1

Deontic Logic: A Historical Survey and Introduction

Risto Hilpinen and Paul McNamara

ABSTRACT. This chapter provides both a historical overview and an introduction to core developments in deontic logic up to the end of the 20th century. The presentation becomes more systematic for the last half century covered, but continues to convey historical developments. In particular, we present some key developments from the Middle Ages through the 19th century, then turn to Meinong and Mally's contributions near the early part of the last century, followed by the full emergence of modern deontic logic with von Wright's work in the early 1950's. We next cover the emergence of the so-called "standard" systems of deontic logic in the 1960s, including the emergence of formal semantics for those systems. Then we cover a wide array of objections to, and limitations of, the standard systems, while also often indicating various lines of response to these challenges. Finally, we turn to the issue of the representation of action and agency in deontic contexts. Supplements to some sections or sub-sections provide the reader with the option of more details on a given topic.

Handbook of Deontic Logic and Normative Systems, 3–136.

The following handbook entry is primarily historical, aimed at providing orientation to the newcomer to deontic logic by conveying a sense of the sweep of themes associated with core material in deontic logic, while at the same time occasionally offering something of interest to those who have already entered the field. We aim to cover the history of deontic logic through roughly the end of the last millennium.[1] Once we turn to the emergence of deontic logic as a full-fledged active area within symbolic logic in the second half of the last century (in Sections 6-9) with the emergence of the so-called "standard" systems of deontic logic, the approach shifts to a more systematic orientation, but with historical information included in the process. We also then must become more selective, and that leads to a focus on core areas and ideas, familiarity with which we think is often presupposed by those actively working in this area. Although there is a narrative structure, especially in the earlier half of this chapter, most sections of the chapter (even the earlier ones) are relatively self-contained despite occasional references backward. Section 8 in particular, which catalogs an array of puzzles and challenges that the standard systems faced, puzzles that often served (some still do) as catalysts for new work, is designed so that the reader can dip into even one of its sub-sections and read about one puzzle more or less independently of the rest. Overall, we think the chapter allows for the acquisition of a narrative sense of the historical core of deontic logic, but without paying the price "all or nothing".

The authors are quite aware that we have had to make choices, and the result falls short of all we had hoped to include, especially as we approach the end of the last millennium; still, it is our hope that such shortcomings

[1]With the exception of reference to some work by Islamic philosophers, the main focus is on the Western tradition. We are not aware of work of relevance in other traditions (although we did make some preliminary inquiries of experts, e.g. about South Asian literature), and we would not be competent to discuss such material unless it was already covered in secondary sources, which we did not succeed in finding.

will be partially overcome by the fine chapters that follow ours on particular areas and issues in deontic logic that our colleagues have written.[2]

Introduction: What is deontic logic?

Deontic logic is an area of logic which investigates normative concepts (deontic concepts), along with closely associated evaluative concepts, norms and norm systems, and normative reasoning. The word 'deontic' is derived from the Greek expression 'déon' (δέον), which means 'what is binding' or 'proper'. Jeremy Bentham used the word 'deontology' for "the science of morality" [Bentham, 1983], and the Austrian philosopher [Mally, 1971], who developed in the 1920's a system of the "fundamental principles the logic of ought", called his theory "Deontik". Normative concepts include the concepts of obligation (duty, requirement), the concept of ought, permission (permissibility, 'may'), prohibition ('may not', 'forbidden'), and related notions, for example, those expressed by the words 'right', 'optional' (normatively contingent), 'claim', 'power', 'immunity', and 'supererogatory'. Deontic logic is also concerned with the relations among normative concepts, axiological concepts (value concepts, e.g., 'good' and 'better than'), and agent-evaluative concepts (e.g., 'blameworthy' and 'praiseworthy'). Thus the formal languages of deontic logic contain, in addition to propositional connectives and quantifiers, logical constants for normative concepts, and in some cases operators representing axiological concepts, praxeological concepts (for agency and action), prohairetic concepts (for preference and interest), aretaic concepts (for agent evaluation) and perhaps other modalities.. The concepts of agency, action, and preference connect deontic logic to the logic of practical reasoning.

1 Early developments. From medieval deontic logic to the 19th century

In his manuscript *Elementa iuris naturalis* [1930] Gottfried Wilhelm Leibniz called the deontic categories of the obligatory (*debitum*), the permitted (*licitum*) and the prohibited (*illicitum*) "modalities of law" (*iuris modalia*), and observed that important basic principles of alethic modal logic hold for these legal modalities. Much of the development of deontic logic during the past half century, and especially in the founding decade of the 1950s, has been based on just such modal analogies, and thus deontic logic has often been studied as a branch of modal logic.[3] In other words, the con-

[2]The authors of this chapter in particular would welcome any corrections or suggestions for improvement.

[3]Deontic necessity was taken to represent what was morally obligatory *all things considered*, and not merely *prima facia* obligatory, a distinction stressed explicitly in W.

cept of obligation has been studied as normative (deontic) necessity, and the concept of permission or permissibility has been construed as normative (deontic) possibility. Moreover, Leibniz suggested that legal (or deontic) modalities can be defined in terms of (or "reduced to") the alethic modalities of necessity and possibility, and one evaluative notion, in a way that is reminiscent of recent approaches in virtue ethics to defining deontic concepts.[4] According to Leibniz, the permitted is "what is possible for a good person to do", and the obligatory is "what is necessary for a good person to do". (Cf. [Hruschka, 1986, p. 35-6]) As it turns out, one strain of the early developments in deontic logic emerging in the mid-Twentieth Century also concurred with the sort of "reductive" approach to the deontic operators endorsed here by Leibniz. Leibniz might thus be said to be prescient regarding some of the first main lines of approach to deontic logic as an area of symbolic logic emerging solidly in the 1950s. It is interesting to note here that five hundred years before Leibniz, Peter Abelard (1079-1144) and other early medieval philosophers often endorsed an inverted form of Leibniz' reduction by defining alethic modal concepts by means of normative or quasi-normative concepts. According to this characterization, necessity is taken to be what nature demands, possibility is identified with what nature allows, and impossibility with what nature forbids. [Knuuttila, 1993, p. 182] Thus analogical links between deontic logic and alethic modal logic have a long and rich history before their widespread reemergence a half century ago in symbolic deontic logic.

In particular, formal analogies between deontic notions and "pure" (alethic) modalities (necessary, possible and impossible) were studied by many 14th century philosophers who regarded deontic logic as a branch of modal logic. They presupposed and used the following equivalences in their discussions on normative concepts: (**O** stands here for the concept of obligation (*obligatum*), **P** for permission, and **F** for prohibition.)

(1.1) (i) $\mathbf{P}p \leftrightarrow \neg\mathbf{O}\neg p$,
(ii) $\mathbf{O}p \leftrightarrow \neg\mathbf{P}\neg p$,
(iii) $\mathbf{O}p \leftrightarrow \mathbf{F}\neg p$, and
(iv) $\mathbf{F}p \leftrightarrow \mathbf{O}\neg p$.

The interest in deontic modalities in late medieval philosophy, especially in the 14th century, was related to the attempts to systematize modal theory and overcome the observed inadequacies in Aristotle's account of modal syl-

D. Ross' seminal [Ross, 1939].

[4][Zagzebski, 1996; Slote, 1997; Hansson, 1999] are representative of this recent approach to analyzing deontic concepts via virtue theoretic concepts in ethical theory.

