *New Studies in Deontic Logic*. Hilpinen, Risto. Dordrecht, Reidel**:** 125-148.

al-Hibri, Azizah (1978). *Deontic Logic: A Comprehensive Appraisal and a New Proposal*. Washington, DC, University Press of America.

Allen, Layman E. (1996). "From the Fundamental Legal Conceptions of Hohfeld to Legal Relations: Refining the Enrichment of Solely Deontic Legal Relations." *Deontic Logic, Agency and Normative Systems*. Brown, Mark A & Jose Carmo. New York, Springer Verlag**:** 1-26.

Anderson, Alan Ross (1956). "The Formal Analysis of Normative Systems." *The Logic of Decision and Action*. Rescher, Nicholas. Pittsburgh, University of Pittsburgh Press**:** 147-213.

--- (1958). "A Reduction of Deontic Logic to Alethic Modal Logic." *Mind* **67**: 100-103.

--- (1967 [1956]). "Some Nasty Problems in the Formal Logic of Ethics." *Nous* **1**: 345-360.

Åqvist, Lennart (1967). "Good Samaritans, Contrary-to-Duty Imperatives, and Epistemic Obligations." *Nous* **1**: 361-379.

--- (2002 [1984]). "Deontic Logic." *Handbook of Philosophical Logic, 2nd Edition*. Gabbay, Dov and F. Guenthner. Dordrecht, Kluwer Academic Publishers. **8:** 147-264.

Åqvist, Lennart and Jaap Hoepelman (1981). "Some Theorems About a 'Tree' System of Deontic Tense Logic." *New Studies in Deontic Logic*. Hilpinen, Risto. Dordrecht, Reidel**:** 187-221.

Asher, Nicholas and Daniel Bonevac (1996). "'Prima Facie' Obligation." *Studia Logica* **57**(1): 19-45.

Bailhache, Patrice (1998). "How to Mix Alethic, Deontic, Temporal, Individual Modalities." *Logica Trianguli* **2**: 3-16.

Bartha, Paul (1999). "Moral Preference, Contrary-to-Duty Obligation and Defeasible Oughts." *Norms, Logics and Information Systems*. McNamara, P. and H. Prakken. Amsterdam, IOS Press**:** 93-108.

Belnap, N, and M. Perloff and M. Xu (2001). *Facing the Future: Agents and Choices in Our Indeterminist World*. New York, Oxford University Press.

Bonevac, Daniel (1998). "Against Conditional Obligation." *Nous* **32**(1): 37-53.

Brewka, Gerhard (1989). "Non-Monotonic Logics - a Brief Overview." *AI Communications: The European Journal of Artificial Intelligence* **2**(2): 88-97.

Brown, Mark A. (1996a). "Doing as We Ought: Towards a Logic of Simply Dischargeable Obligations." *Deontic Logic, Agency and Normative Systems*. Brown, Mark A & Jose Carmo. New York, Springer Verlag**:** 47-65.

--- (1996b). "A Logic of Comparative Obligation." *Studia Logica* **57**(1): 117-137.

Carmo, Jose and Andrew Jones (2002). "Deontic Logic and Contrary-to-Duties." *Handbook of Philosophical Logic*. Geunther, D. Gabay and F. Dordrecht, Kluwer Academic Publishers. **8:** 265-343.

Castañeda, Hector Neri (1981). "The Paradoxes of Deontic Logic: The Simplest Solution to All of Them in One Fell Swoop." *New Studies in Deontic Logic*. Hilpinen, Risto. Dordrecht, Reidel**:** 37-85.

Chellas, Brian F. (1969). "The Logical Form of Imperatives." Philosophy Department. Stanford University. Stanford: Perry Lane Press.

--- (1974). "Conditional Obligation." *Logical Theory and Semantic Analysis*. Stenlund, Soeren. Dordrecht, D. Reidel Publishing Co.**:** 23-33.

--- (1980). *Modal Logic: An Introduction*. Cambridge, Cambridge University Press.

Chisholm, Roderick M. (1963a). "Contrary-to-Duty Imperatives and Deontic Logic." *Analysis* **24**: 33-36.

--- (1963b). "Supererogation and Offence: A Conceptual Scheme for Ethics." *Ratio* **5**: 1-14.

--- (1964). "The Ethics of Requirement." *American Philosophical Quarterly* **1**: 147-153.

Chisholm, Roderick M. and Ernest Sosa (1966). "Intrinsic Preferability and the Problem of Supererogation." *Synthese* **16**: 321-331.

Da Costa, Newton C. A. and Walter A. Carnielli (1986). "On Paraconsistent Deontic Logic." *Philosophia* **16**: 293-305.

Danielsson, Sven (1968). *Preference and Obligation: Studies in the Logic of Ethics*. Uppsala, Filosofiska föreningen.

DeCew, Judith Wagner (1981). "Conditional Obligation and Counterfactuals." *Journal of Philosophical Logic* **10**: 55-72.

Elgesem, Dag (1993). "Action Theory and Modal Logic." Department of Philosophy. University of Oslo. Oslo.

--- (1997). "The Modal Logic of Agency." *Nordic Journal of Philosophical Logic* **2**(2): 1-46.

Fehige, Christoph (1994). "The Limit Assumption." *Deontic (and Prohairetic) Logic in Analyomen 1, Meggle, Georg (Ed)*. Hawthorne, de Gruyter.

Feldman, Fred (1986). *Doing the Best We Can : An Essay in Informal Deontic Logic*. Dordrecht ; Boston, D. Reidel Publishing Company.

Føllesdal, Dafgin and Risto Hilpinen (1971). "Deontic Logic: An Introduction." *Deontic Logic: Introductory and Systematic Readings*. Hilpinen, R. Dordrecht, D. Reidel.

Forrester, James William (1984). "Gentle Murder, or the Adverbial Samaritan." *Journal of Philosophy* **81**: 193-196.

Forrester, Mary (1975). "Some Remarks on Obligation, Permission, and Supererogation." *Ethics* **85**: 219-226.

Goble, Lou (1990a). "A Logic of Good, Should, and Would: Part I." *Journal of Philosophical Logic* **19**(2): 169-199.

--- (1990b). "A Logic of Good, Should, and Would: Part Ii." *Journal of Philosophical Logic* **19**(3): 253-276.

--- (1991). "Murder Most Gentle: The Paradox Deepens." *Philosophical Studies* **64**(2): 217-227.

--- (Forthcoming-a). "A Logic for Deontic Dilemmas." *Journal of Applied Logic*.

--- (Forthcoming-b). "Preference Semantics for Deontic Logic Part I - Simple Models." *Logique et Analyse* **46**(183-184): 383-418.

Goldman, Holly S. (1976). "Dated Rigthness and Moral Perfection." *The Philosophical Review* **85**(4): 449-487.