logisms. Many logicians thought that the logic of alethic modalities (modalities of truth or being), such as those of necessity and possibility, could be used as a formal model for other concepts which show apparent similarities to modal concepts, such as knowing, believing, having an opinion, doubting, appearing, and being obligatory, permitted, or forbidden. These interpretations of modal logic led to the development of the elements of epistemic and deontic logic in the fourteenth century [Boh, 1985; Boh, 1993; Knapp, 1981], and to critical discussions of the applicability of the basic principles of modal logic to epistemic and normative concepts. The principles investigated include the following inference patterns

$$\frac{\mathbf{N}(p \to q)}{\mathbf{N}p \to \mathbf{N}q} \tag{1.2}$$

and

$$\frac{\mathbf{N}(p \to q)}{\mathbf{M}p \to \mathbf{M}q} \tag{1.3}$$

where **N** and **M** represent the concepts of alethic necessity and possibility. (1.2) is equivalent to the principle K of contemporary modal logic, and is a fundamental principle for normal modal logics. (See [Chellas, 1980, p. 114]) A number of medieval philosophers discussed the epistemic and deontic variants of these principles, and concluded that the following rules do not hold for deontic (nor epistemic concepts) without restrictions:

$$\frac{\mathbf{N}(p \to q)}{\mathbf{O}p \to \mathbf{O}q} \tag{1.4}$$

and

$$\frac{\mathbf{N}(p \to q)}{\mathbf{P}p \to \mathbf{P}q} \tag{1.5}$$

Principles (1.4) and (1.5) were discussed already in the 12th century as principles concerning the logic of will, and the counterexamples to them were formulated in terms of the concept of willing. Peter of Poitiers (1130-1205) gave the following example: If a sinner repents of a sin, he is guilty of sin, but if a sinner wills to repent of a sin, it does not follow that he wills to be guilty of sin. The example can be expressed in normative terms as follows: Necessarily, if a person R repents of a sin, R is guilty of sin, but it does not follow that R ought to be guilty of sin if he ought to repent of a sin. Stephen Langton's (1150-1228) counter-example was similar: Necessarily, if a man visits his sick father, the father is sick. But it does not follow that if this man wills to visit his sick father, then he wills the father to be sick. If the concept of willing is replaced by the concept of *ought*, we get the following counterexample to (1.4): Necessarily, if this man visits his sick father, the father is sick. But it does not follow that if this man ought to visit his sick father, then his father ought to be sick. For surely it ought not to be the case that this man's father is sick. (See [Knuuttila and Hallamaa, 1995, p. 77].) The 14the century philosopher Roger Roseth reformulated Peter of Poitiers's example as follows:

> "There are consequences which are good and known to be good the antecedent of which I am permitted to will, without being permitted to will be consequence, For example, this consequence is good and known to be good: I repent of my sin, therefore I am in sin. I am permitted to will the antecedent but I am not permitted to will the consequent, because I am permitted to repent of my sins, but I am not permitted to will to be in sin." [Knuuttila and Hallamaa, 1995, p. 77]

This example serves as a counterexample to principle (1.5); a permission to repent of one's sins does not entail a permission to sin. These medieval authors in effect argued that deontic logic is not a *normal* modal logic, that is, a logic satisfying, among other conditions, the rules

(1.6)
$$\frac{p \models q}{\mathbf{O}p \models \mathbf{O}q}$$

and

(1.7)
$$\frac{p \models q}{\mathbf{P}p \models \mathbf{P}q}$$

where $\models$ represents the concept of logical consequence. (1.6) is a deontic version of the modal rule usually called RM (see [Chellas, 1980, p. 114]); (1.6) and (1.7) may be called the consequence rules for **O** and **P**. In the 20th century deontic logic counterexamples to the consequence principle (1.6) reappeared in various forms, as Ross's paradox (the paradox of disjunctive obligation), the paradox of disjunctive permission ("free choice permission"), the paradox of the Good Samaritan, and the paradox of Epistemic Obligation. These paradoxes will be discussed below in Section 8.

In the 17th and 18th century literature on normative discourse and the logic of norms, some authors regarded normative concepts as analogous to modal concepts, like in the medieval literature, as was observed above in the particular case of Leibniz. Thus it was assumed that the concepts of obligation, permission and prohibition were related to each other in the same way as the modal concepts of necessity, possibility and impossibility ([Hruschka, 1986, pp. 39-43] and [Knebel, 1991]). Moreover, deontic concepts were usually applied to actions, thus deontic modalities were regarded as *action modalities*. Like the English word 'action', the expression 'actio' used in the 17th and 18th century literature did not refer only to human actions, but also to events which take place as a result of natural necessity; human actions (or actions in the narrow sense) were distinguished from other actions and events by the attribute 'liber'; thus the philosophers of this period made a distinction between *actio libera* and *actio physice necessaria* [Hruschka, 1986, p. 10]. Normative concepts were regarded as being properly applicable only to the former ('free') actions. The concept of obligatory action (*actio obligatoria*) was usually understood as *legally determined*action; an obligatory action in this sense may be an action which must be performed (*actio praecepta*, a morally necessary action) or an action which must be omitted (*actio prohibita*) [Hruschka, 1986, pp. 17-22]. In this context 'an action' meant an act-type or a kind of action rather than an individual action. These features are present in Leibniz's deontic logic. Leibniz defined the permitted (*licitum*) as "that which is possible for a good man [person] to do", and the obligatory (*debitum*, duty) as "that which is necessary for a good person to do". He called deontic modalities "iuris modalia", "modalities of law", and observed that the basic principles of the Aristotelian modal logic hold for the "iuris modalia" (modalities of law) as well as for the other modalities [Leibniz, 1930, p. 466].

2 Alexius Meinong on normative concepts and value concepts

In his work *Psychologisch-ethische Untersuchungen zur Werth-Theorie* [1968b] the Austrian philosopher Alexius Meinong (1853-1920) divided acts from the normative and evaluative point of view into four mutually exclusive "value-classes" (88-93):

(2.1) (a) Meritorious (*verdienstlich*),
(b) Correct (*correct*),
(c) (Merely) Excusable (*zulässig*), and

(d) Reprehensible, inexcusable (*verwerflich*).[5]

(See [Chisholm, 1982, pp. 104-5] and [Sajama, 1988, pp. 71-2].)The actions in classes (a) and (b) have a positive value, whereas those in (c) and (d) lie below the zero point of the "value line" (1894/1968, 90) Using what Meinong called the "vulgar expressions" 'good' and 'bad', the actions in categories (a) and (b) might be characterized as good, and those in (c) and (d) as bad [Meinong, 1968b, p. 92], even though "zulässige" acts can be said to be "bad" only in a rather weak sense (cf. [Sajama, 1988, p. 71]). Meinong's value classes can be correlated with a comparative concept of goodness in the following way: If an action is meritorious or correct, it is better to perform it than not perform it, and if it is reprehensible or excusable, omitting the action is preferable to doing it. Meinong's schema does not include (morally or normatively) indifferent actions; they are obviously actions whose commission is neither more nor less preferable than their omission. According to the simple act-utilitarian or optimizing consequentialist account of the concept of obligation, an act is obligatory (required) whenever it is better (or has better consequences) than its omission, but this analysis of the concept of obligation leaves no room for meritorious but optional (that is, supererogatory) actions, and this has often been regarded as one of the weaknesses of act-utilitarianism. In this respect the act-utilitarian account does not agree with our conception of the relations between normative concepts and value concepts. Some attempts to define obligation in terms of 'better' and rely on a logic for the latter to derive a logic for former, are subject to the same difficulty; for example, Lennart Åqvist's [1963, p. 286] definition

(DOÅqv) $\mathbf{O}p \leftrightarrow p\mathbf{B}\neg p.$

Meritorious, correct, and excusable actions are actions an agent may perform, and belong to the deontic category of the permitted (permitted actions); inexcusable actions are not permitted (that is, are prohibited), and Meinong's concept of 'correct' can be regarded as equivalent to the

[5]Regarding Meinong's terminology, the intention is clearly that the four categories are to be mutually exclusive. We add "merely" in front of "Excusable" because if (c) is simply the complement of (d), the inexcusable, then it must include (b) and (a), which are also not inexcusable surely; but the intention of (c) seems to be a category of offense or suberogation—permissible yet blameworthy. Regarding (b), surely the meritorious is also "correct", but Meinong seems to intend what is obligatory but not meritorious, else (b) overlaps with (a), again violating mutual exclusivity. (a) seems to be intended for a category like the supererogatory, something that is meritorious and optional. There are a variety of interesting subtleties here we must pass over. See [McNamara, 1996a; McNamara, 1996b; McNamara, 1996c; McNamara, 2011a; McNamara, 2011b] for closely related issues.

deontic concept of obligation or requirement, not of permitted as might be suggested by "correct".

Meinong's deontological-axiological categories are represented in Table 1. The category of indifferent actions has been added here to Meinong's schema. 'A' represents an action type ("generic action"), $V(A)$ is the value of A (in a given situation), $\mathbf{om}(A)$ represents the omission of A, and deontic and axiological concepts are symbolized as follows.