Gowans, Christopher (1987). *Moral Dilemmas*. Oxford, Oxford University Press.
Greenspan, P. S. (1975). "Conditional Oughts and Hypothetical Imperatives." *Journal of Philosophy* **72**: 259-276.
Hansen, Jorg (2004). "Problems and Results for Logics About Imperatives." *Journal of Applied Logic* **2**(1): 39-61.
Hansson, Bengt (1969). "An Analysis of Some Deontic Logics." *Nous* **3**: 373-398.
Hansson, Sven Ove (1990). "Preference-Based Deontic Logic (Pdl)." *Journal of Philosophical Logic* **19**(1): 75-93.
--- (2001). *Structures of Values and Norms*. Cambridge, Cambridge University Press.
Hedenius, Ingemar (1963 [1941]). *Om Ratt Och Moral ('on Law and Morals')*. Stockholm.
Herrestad, Henning and Christen Krogh (1995). "Obligations Directed from Bearers to Counterparts." *Proceedings of the Fifth International Conference on Artificial Intelligence and Law (International Conference on Artificial Intelligence and Law)*. College Park, Maryland, USA, ACM Press**:** 210-218.
Hilpinen, Risto (1997a). "On Action and Agency." *Logic, Action and Cognition - Essays in Philosophical Logic*. Ejerhed, E. and S. Lindstrom. Dordrecht, Kluwer Academic Publishers**:** 3-27.
--- (1997b). "On States, Actions, Omissions and Norms." *Contemporary Action Theory (Individual Action)*. Holmstrom-Hintikka, Ghita and Raimo Tuomela. Dordrecht, Kluwer Academic Publishers. **1:** 83-107.
--- (2001a). "Deontic Logic." *The Blackwell Guide to Philosophical Logic*. Goble, Lou. Malden, Blackwell Publishers, LTD.**:** 159-182.
--- (2001b). "Stig Kanger on Deontic Logic." *Collected Papers of Stig Kanger with Essays on His Life and Work, Vol. Ii*. Holmstrom-Hintikka, G., S. Lindstrom and R. Sliwinski. Dordrecht, Kluwer Academic Publishers**:** 131-149.
Hintikka, Jaakko (1957). "Quantifiers in Deontic Logic." *Societas Scientiarum Fennica, Commentationes Humanarum Litterarum* **23**(4): 1-23.
--- (1971). "Some Main Problems of Deontic Logic." *Deontic Logic: Introductory and Systematic Readings*. Hilpinen, R. Dordrecht, D. Reidel Publishing company**:** 59-104.
Hohfeld, Wesley Newcomb (1919). *Fundamental Legal Conceptions as Applied in Judicial Reasoning (Edited by Walter Wheeler Cook)*. New Haven, Yale University Press.
Horty, John F. (1994). "Moral Dilemmas and Nonmonotonic Logic." *Journal of Philosophical Logic* **23**(1): 35-65.
--- (1996). "Agency and Obligation." *Synthese* **108**(2): 269-307.
--- (2001). *Agency and Deontic Logic*. Oxford England ; New York, Oxford University Press.
--- (2003). "Reasoning with Moral Conflicts." *Nous* **37**(4): 557-605.
Hrushka, Joachim and Jan C Joerden (1987). "Supererogation: Vom Deontologischen Sechseck Zum Deontologischen Zehneck." *Archiv fu Rechts und Sozialphilosophie* **73**: 93-123.
Humberstone, I. L. (1974). "Logic for Saints and Heroes." *Ratio* **16**: 103-114.
Jackson, Frank (1985). "On the Semantics and Logic of Obligation." *Mind* **94**: 177-195.
--- (1988). "Understanding the Logic of Obligation." *Aristotelian Society* **SUPP 62**: 255-270.
Jackson, Frank and Robert Pargetter (1986). "Oughts, Options, and Actualism." *Philosophical Review* **95**: 233-255.
Jones, Andrew (1990). "Deontic Logic and Legal Knowledge Representation." *Ratio Juris* **3**(2): 237-244.
Jones, Andrew and Ingmar Porn (1985). "Ideality, Sub-Ideality and Deontic Logic." *Synthese* **65**:

275-290.
--- (1986). "'Ought' and 'Must'." *Synthese* **66**: 89-93.
Jones, Andrew and Marek Sergot (1993). "On the Characterization of Law and Computer Systems." *Deontic Logic in Computer Science*. Meyer, John-Jules Ch. and Roel J. Wieringa. Chichester, John Wiley and Sons**:** 275-307.
--- (1996). "A Formal Characterization of Instituionalized Power." *IGPL* **4**(3): 429-445.
Jorgensen, Jorgen (1937). "Imperatives and Logic." *Erkenntnis* **7**: 288-296.
Kamp, Hans (1974). "Free Choice Permission." *Proceedings of the Aristotelian Society* **74**: 57-74.
--- (1979). "Semantics Versus Pragmatics." *Formal Semantics and Pragmatics for Natural Languages*. Guenthner, F. and S. J. Schmidt. Dordrecht, D. Reidel**:** 255-287.
Kanger, Stig (1971 [1957]). "New Foundations for Ethical Theory." *Deontic Logic: Introductory and Systematic Readings*. Hilpinen, R. Dordrecht, D. Reidel Publishing Co.**:** 36-58.
--- (1972). "Law and Logic." *Theoria* **38**: 105-132.
Kempson, Ruth M. (1977). *Semantic Theory*. Cambridge, Cambridge University Press.
Knuuttila, Simo (1981). "The Emergence of Deontic Logic in the Fourteenth Century." *New Studies in Deontic Logic*. Hilpinen, Risto. Dordrecht, Reidel**:** 225-248.
Krogh, Christen and Henning Herrestad (1996). "Getting Personal: Some Notes on the Relationship between Personal and Impersonal Obligation." *Deontic Logic, Agency and Normative Systems*. Brown, Mark A & Jose Carmo. New York, Springer Verlag**:** 134-153.
Lemmon, E. J. (1957). "New Foundations for Lewis' Modal Systems." *The Journal of Symbolic Logic* **22**: 176-186.
--- (1962a). "On Sentences Verifiable by Their Use." *Analysis* **22**: 86-89.
--- (1962b). "Moral Dilemmas." *Philosophical Review* **71**: 139-158.
Lemmon, E. J. and Dana Scott (1977). *The 'Lemmon Notes': An Introduction to Modal Logic*. Oxford, Basil Blackwell.
Lewis, David (1978). "Reply to Mcmichael's "Too Much of a Good Thing: A Problem in Deontic Logic"." *Analysis* **38**: 85-86.
Lewis, David K. (1973). *Counterfactuals*. Oxford, Blackwell.
--- (1974). "Semantic Analyses for Dyadic Deontic Logic." *Logical Theory and Semantic Analysis: Essays Dedicated to Stig Kanger on His Fiftieth Birthday*. Stenlund, Soeren. Dordrecht, D. Reidel Publishing Co.**:** 1-14.
Lindahl, Lars (1977). *Position and Change*. Dordrecht, D. Reidel Publishing co.
--- (2001). "Stig Kanger's Theory of Rights." *Collected Papers of Stig Kanger with Essays on His Life and Work, Volume Ii*. Holmstrom-Hintikka, G., S. Lindstrom and R. Sliwinski. Dordrecht, Kluwer Academic Publishers**:** 151-171.
Loewer, Barry and Marvin Belzer (1983). "Dyadic Deontic Detachment." *Synthese* **54**: 295-318.
Lokhorst, Gert-Jan (2004). "Mally's Deontic Logic." *The Stanford Encyclopedia of Philosophy (Winter 2004 Edition)*. Zalta, Edward N. Stanford, http://plato.stanford.edu/archives/win2004/entries/mally-deontic/.
Makinson, David (1993). "Five Faces of Minimality." *Studia Logica* **52**(3): 339-379.
--- (1999). "On the Fundamental Problem of Deontic Logic." *Norms, Logic and Information Systems*. McNamara, P. and H. Prakken. Amsterdam, IOS Press**:** 29-53.
Makinson, David and Leendert W. N. van der Torre (2003). "What Is Input/Output Logic?" *Foundations of the Formal Sciences Ii: Applications of Mathematical Logic in*