$\mathbf{L}A = A$ is laudable (meritorious, supererogatory),
$\mathbf{O}A = A$ is obligatory (required),
$\mathbf{I}A = A$ is normatively indifferent,
$\mathbf{E}A = A$ is excusable (suberogatory),
$\mathbf{P}A = A$ is permitted, that is, not forbidden, and
$\mathbf{F}A = A$ is forbidden (not permitted, reprehensible).

These categories are represented in Table 1 and linked to corresponding value notions.

<table>
<tr><td rowspan="4">$\mathbf{P}A$: A is permitted</td><td>$\mathbf{L}A$: A is meritorious (laudable, supererogatory, verdienstlich)</td><td rowspan="2">A is good, an action with a positive value $V(A) > V(\mathbf{om}A)$</td></tr>
<tr><td>$\mathbf{O}A$: A is obligatory (required, correct)</td></tr>
<tr><td>$\mathbf{I}A$: A is indifferent</td><td>A is indifferent $V(A) = V(\mathbf{om}A)$</td></tr>
<tr><td>$\mathbf{E}A$: A is excusable (zulässig)</td><td rowspan="2">A is bad, an action with a negative value, an undesirable action $V(A) < V(\mathbf{om}A)$</td></tr>
<tr><td>$\mathbf{F}A$: A is forbidden (prohibited)</td><td>$\mathbf{F}A$: A is forbidden (reprehensible, verwerflich)</td></tr>
</table>

Table 1: Meinong's deontological-axiological action categories

To say that an action has positive value means here that it is preferred to its omission. The value of an action, $V(A)$, need not be regarded as interval measurable; $V(A) < V(B)$ may be taken to mean only that B is strictly preferred to A, $V(A) = V(B)$ means that there is no noticeable value-difference between A and B. The arrangement of the five categories in Table 1 does not mean that supererogatory actions are invariably better (more valuable) than obligatory (required) actions, but it is clear that both are better than normatively indifferent actions.

Meinong formulated a "law of omission" concerning the four main deontological-axiological categories, according to which an action A is meritorious if and only if its omission is excusable, excusable if and only if its omission is meritorious, correct (obligatory) if and only if its omission is reprehensible (forbidden), and reprehensible if and only if its omission is correct. ([1968b, p. 89] and [1968a, p. 32]) These laws are expressed by the following formulas:

(2.2) $$\mathbf{L}A \leftrightarrow \mathbf{Eom}A$$

(2.3) $$\mathbf{E}A \leftrightarrow \mathbf{Lom}A$$

(2.4) $$\mathbf{E}A \leftrightarrow \mathbf{Lom}A$$

(2.5) $$\mathbf{O}A \leftrightarrow \mathbf{Fom}A$$

Moreover, according to the definability of $\mathbf{P}$ in terms of $\mathbf{F}$, we have

(2.6) $$\mathbf{P}A \leftrightarrow \neg\mathbf{F}A \leftrightarrow \neg\mathbf{Oom}A$$

where '$\neg$' is the sign of propositional negation. (2.4)-(2.6) are analogous to the standard interdefinability principles for deontic operators. If the **om**–operator is formally analogous to negation and satisfies the principle of "double omission (negation)", then (2.3) follows from (2.2) and (2.5) follows from (2.4). Meinong does not accept this principle; according to him, the omission of an act requires an opportunity to perform the act [1968a, pp. 691-2]. However, if the concept of omission is interpreted as not-doing or if we consider only actions which are possible for an agent to perform or not to perform in a given situation, the **om**-operator is analogous to propositional negation, and subject to the principle of "double omission".

According to Meinong, normative and axiological concepts can be defined in terms of the praiseworthiness or blameworthiness of actions and their omissions. The omission of a correct (obligatory, required) act is always forbidden and deserves blame, which can be regarded as a social or moral sanction associated with the action. The performance of a meritorious act is praiseworthy, that is, deserves praise or a reward. An act is excusable if its performance does not deserve blame or praise; in this respect excusable actions do not differ from indifferent actions, but the omission of an excusable action is praiseworthy. Thus Meinong's definitions suggest the following analysis of the main axiological and deontological concepts in terms of a reward ($\mathbf{R}$) or sanction or punishment ($\mathbf{S}$).

(2.7) $\mathbf{L}A \leftrightarrow A \ni \mathbf{R}$,
(2.8) $\mathbf{O}A \leftrightarrow \mathbf{om}A \ni \mathbf{S}$,
(2.9) $\mathbf{E}A \leftrightarrow \mathbf{om}A \ni \mathbf{R}$,

and

(2.10) $\mathbf{F}A \leftrightarrow A \ni \mathbf{S}$,

where the letter '∋' (the Cyrillic 'E') signifies the association between an action type and a sanction **S** or reward **R**. If 'A' and '**om**A' are read as the propositions that A is performed or omitted, the sign ∋ may be read as a sign of a defeasible or non-defeasible conditional. If '∋' is read as a strict (necessary) conditional and '**om**' is read as a sign for propositional negation, (2.8) becomes equivalent to Alan Ross Anderson's [1956, p. 169] and [1958] proposal to reduce deontic logic to alethic modal logic by means of the translation

(2.11) $\mathbf{O}A \leftrightarrow \mathbf{N}(\neg A \rightarrow \mathbf{S})$.

According to Meinong's interpretation of (2.1)-(2.4), **S** represents strong negative value-feelings, and **R** stands for positive value-feelings [Meinong, 1968b, pp. 73-4] and [Sajama, 1988, p. 75].

$\mathbf{O}A$ implies that doing A is better than its omission, but the converse does not hold. According to Meinong's analysis of value and obligation, the principle

(2.12) $V(A) \leq V(A') \rightarrow (\mathbf{O}A \rightarrow \mathbf{O}A')$

does not hold, a meritorious act may be better than an obligatory act. An obligatory or normatively required action need not be the best or optimal action available to the agent in a given situation.[6] (2.12) may be called the principle of value-positivity; Sven Ove Hansson calls it the principle of preference positivity (≥'-positivity, where ≥' is the preference relation associated with **O**; see [Hansson, 2001, p. 115]). Hansson gives this counter-example to (2.12): Serving a fine dinner to unexpected guests may be better than offering them something to eat and drink, but the former need not be a moral or social requirement if the latter is so [Hansson, 2001, p. 146]. On the other hand, Meinong's system is consistent with the rule of value-positivity (preference positivity) for the (standard) concept of permission (as defined in Table 1), that is,

(2.13) $V(A) \leq V(B) \rightarrow (\mathbf{P}A \rightarrow \mathbf{P}B)$.

[6] Although the best act will then be one that fulfills the obligation, and so the obligatory act (e.g. providing some help) will be done if the more specific best act (providing maximal help) is done.

Given the definition of $\mathbf{F}$ as $\neg\mathbf{P}$, (2.13) entails the principle of value-negativity (preference-negativity),

$$(2.14) \qquad V(A) \leq V(B) \rightarrow (\mathbf{F}B \rightarrow \mathbf{F}A).$$

(2.13) and (2.14) are based on the plausible assumption that a permissible action (including an excusable one) cannot be worse than an inexcusable (prohibited) action.

It is interesting to note here that in the 11th century, the Islamic rationalist philosopher Abd-al-Jabbār (935-1025) presented a schema essentially similar to that proposed by Meinong and distinguished normative categories on the basis of whether a given act or its omission deserves blame or praise. Like Meinong, he distinguished four main normative categories ([Hourani, 1975, pp. 100-1], [Hilpinen, 1985, pp. 100-1] and [Sajama, 1988, p. 80])

(2.15) (i) An act A is an act of grace (*tafaddul*) or recommended (*nadb*) if and only if the doer deserves praise, the omitter does not deserve blame.
(ii) A is obligatory (*wājib*) if and only if the omitter deserves blame.
(iii) A is merely permissible (optional, *mubāh*) if and only if neither the doer nor the omitter deserves blame or praise.
(iv) A is evil (*qabīh*) if and only if the doer deserves blame.

The actions in categories (i)-(iii) are described as "good" (*hasan*) actions (Hourani 1985, 101); the word 'acceptable' might be more suitable. Hourani's use of 'permissible' for category (iii) may be misleading; 'optional' seems more apt. The main difference between (2.15) and Meinong's definitions (2.7)-(2.10) is that (2.15.iii), a "permissible" (i.e., optional) action corresponds to the category of normative indifference ($\mathbf{I}$), and the category of excusable actions ($\mathbf{E}$) is missing from Abd-al-Jabbār's schema. The omission of a meritorious (laudable) action does not deserve blame, i.e.,

$$(2.16) \qquad \mathbf{L}A \rightarrow \neg(\mathbf{om}A \ni \mathbf{S}),$$

(This is obvious if $\mathbf{L}$ represents supererogatory actions.) An analogous principle holds for the other categories, for example, the commission of an obligatory action does not generally merit a reward. Thus we may adopt the following principles:

$$(2.17) \qquad A \ni \mathbf{S} \rightarrow \neg(\mathbf{om}A \ni \mathbf{R})$$
$$(2.18) \qquad \mathbf{om}A \ni \mathbf{S} \rightarrow \neg(A \ni \mathbf{R})$$

(2.19) $A \ni \mathbf{R} \to \neg(\mathbf{om}A \ni \mathbf{S})$
(2.20) $\mathbf{om}A \ni \mathbf{R} \to \neg(A \ni \mathbf{S})$

(2.17) does not mean that an agent cannot be rewarded after fulfilling his duty, only that such a reward is contingent, and not associated with the duty by a general rule or custom. According to [Sajama, 1988, pp. 77-8, n. 11], some formulations of the *sharia* law violate (2.18) and (2.17) by defining an obligatory action (a duty) as an action whose performance deserves reward and omission a punishment, and a forbidden action as one whose performance deserves punishment and omission a reward. (Sajama refers to [Hartmann, 1987, p. 60].)

3 Ernst Mally's *Deontik*

Meinong gave a conceptual analysis of some axiological and normative concepts and investigated their interrelations, but apart from the formulation of the Laws of Omission, he did not attempt to develop a systematic logical theory in the field. Such an attempt was made by his student Ernst Mally [1971] who was inspired by the formal axiom systems of logic developed in the early 20th century, especially by that of Russell and Whitehead's *Principia Mathematica*. He wanted to develop a formal logic for the concepts of ought (*Sollen*) and the concept of willing something (*Wollen*).

According to Mally, judging (*Urteilen*) and willing are distinct attitudes towards states of affairs. Classical logic is concerned with judgments, and Mally proposed to develop a parallel logic for the attitude of willing. Willing that a certain state of affairs p should obtain was expressed by sentence of the form 'It ought to be that p' (p *soll sein*); and Mally thought that Deontik, the logic of ought, can also serve the logic of will (or willing). [Mally, 1971, p. 241] The non-logical signs of Mally's system represent (possible) states of affairs, not actions (or action types), thus his logic may be conceived as a logic of the ought-to-be (*Seinsollen*) rather than a logic of the ought-to-do (*Tunsollen*). In the following discussion we shall frequently not make a distinction between the concepts of *ought* and *being obligatory* (or the concept of obligation), even though these terms are often not interchangeable in ordinary speech [McNamara, 1990; McNamara, 1996c]. Occasionally though we will note issues that arise in assuming they are equivalent.

Mally's deontic logic is based on classical propositional logic. Its vocabulary consists of a sign for the concept of ought, the standard truth-functional connectives, and sentence letters. Here we shall use p, q, r, ... as sentence letters, and the usual symbol $\mathbf{O}$ as the ought-operator (instead of Mally's '!'). In addition, Mally uses propositional constants for the "unconditionally" or "actually" ("tatsächlich") obligatory, here expressed as '**u**'; for the "negation" of **u** ("das Sollenswidrige"), '**n**'; for what is the case (the facts, '**w**'), and for what is not the case ("das Untatsächliche"), here expressed by the letter '**m**' [Mally, 1971, p. 239, pp. 249-50], as well as an existential quantifier over propositions. [Mally, 1971, pp. 249-50]

Mally reads '$p \to \mathbf{O}q$', as 'p requires q' ('p fordert q'), and abbreviates it '$p\boldsymbol{f}q$', that is,

(Df.$\boldsymbol{f}$) $\quad (p\boldsymbol{f}q) \leftrightarrow (p \to \mathbf{O}q)$.

The arrow '$\to$' is a sign for the truth-functional conditional; '$p \to q$' means that "it is not the case that p and not q." [Mally, 1971, p. 243] Mally adopts the following axioms for the **O**-operator [Mally, 1971, pp. 246-50]:

(MA1) $\quad ((p \to \mathbf{O}q) \,\&\, (q \to r)) \to (p \to \mathbf{O}r)$

(MA2) $\quad ((p \to \mathbf{O}q) \,\&\, (p \to \mathbf{O}r)) \to (p \to \mathbf{O}(q\&r))$

(MA3) $\quad (p \to \mathbf{O}q) \leftrightarrow \mathbf{O}(p \to q)$

(MA4) $\quad (\exists \mathbf{u})\, \mathbf{Ou}$

(MA5) $\quad \neg(\mathbf{u} \to \mathbf{On})$

Mally takes (MA4) to mean that there is an unconditionally obligatory state of affairs **u** (ibid., 249), but if '**u**' is a constant, the quantifier in (MA4) is superfluous, and it can be simplified to

(MA4') $\quad \mathbf{Ou}$.

Since **n** is a state of affairs logically incompatible with **u**, (MA5) can be expressed as

(MA5') $\quad \neg(\mathbf{u} \to \mathbf{O}\neg\mathbf{u})$.

Mally calls (MA5) the principle of the consistency of the unconditionally (or actually) obligatory. [Mally, 1971, p. 250]. For the constants **w** and **m**, Mally adopts the principles (or schemata) [Mally, 1971, p. 239]:

(3.1) $\quad$ For any $p, p \to \mathbf{w}$,

and

(3.2) $\quad$ For any $p, \mathbf{m} \to p$.

Mally's attempt to systematize deontic logic was unsuccessful, as some of his critics were quick to point out. Karl Menger was the first to show that the consequences of Mally's axioms include the theorem

(3.3) $\mathbf{O}p \leftrightarrow p.$

(For proofs of (3.3) from Mally's axioms, see [Føllesdal and Hilpinen, 1971, p. 4], [Lokhorst and Goble, 2004, pp. 45-6] and [Lokhorst, 2008].) Menger observed that because of (3.1), the introduction of the sign $\mathbf{O}$ for the concept of ought is "superfluous in the sense that it may be cancelled or inserted in any formula at any place you please". ([Menger, 1939, p. 58] quoted from [Lokhorst and Goble, 2004, p. 46] and [Lokhorst, 2008].) We also get the theorem

(3.4) $\mathbf{Ow}$,

which Mally takes to mean that a state of affairs that actually obtains ought to obtain, or that "the facts are unconditionally required" [Mally, 1971, p. 266]. Menger also pointed out that Mally's system is incompatible with his own informal remarks on the concept of ought, for example, that $\mathbf{O}(p \vee q)$ is not equivalent to $\mathbf{O}p \vee \mathbf{O}q$; the latter entails the former but the converse does not hold. ([Mally, 1971, p. 260] and [Menger, 1939]) Mally himself thought that some consequences of his axioms are counter-intuitive or "strange", including (3.3) and (3.4), but instead of revising the system, he tried to interpret his theory as the theory of "correct willing", willing in accordance with the facts ("richtiges Wollen"; [1971, pp. 286ff.] and [Lokhorst, 2008].) A. Hofstadter and J. C. C. McKinsey's attempt to develop a logic of imperative discourse (1939) was subject to the same problem as Mally's system; in their system an imperative '$!p$' (where ! is the imperative operator) turned out to be equivalent to 'p', and the imperative sign was logically superfluous [1939, p. 453].

The interpretation of the constants $\mathbf{u}$, $\mathbf{n}$, $\mathbf{w}$, and $\mathbf{m}$ is unclear. Principles (3.1) and (3.2) suggest that $\mathbf{w}$ should be regarded as the constant *Verum* (or $\top$), a sentence which is true in every possible situation, that is, any tautology. According to this interpretation, (3.4) holds in any normal system of modal (deontic) logic, and is not necessarily objectionable or "surprising", since it then says that whatever cannot not be ought to be. On the other hand, as Menger pointed out, (3.3) would reduce deontic logic to classical propositional logic, and consequently trivialize it.