Philosophy and Linguistics. Löwe, Benedikt , Wolfgang Malzkorn and Thoralf Räsch. Berlin, Springer-Verlag. **17:** 163-174.
Mally, Ernst (1926). *Gurndgesetz Des Sollens. Elemente Der Logik Des Willens*. Graz, Leuschner & Leubensky.
Mares, Edwin D. (1992). "Andersonian Deontic Logic." *Theoria* **58**(1): 3-20.
Mares, Edwin D. and Paul McNamara (1997). "Supererogation in Deontic Logic: Metatheory for Dwe and Some Close Neighbours." *Studia Logica* **59**(3): 397-415.
McMichael, Alan (1978). "Too Much of a Good Thing: A Problem in Deontic Logic." *Analysis* **38**: 83-84.
McNamara, Paul (1990). "The Deontic Quaddecagon." Philosophy Department. University of Massachusetts. Amherst.
--- (1995). "The Confinement Problem: How to Terminate Your Mom with Her Trust." *Analysis* **55**(4): 310-313.
--- (1996a). "Doing Well Enough: Toward a Logic for Commonsense Morality." *Studia Logica* **57**(1): 167-192.
--- (1996b). "Making Room for Going Beyond the Call." *Mind* **105**(419): 415-450.
--- (1996c). "Must I Do What I Ought? (or Will the Least I Can Do Do?)." *Deontic Logic, Agency and Normative Systems*. Brown, Mark A. & Jose Carmo. New York, Springer Verlag**:** 154-173.
--- (1999). "Doing Well Enough in an Andersonian-Kangerian Framework." *Norms, Logics and Information Systems: New Studies in Deontic Logic and Computer Science*. McNamara, Paul and Henry Prakken. Washington, DC, IOS Press**:** 181-198.
--- (2004a). "Agential Obligation as Non-Agential Personal Obligation Plus Agency." *Journal of Applied Logic* **2**(1): 117-152.
--- (2004b). "[Review of] Agency and Deontic Logic." *Mind* **113**(447): 179-185.
--- (2005 (Fall)). "Deontic Logic." *The Stanford Encyclopedia of Philosophy (Fall 2005 Edition)*. Zalta, Edward N. Stanford, http://plato.stanford.edu/archives/fall2005/entries/logic-deontic/.
Meinong, Alexius (1972 [1917]). *On Emotional Presentation*. Evanston, Northwestern University Press.
Meyer, J. J. C. (1988). "A Different Approach to Deontic Logic: Deontic Logic Viewed as a Variant of Dynamic Logic." *Notre Dame Journal of Formal Logic* **29**: 109-136.
Morreau, Michael (1996). ""Prima Facie" and Seeming Duties." *Studia Logica* **57**(1): 47-71.
Mott, Peter L. (1973). "On Chisholm's Paradox." *Journal of Philosophical Logic* **2**: 197-210.
Nowell Smith, P. H. and E. J. Lemmon (1960). "Escapism: The Logical Basis of Ethics." *Mind* **69**: 289-300.
Nute, Donald, Ed. (1997). *Defeasible Deontic Logic*. Synthese Library ; V. 263. Dordrecht ; Boston, Kluwer Academic.
Porn, Ingmar (1970). *The Logic of Power*. Oxford, Blackwell.
--- (1977). "Action Theory and Social Science." *Synthese Library* **120**.
--- (1989). "On the Nature of a Social Order." *Logic, Methodology and Philosophy of Science*. Fenstad, J.E., T. Frolov and R. Hilpinen. New York, Elsevier Science. **VIII:** 553-67.
Prakken, Henry (1996). "Two Approaches to the Formalisation of Defeasible Deontic Reasoning." *Studia Logica* **57**(1): 73-90.
Prakken, Henry and Marek Sergot (1996). "Contrary-to-Duty Obligations." *Studia Logica* **57**(1): 91-115.