It is easy to see that Mally's axioms (MA1) and (MA3), unlike (MA2), are not intuitively valid. According to (MA3), a wide-scope obligation $\mathbf{O}(p \rightarrow q)$

is equivalent to a narrow-scope obligation ($p \to \mathbf{O}q$), but this is clearly not the case. $\mathbf{O}q$ follows logically from the assumptions p and $p \to \mathbf{O}q$ by Modus Ponens, but not from p and $\mathbf{O}(p \to q)$. For example, rationality presumably requires that if you believe that the world was made in six days, you believe that it was made in less than a week. But if you believe that the world was made in six days, it does not follow that you ought to believe that it was made in less than a week, on the contrary, you ought not to believe that, because it is false, and you ought not to believe what you actually believe, viz., that the world was made in less than a week. (This example is from [Broome, 2004, p. 29].) It is also easy to find counterexamples to Mally's first axiom. For example, assume that Brutus is your neighbor's bad-tempered dog which is sometimes let outside on his yard. Then the following sentences may be assumed to be true:

(3.5) (i) If Brutus is outside, the gate ought to be closed.
(ii) If the gate is closed, I am not afraid of Brutus.[7]

Then, according to (MA1),

(3.5) (iii) If Brutus is outside, I ought not to be afraid of him.

However, (3.5.iii) is false in a situation in which the gate is not closed, that is, if the requirement in the consequent of (3.5.i) is not satisfied. As [Hintikka, 1971, p. 82] has noted, "One can 'escape' the obligation that r [in (MA1)] simply by failing to carry out the duty expressed by $\mathbf{O}q$."

Some of Mally's successors made similar mistakes in judging the validity of deontic formulas. In a short paper on the logic of imperatives and deontic propositions Kurt Grelling [1939, p. 45] put forward the following rule of deontic logic:

(3.6) If r follows from p and q, the conjunction of p and $\mathbf{O}q$ implies $\mathbf{O}r$

If the expression 'follows' (Grelling's "folgt") is interpreted as a truth-functional conditional, (3.6) can be formalized as

(3.7) $((p \,\&\, q) \to r) \to ((p \,\&\, \mathbf{O}q) \to \mathbf{O}r)$,

[7]In examples throughout, as well as in textual discussions of examples and in our expositions of positions, we will often use the first person singular "I" rather than the more cumbersome "we". It is an invitation for the reader to identify with the position being discussed or agent in focus as an expositional tool.

that is,

(3.8) $((p \to (q \to r)) \to (p \to (\mathbf{O}q \to \mathbf{O}r))$.

This schema seems invalid. It may be true that if Brutus is outside, I am not afraid of Brutus if the gate is closed, but false that if Brutus is outside, I ought not to be afraid of Brutus if the gate ought to be closed; on the contrary, I ought to be afraid if the gate is in fact not closed even if it ought to be. (Cf. [Hintikka, 1971, p. 83].) In a note on Grelling's paper Karl Reach [1939, p. 72] observed that if q is replaced by $\neg p$ and r by p, (3.7) becomes

(3.9) $((p \,\&\, \neg p) \to p) \to ((p \,\&\, \mathbf{O}\neg p) \to \mathbf{O}p)$.

The antecedent is a logical truth, thus Grelling's rule implies

(3.10) $(p \,\&\, \mathbf{O}\neg p) \to \mathbf{O}p$,

which means that if something that ought not to be the case is the case, it ought to be the case.

The problems with Mally's system are mainly due to his failure to distinguish wide-scope *oughts* (obligations) from narrow scope *oughts*, and partly to a problematic interpretation of conditionals and the word 'implies'. Mally reads the truth-functional conditional '$p \to q$' as 'p implies ("impliziert") q' [1971, p. 238], but this word often means some kind of strict implication. According to Mally's axiom (MA3), $p \to \mathbf{O}q$ is equivalent to $\mathbf{O}(p \to q)$, and it should be possible to express (3.5.i) as

(3.11.i) It ought to be that if Brutus is outside, the gate is closed,

using the somewhat artificial construction 'It ought to be that' to indicate a wide-scope ought. In the same way, the suggested conclusion may be expressed as

(3.11.iii) It ought to be that if Brutus is outside, I am not afraid of him.

If the second premise (3.5.ii) is interpreted as a strict (necessary) conditional or as a *deontic* implication of the form $\mathbf{O}(q \to r)$, that is, as

(3.11.iin) In all possible circumstances (situations), if the gate is closed, then I am not afraid of Brutus, or as

(3.11.iid) It ought to be that if the gate is closed, then I am not afraid of Brutus.

the conclusion follows from (3.11.i) and (3.11.ii). If things are the way they ought to be, the gate is closed if Brutus is outside, and I need not be afraid of him. Here the possible circumstances should be taken as the situations which differ from the actual situation only with respect to Brutus's location (in the house or outside on the yard), the gate's being open or closed, and my state of fear or lack of fear. The wide-scope ought $\mathbf{O}(p \rightarrow q)$ is *prima facie* a more plausible representation (or partial representation) of the normative relation of requirement between possible states of affairs (Mally's '$p\boldsymbol{f}q$') than the narrow-scope ought, and as we have seen, they are not equivalent. (Cf. [Broome, 1999, pp. 401-5] and [Broome, 2004, p. 29].)

4 On the interpretation of deontic logic

Mally's and Grelling's failure to formulate workable principles of the logic of norms may have reinforced the skepticism expressed by some authors in the late 1930's and early 1940's about the very possibility of the logic of norms and imperatives.

In the late 1930's Jørgen Jørgensen and a number of other philosophers considered the following problem concerning the logic of imperatives and directives. According to the standard conception of logical entailment, a conclusion follows logically from certain premises if and only if the conclusion cannot be false if the premises are true. Thus it is essential for logical inference that the premises and the conclusion are sentences which can be true or false. But since imperatives cannot be said to be true or false, they cannot function as the premises or conclusions of logical inferences, and it is therefore in principle impossible to justify an imperative by means of logical reasoning. [Jørgensen, 1938, p. 184] On the other hand, Jørgensen notes that it seems equally evident that there are "inferences in which one or both premises as well as the conclusion are imperative sentences, and yet the conclusion seems just as inescapable as the conclusion in any syllogism containing sentences in the indicative mood only." [Jørgensen, 1937 and 1938, p. 290] Jørgensen gives the following example (*loc. cit.*):

Love your neighbor as yourself!
Love yourself!
(Therefore:) Love your neighbor!

This seems to be an example of logically valid reasoning with imperatives.

Jørgensen's countryman Alf Ross called this problem "Jørgensen's dilemma" [Ross, 1941, p. 55]. The word 'imperative' should be taken here to refer to an imperative speech act or what is expressed by it, for example, a command or a directive, not to the grammatical mood of a sentence. The word may be regarded here as interchangeable with 'directive' or 'command'. (For imperatives and commands, see [Aikhenvald, 2010, p. 1-16] and [Lyons, 1977].) It is clear that Jørgensen's dilemma concerns normative discourse in general. Norms cannot be said to be true or false, and if deontic logic is defined as the logic of norms, Jørgensen's dilemma is a problem for deontic logic.

This problem continues to engage philosophers. G. H. von Wright published in the 1990's a paper entitled 'Is There a Logic of Norms?' [von Wright, 1996], and [Makinson, 1999, p. 29] has called Jørgensen's dilemma "a fundamental problem of deontic logic". Von Wright formulated the problem as follows: "Since norms are usually thought to lack truth-value, how can logical relations such as contradiction and entailment (logical consequence) obtain between norms?" [1996, p. 35] (but see also [von Wright, 1983, pp. 130-1]). Makinson observed that "there is a singular tension between the philosophy of norms and the formal work of deontic logicians", because "the usual presentations of deontic logic .... treat norms as if they could bear truth-values", but "it makes no sense to describe norms as true or false" [1999, p. 29-30].

[Jørgensen, 1937 and 1938, p. 290] suggests two possible ways out of this dilemma.

1. We may widen the concept of logical consequence in such a way that it need not be defined in terms of the concept of truth, but some semantic feature of norms or imperatives which can be regarded as analogous to truth. (Cf. [Grue-Sörensen, 1939, p. 197].) According to this proposal, logic can be said to have "a wider reach than truth" [von Wright, 1957, vii].