--- (1997). "Dyadic Deontic Logic and Contrary-to-Duty Obligations. In Nute."
Prior, A. N. (1954). "The Paradoxes of Derived Obligation." *Mind* **63**: 64-65.
--- (1958). "Escapism: The Logical Basis of Ethics." *Essays in Moral Philosophy*. Melden, A. I. Seattle, University of Washington Press**:** 135-146.
--- (1962 [1955]). *Formal Logic*. Oxford, Oxford University Press.
Quine, Willard (1976 [1962]). "The Ways of Paradox." *The Ways of Paradox and Other Essays*. Cambridge, Harvard University Press**:** 1-18.
Ross, Alf (1941). "Imperatives and Logic." *Theoria* **7**: 53-71.
Ross, W. D. (1939). *Foundations of Ethics*. Oxford, Oxford University Press.
Sajama, Seppo (1985). "Supererogation and High Values." *Theoria* **51**: 77-88.
Santos, Filipe and Jose Carmo (1996). "Indirect Action, Influence and Responsibility." *Deontic Logic, Agency and Normative Systems*. Brown, Mark A. & Jose Carmos. New York, Springer Verlag**:** 194-215.
Schotch, Peter K. and Raymond E. Jennings (1981). "Non-Kripkean Deontic Logic." *New Studies in Deontic Logic*. Hilpinen, Risto. Dordrecht, Reidel**:** 149-162.
Schurz, Gerhard (1997). *The Is-Ought Problem: An Investigation in Philosophical Logic*. Dordrecht ; Boston, Kluwer Academic.
Segerberg, Krister (1982). "A Deontic Logic of Action." *Studia Logica* **41**: 269-282.
--- (1997). "Delta Logic and Brown's Logic of Ability." *Logic, Action and Cognition - Essays in Philosophical Logic*. Ejerhed, E. and S. Lindstrom. Dordrecht, Kluwer Academic Publishers**:** 29-45.
Sergot, Marek (1999). "Normative Positions." *Norms, Logic and Information Systems*. McNamara, P. and H. Prakken. Amsterdam, IOS Press**:** 289-308.
Sinnott-Armstrong, Walter (1985). "A Solution to Forrester's Paradox of Gentle Murder." *Journal of Philosophy* **82**: 162-168.
Smiley, Timothy (1963). "The Logical Basis of Ethics." *Acta Philosophica Fennica* **16**: 237-246.
Smith, Tina (1993). "Violation of Norms." *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*. Amsterdam, The Netherlands**:** 60 - 65.
--- (1994). "Legal Expert Systems: Discussion of Theoretical Assumptions." Faculty of Law. University of Utrecht. Utrecht.
Thomason, Richmond H. (1981a). "Deontic Logic and the Role of Freedom in Moral Deliberation." *New Studies in Deontic Logic*. Hilpinen, Risto. Dordrecht, Reidel**:** 177-186.
--- (1981b). "Deontic Logic as Founded on Tense Logic." *New Studies in Deontic Logic*. Hilpinen, Risto. Dordrecht, Reidel**:** 165-176.
Tomberlin, James E. (1981). "Contrary-to-Duty Imperatives and Deontic Logic." *Nous* **15**: 357-375.
Urmson, J.O. (1958). "Saints and Heroes." *Essays in Moral Philosophy*. Melden, A. I. Seattle, University of Washington Press**:** 198-216.
van der Torre, Leendert (1997). "Reasoning About Obligations: Defeasibility in Preference-Based Deontic Logic." Tinbergen Institute. Erasmus University Rotterdam. Amsterdam.
van der Torre, Leendert and Yao Hua Tan (1997). "The Many Faces of Defeasibility in Defeasible Deontic Logic. In Nute 1997."
--- (2000). "Two-Phase Deontic Logic." *Logique et Analyse* **171-172**: 411-456.
van Eck, J. (1982). "A System of Temporally Relative Modal and Deontic Predicate Logic and Its Philosophical Applications (2)." *Logique et Analyse* **25**: 339-381.

van Fraassen, Bas C. (1972). "The Logic of Conditional Obligation." *Journal of Philosophical Logic* **1**: 417-438.
--- (1973). "Values and the Heart's Command." *Journal of Philosophy* **70**: 5-19.
von Wright, G. H. (1951). "Deontic Logic." *Mind* **60**: 1-15.
--- (1953). *An Essay in Modal Logic*. New York, Humanities Press.
--- (1956). "A Note on Deontic Logic and Derived Obligation." *Mind* **65**: 507-509.
--- (1963). *Norm and Action: A Logical Enquiry*. New York, Humanities Press.
--- (1968). *An Essay in Deontic Logic and the General Theory of Action*. Amsterdam, North Holland Publishing Co.
--- (1971). "A New System of Deontic Logic." *Danish Yearbook of Philosophy* **1**: 173-182.

Paul McNamara
paulm@cisunix.unh.edu