2. We might also try to solve the puzzle by defining the concept of validity for reasoning about norms (or imperatives) indirectly, in terms of the truth-values of propositions which are related to norms in a suitable way. In this way of dealing with the puzzle, the logical relations among norms and imperatives are regarded as being constituted by relations among certain propositions associated with them.

[Hofstadter and McKinsey, 1939, p. 447] adopted the first approach, and suggested that the concept of *satisfaction* can replace the concept of truth in the definition of validity and inconsistency for imperatives. An imperative or a directive cannot be said to be true or false, but it can be satisfied or not satisfied by the actions of the addressee. An imperative is satisfied if (and only if) what is commanded is the case. G. H. von Wright has made a similar proposal concerning the logic of norms, and suggested that deontic logic can be understood as "a logic of norm-satisfaction" [von Wright, 1983, p. 130, pp. 138-142]. In another variant of this approach, logical relations among directives are defined in terms of the "validity" (or "correctness") of a directive or a norm so that the concept of (norm) validity plays the same role in the analysis of normative reasoning as the concept truth in "indicative" reasoning. To distinguish this use of the word 'valid' from the concept of logical validity used in the evaluation of an argument ('argument validity'), it may be called 'norm validity'. For example, Alf Ross has argued that our conception of logically valid normative reasoning is based on the concept of norm validity: "The logical deduction of [a directive] I_2 from I_1 then means that I_2 has objective validity in case I_1 has objective validity." [Ross, 1941, p. 59] The validity of a norm means its " 'existence' or 'being in force' – however these expressions are to be understood." [Ross, 1968, p. 175] It has also been suggested that it is possible to distinguish two logics of imperatives and norms, the logic of satisfaction and the logic of validity, which are not the same. ([Segerberg, 1990, p. 203] and [Ross, 1944, pp. 39-43].)

Jørgensen prefers the second approach, following a proposal made by Walter Dubislav. According to [Dubislav, 1937, p. 341], every directive ("Forderungssatz") D is related to a certain statement ("Behauptungssatz") s(D) in such a way that our judgments about the logical relations among directives are determined by the logical relations among the corresponding statements. This proposal may be also be expressed by using the word 'proposition': a directive G can be inferred from D is and only if the proposition s(G) associated with G is a logical consequence of s(D). A set of directives or norms is regarded as inconsistent if and only if the set of the corresponding proposition is inconsistent. What we are inclined to take as logical relations among norms or imperatives are really relations among the propositions associated with the norms, or derived from such relations.

The philosophers who have adopted this conception of the logic of norms have understood the relevant proposition associated with a norm in different ways. According to Jørgensen, an imperative (or directive) can be analyzed into two parts, the "imperative factor" and the "indicative factor". The former indicates that something is commanded or requested, and the latter describes what is commanded, the content of the command [Jørgensen, 1937 and 1938, p. 291]. The indicative factor of the directive

(4.1) Bertie, pinch the cow-creamer!

can be taken to be the proposition that Bertie pinch the cow-creamer. To indicate that a proposition is not asserted, it may be expressed in a subjunctive form or by an infinitive clause:

(4.2) Bertie to pinch the cow-creamer.

As C. S Peirce's observed, "the proposition in the sentence, 'Socrates est sapiens', strictly expressed, should be 'Socratem sapientem esse'." [Peirce, 1998, p. 312] The content may also be expressed by an indicative sentence; hence the term "indicative factor". If the imperative factor (or directive factor) is expressed by the exclamation mark '!', (4.1) has (according to Jørgensen) the form

(4.3) !(Bertie to pinch the cow-creamer).

The distinction between the content and the directive factor of a directive is a special case of the distinction between the illocutionary character and the content of a speech act. If D $= !p$, where p is a proposition, p is the "indicative" (the proposition or statement) s(D) which determines the logical relations of D to other directives, that is,

(4.4) $\mathrm{s}(!p)=p$.

According to (4.4), imperative reasoning (or reasoning with directives) as reasoning about their propositional contents:

(4.5) An imperative $!q$ is said to be derivable from $!p$ if and only if the proposition q is derivable from p.

Here "the imperative factor is so to speak put outside the brackets much as the assertion-sign in the ordinary logic [logic of statements], and the logical operations are only performed within the brackets" [Jørgensen, 1937 and 1938, p. 292]. The logic of imperatives is thus reduced to the logic of statements for which the concept of logical consequence can be defined in the usual way, and "there seems to be no reason for, and hardly any possibility of, constructing a special 'logic of imperatives'." [Ross, 1941, p. 57].

In this way of analyzing imperative inference, the "indicative" s(D) associated with a given directive or norm is assumed to express the propositional content of the directive. Instead of the "indicative factor" we may use the term 'semantic component' to refer to the content which determines the logical relations among directives and norms.

According to Jørgensen, the logic of imperatives can also be based on another way of transforming imperatives into indicatives. In this method, imperative sentences are transformed into statements which say that "the ordered actions are to be performed, resp. the wished state of affairs is to be produced." [Jørgensen, 1937 and 1938, p. 292] According to this proposal, the content of the command "Pinch the cow-creamer!" may be expressed as "The cow-creamer is to be stolen." Thus the semantic content of the command (4.1) is expressed by

(4.6) Bertie is to pinch the cow-creamer.

The sentence (4.6) does not differ much from (4.2), but it is possible to see a significant difference in meaning. Unlike (4.2), (4.6) can be interpreted as having normative (deontic) content, and can be regarded as equivalent to 'Bertie must pinch the cow-creamer' (or 'Bertie ought to pinch the cow-creamer'), where the word 'must' functions as a deontic operator. If the requirement (or obligation) expressed or created by a command is expressed by the deontic O-operator, (4.6) can be written as

(4.7) **O**(Bertie to pinch the cow-creamer)

According to this construal of the logic of directives,

(4.8) $\mathrm{s}(!p) = \mathbf{O}p$,

where $\mathbf{O}p$ is a deontic proposition. This way of correlating norms and directives with deontic sentences (understood as "indicatives") helps to solve Jørgensen's problem if it is supplemented by an account the truth-conditions or "truth-makers" of deontic propositions. (For the concept of a truth-maker, see [Mulligan *et al.*, 1984; Armstrong, 2004].) How should the meaning of such sentences be understood? Like many logical empiricists of his time, Jørgensen formulated this question as a question about the verifiability of deontic sentences: "How is a sentence of the form 'Such and such is to be so and so' to be verified?" [Jørgensen, 1937 and 1938, p. 292]. His answer was that the phrase "is to be etc." describes a "quasi-property" ascribed to an action or a state of affairs when "a person is willing or commanding the action to be performed, resp. the state of affairs to be produced." According to him, the sentence "Such and such action is to be performed" may be regarded as an abbreviation of the sentence form

(4.9) There is a person who is commanding that such and such action is to be performed.

Sentences of this form state only that some normative source is issuing a certain command. According to this proposal, (4.6) corresponds to

(4.10) It is commanded that Bertie pinch the cow-creamer.

However, it is clear that (4.10) is understood, it is not equivalent to (or synonymous with) (4.6) or (4.7). It does not necessarily have any normative import.

Nevertheless Jørgensen's proposal suggests a possible truth-maker (or "falsity-maker") for deontic propositions: Often certain speech acts or other actions, for example, the actions of legislative bodies, judicial decisions, and contracts between individuals, function as truth-makers of legal ought-sentences. This does not hold for all deontic propositions, for example for moral *ought*s and obligations. In this case different metaethical theories can be regarded as theories about the truth-makers of deontic propositions. A moral realist may hold the view that "there are objective normative facts, existing independently of our conceptualization and thinking" [Tannsjö, 2010, p. 38], or we may say that an agent has a moral obligation to perform an action if and only if its omission would violate the interests of the persons affected by the action. If the word 'ought' (or 'must') is regarded as an expression of practical necessity or of a prudential ought, (4.6) is true if and only if Bertie's theft of the cow-creamer is necessary for satisfying Bertie's or some other person's current interests or the best way of satisfying such interests. The nature of the truth-makers of ought-sentences depends on the kind of ought (or obligation) under consideration.

Deontic logicians have often made a distinction between two interpretations of deontic sentences. It has been suggested that a deontic sentence of the form $\mathbf{O}p$ can be interpreted normatively (or prescriptively) as expressing a mandatory norm, or descriptively as a statement that it is obligatory that p according to some unspecified system of norms. ([von Wright, 1963, viii, pp. 104-5], [Stenius, 1963, pp. 250-1], [Alchourrón, 1969, pp. 243-5], [Alchourrón and Bulygin, 1971, p. 121], [Hansson, 1971, p. 123], [Bulygin, 1982, pp. 127ff] and [Alchourrón and Bulygin, 1993, p. 285].) According to [von Wright, 1963, viii]:

> "The deontic sentences of ordinary language ... exhibit a characteristic ambiguity. Tokens of the same sentence are used, sometimes to enunciate a certain prescription (*i.e.*, to enjoin, permit, or prohibit a certain action), sometimes again to express a

> proposition to the effect that *there is* a prescription enjoining or permitting or prohibiting a certain action."

For example, the deontic sentence 'The Florida Keys must be evacuated' can function as an evacuation order given by the authorities of Monroe County to the inhabitants of the Keys before the arrival of a hurricane, or as a proposition which gives information about current evacuation orders in South Florida. The announcement 'You may return to your homes' can likewise issue the permission abrogating the prior order, or instead report the fact that the order has already been canceled. Von Wright calls propositions of the latter kind *norm-propositions* or *normative statements* [1963, viii, p. 105]. We shall use for this purpose below the term 'norm-statement'. Norm-statements, unlike the norms themselves, can be said to be true or false, and the logical relationships among them can therefore be understood in the usual way in terms of the concept of truth. The descriptive interpretation of deontic sentences and formulas is essentially the same as Jørgensen's second method of associating indicatives with imperatives. This distinction solves Jørgensen's problem for deontic logic if it is regarded as the logic of normative statements, statements about the existence of norms. ([Stenius, 1963, p. 251] and [Hansson, 1971, p. 123].) However, Carlos Alchourrón and Eugenio Bulygin have made a distinction between the logic of norms and the logic of normative statements, and argued that the logic of descriptively interpreted normative statements differs from the logic of norms (deontic logic proper), and therefore cannot serve as a substitute for the latter, nor can the latter be derived from the former. ([Alchourrón, 1969], [Alchourrón and Bulygin, 1971, pp. 121-7] and [Alchourrón and Bulygin, 1993, p. 285].) They distinguish two sets of deontic operators, the "prescriptive" operators O and P, and the "descriptive" operators $\mathbf{O}_\alpha$ an $\mathbf{P}_\alpha$, where α refers to a system of norms and rules, for example, a certain system of legal or moral norms. According to Alchourrón and Bulygin, the principle of consistency

(4.11) $\mathbf{O}_\alpha p \rightarrow \neg\mathbf{O}_\alpha\neg p$

does not hold for the descriptive **O**-operator for a norm system α: $\mathbf{O}_\alpha p\,\&$ $\mathbf{O}_\alpha\neg p$ should be regarded as consistent, because a legislator can promulgate two incompatible norms and a norm system may generate norm conflicts; the existence of such a system is not logically impossible [1993, pp. 290-1] On the other hand, the counterpart of (4.11) for the "prescriptive" Ought is valid, because a norm (prescription) commanding that p be the case is inconsistent with a norm commanding that $\neg p$ be the case [1993, p. 283]. If the logic of normative statements cannot serve as the foundation for the logic of norms, the possibility of (apparent) logical relations between norms and directives must be explained in some other way. Alchourrón and Bulygin introduce for this purpose the concept of *norm-lekton* as the content of a possible prescription. A norm-lekton is related to a prescription in the same way as a proposition to an assertion; the content of a possible assertion is a proposition [1993, pp. 275-6]. The consistency and other logical properties of norms are constituted by the logical properties of norm-lekta and the relations between them; thus the logic of norms (deontic logic proper) is, strictly speaking, the logic of norm-lekta. Moreover, consistency is not a necessary condition for the existence of norms, because a norm-authority can promulgate incompatible norms (i.e., norms with mutually inconsistent lekta) or prescriptions which can lead to conflict situations; hence the difference between the logic of norms and the logic of normative statements (statements about norms). [Alchourrón and Bulygin, 1993, pp. 281-2]

The term 'lekton' ($\lambda\epsilon\kappa\tau\acute{o}\nu$) was used in Stoic logic to refer to the sense of an expression or utterance. The lekta were divided into incomplete and complete lekta. The latter were the contents of complete speech acts, and were divided further into propositions, questions, commands (imperatives), and other kinds. [Mates, 1965, pp. 16-9] Alchourrón and Bulygin's notion of a norm-lekton as the semantic content of a norm agrees in this respect with the Stoic account of lekta.

Alchourrón and Bulygin call the view that norms and directives have norm-lekta as their semantic content the "hyletic conception of norms". According to an alternative view, the "expressive conception" [1981, pp. 95-9][1993, pp. 273-4], the normative component of a norm is not part of its semantic content, but indicates only how the content is presented, that is, as a command or prescription rather than a statement about matters of fact; thus there are no special "norm-lekta" distinct from ordinary descriptive propositions. In the discussion of Jørgensen's problem above, proposal (4.4) represents the expressive conception, (4.8) exemplifies the hyletic conception. According to schema (4.8), the content of the directive or "norm" that Bertie must pinch the cow-creamer (or 'Bertie, pinch the cow-creamer!') is expressed by the deontic sentence **O**(Bertie to pinch the cow-creamer). Instead of calling the content expressed by such a sentence a "norm-lekton" we may call it a deontic proposition. Some philosophers have been reluctant to recognize the possibility that such propositions can have truth-conditions or have confused them with norms, and this has given rise to Jørgensen's problem for deontic logic.

As von Wright notes in the passage quoted above, the distinction between the normative and the descriptive "interpretation" of deontic sentences can be understood as a distinction between two ways of *using* such sentences: they can be used normatively, to create norms, or assertorically to inform the hearer about the content of some system of norms. Jeremy Bentham distinguished between *authoritative* and *unauthoritative* books of "expository jurisprudence": a book is authoritative when it is composed by the legislator, and unauthoritative when it is the work of any other author. [Bentham, 1948, pp. 323-4] [Hedenius, 1941, pp. 65-6] makes a similar distinction between "genuine" and "spurious" legal sentences, and [Kelsen, 1967, p. 355] distinguishes an "authentic" interpretation of law by legal organs from a jurisprudential ("nonauthentic") interpretation: only the former can create law.

The distinction between two ways of using norm sentences can be regarded as a distinction between two kinds of *utterances* of sentences; the "tokens" of a sentence mentioned by von Wright in the passage quoted above are utterances or inscriptions of the sentence. Deontic propositions can be uttered either *performatively*, for creating norms (bringing about an obligation or requirement) or assertorically. [Kamp, 1979, pp. 263-4] In the former case the utterance of the proposition in the appropriate circumstances by a proper norm authority has normative force, and is sufficient to make the deontic proposition in question true, but the truth of an assertoric utterance of the same sentence depends on whether it fits a norm system whose content is determined independently of the utterance in question. The utterer of a deontic proposition can make the intended normative force of the utterance evident by expressing the proposition in the (grammatically) imperative mood or by adding to the utterance the word 'hereby', as in 'You are hereby ordered to pinch the cow-creamer.' Adding the word 'hereby' to the utterance does not change its content. In the case of legal norms and directives, normative utterances include the written inscriptions (occurrences) of norm sentences in authoritative legal texts and documents.

The authoritative (performative) utterances of norm sentences determine the truth-conditions of the deontic propositions which constitute the content of a norm system, and the system derives its normative force from the authoritative utterances of norm sentences which identify the system and tie it to reality. The sense of a deontic proposition can be understood independently of the system to which it belongs, and the same deontic proposition can belong to different systems. Sameness of content is not enough to determine the identity of normative systems; even if α_1 and α_2 contain the same deontic propositions, they are distinct systems if they originate from different normative sources.

The purpose of an "unauthoritative" utterance of a deontic sentence is presumably to convey the content of an existing norm to an audience, and to do this, it must express the same deontic proposition as the original "authoritative" utterance. We may also say that if the former is a replica of the latter, it informs the audience about the normative force of the authoritative utterance. There are no performative utterances of the deontic propositions of general morality, and as was noted earlier, their truth-conditions are determined by the morally relevant objective facts, for example, by the interests of moral subjects.

According to this view, $\mathbf{O}p$ is a complete deontic proposition, and its sense can be grasped independently of the system to which it belongs; thus the same deontic proposition can belong to different systems. The present use of the expression 'deontic proposition' differs from von Wright's and Alchourrón and Bulygin' notion of norm-proposition (norm-statement, a proposition about the existence of a norm). We have to distinguish here the following entities and signs:

(4.12) (i) A norm N (directive, command, imperative).
(ii) '$\mathbf{O}p$': a deontic proposition (norm-lekton, norm-content).
(iii) 'According to a norm system α, $\mathbf{O}p$' or '$\mathbf{O}p$ is part of the content of α'; a normative statement; a proposition which states that a certain norm is part of a norm-system, and conveys the content of the norm.

(For an earlier discussion of (i)-(iii), see [Hansson, 2006].) In the present paper we shall not discuss the nature or existence of norms, except to remark that there is a clear conceptual difference between a norm and its content (a lekton or a deontic proposition). Norms, unlike deontic propositions (norm-lekta), are temporal entities which come into existence by the establishment of a customary rule or (in the case of many legal norms) by acts of promulgation, and they cease to exist by acts of derogation or by being replaced with new customary rules. (Cf. [Alchourrón and Bulygin, 1993, pp. 276-8].) Norms cannot be said to be true or false, but as noted earlier, deontic propositions can have truth-makers, and are capable of being true or false.

If normative and descriptive utterances of a deontic proposition $\mathbf{O}p$ have the same sense, the reference to a specific norm system α and the truth-value of the proposition are determined by the context of utterance. Deontic propositions are true or false relative to a context of normative utterances which determine the identity of the relevant normative system. The reference to a system α can be added to $\mathbf{O}p$; in this way we get norm-statements of the form (12.iii).

The logic of norms can be understood as the logic of norm-statements in the way suggested by Alchourrón and Bulygin as determining the possibility of the existence of norms and norm systems, but the principles of the logic of norms can also be regarded as conditions of the consistency or rationality of norm systems. According to the latter interpretation, deontic logic (the logic of deontic propositions) functions simultaneously as the logic of norms and the logic of norm-statements. The logic of imperatives can be understood in the same way if the semantic content of an imperative is formulated by (4.8), not by (4.4); understood in this way, the logic imperatives is the same as the logic of deontic propositions. This provides a solution to Jørgensen's problem.

[Kamp, 1979, p. 264] has observed that the assertoric use of deontic sentences depends on their performative use. This is true in the sense that the performative utterances of deontic sentences determine the truth-conditions of deontic propositions and their assertoric utterances. Therefore the proposal that logical relations among norms can be understood by studying statements of the form (4.9) or (4.12.iii) puts the cart before the horse. Performative utterances of deontic propositions constitute their own "truth-makers", and they also constitute the truth-makers of assertoric (descriptive) utterances of the same propositions. In their performative use, the function of O- and F-sentences (obligation and prohibition-sentences) representing all things considered norms, is to restrict the range of normatively acceptable options ("the field of permissibility") available to a norm-subject (the addressee), whereas permission sentences have the opposite effect; they enlarge the set of normatively acceptable possibilities. An O-sentence $\mathbf{O}p$ excludes all possibilities in which p does not hold, and a permissive utterance $\mathbf{P}p$ enlarges the set of acceptable options in such a way that they include some possibilities in which p is true. ([Lewis, 1979, p. 166] and [Kamp, 1979, p. 264].) Often it does not matter whether a deontic sentence is used performatively or assertorically, because a true assertoric utterances convey the normative content of a norm or directive to an audience, and can guide the agent's actions in the same way as their performative utterances. For example, in the case of a permission sentence, "either the utterance is a performative and creates a number of new options, or else it is an assertion; but then if it really is appropriate it must be true; and its truth then guarantees that these very same options already exist" [Kamp, 1979, p. 264]. The two kinds of utterances are informationally equivalent.

According to the view that the logic of norms is the same as the logic of deontic propositions, the validity conditions of norms are the truth-makers of deontic propositions. These conditions depend on the kind of directive under consideration. In some cases we may assume that "saying makes it so" [Lewis, 1979, p. 166], and the utterance of an imperative by an authority, (for example, "Bertie, pinch the cow-creamer!" uttered by Stiffy Byng) is enough to make it valid and the corresponding deontic proposition 'Bertie is required to pinch the cow-creamer' true. In the case of legal norms the mere utterance of a deontic proposition does not ensure the validity of the norm (or directive) if the utterer does not have the competence to issue the norm in question. The question about the validity conditions of legal norms is one of the central questions of legal philosophy, but it is not a question for the logic of norms; in the logic of norms the possible validity of normative utterances and norms is presupposed.

5 G. H. von Wright's deontic interpretation of modal logic

In the early 1950's Georg Henrik von Wright [1951] revived the old conception of normative concepts as modal concepts, and presented a system of deontic logic in which the concepts of obligation (ought) and permission (may) were represented by modal operators.

In von Wright's [1951] system the usual deontic operators **P** (for 'permitted'), and **O** (for 'obligatory' or 'ought') are prefixed, not to propositional expressions, but to expressions for action-types or (in von Wright's terminology) "act-names". His syntax contains no propositional letters. In this respect he interprets deontic operators in the same way as the 17th and 18th century authors mentioned above in Section 1. Such expressions can be predicated of individual acts, and may be called act-predicates. Thus the deontic operators of von Wright's system are expressions which turn act-predicates into deontic propositions. This interpretation has certain syntactical consequences. If deontic operators are prefixed to names of generic acts (or act-predicates), the iteration of the operators is ungrammatical: '$\mathbf{P}A$' and '$\mathbf{O}A$' are not act-predicates, and therefore '$\mathbf{OP}A$' and '$\mathbf{OO}A$' are not well-formed formulas. For the same reason "mixed" sentences such as '$A \rightarrow \mathbf{P}A$', in which logical connectives are used to combine deontic and non-deontic components, are not well-formed, since A is (in effect) a predicate, not a proposition. However, mixed sentences and unmodalized propositional expressions are needed, for example, for the representation of conditional norms.

The act-predicates of von Wright's system can be simple (atomic) or complex; and he assumes that complex act-properties are built from atomic predicates by "act-connectives" which are analogous to the classical propositional connectives; thus von Wright speaks of the "negation-act of a given act" and "the conjunction-, disjunction-, implication-, and equivalence-act of two given acts". He assumes that act-predicates have "performance values" analogous to the truth-values of propositions, and the performance-value of a complex act-predicate is determined by the performance-values of its constituents in the same way as the value of a complex propositions is determined by the truth-values of its constituent propositions; for example, a "conjunction-act" $A\&B$ has the value *performed* if and only if A and B

each have the value *performed*.

Von Wright uses the same signs for performance-functions and truth-functions, and adopts the standard interdefinability principles (1.1)-(1.6) (see above) for deontic operators, for example,

(DP0) $\mathbf{O}A \leftrightarrow \neg\mathbf{P}\neg A$,

and two additional principles which he calls the *Principle of Deontic Distribution* and the *Principle of Permission*, using **P** as his deontic primitive:

(DP1) $\mathbf{P}(A \vee B) \leftrightarrow \mathbf{P}A \vee \mathbf{P}B$,

and

(DP2) $\mathbf{P}A \vee \mathbf{P}\neg A$.

Schemata analogous to (DP1) and (DP2) hold for the "ordinary" alethic notion of possibility; thus the P-operator may be said to represent the concept of deontic or normative possibility

Moreover, von Wright adopts the standard inference rules of propositional logic and a modal rule he calls the *Rule of Extensionality*:

(DRE) If A and B are logically equivalent, $\mathbf{P}A$ and $\mathbf{P}B$ are logically equivalent.

According to the customary use of 'intension', logically equivalent expressions have the same intension, not only the same extension; therefore this rule may be called the Rule of Intensionality. In addition, von Wright accepts a third principle which he calls the *Principle of Deontic Contingency*, according to which

(PDC) $\mathbf{O}(A \vee \neg A)$ and $\neg\mathbf{P}(A \& \neg A)$ are not theorems.

If von Wright's system is formulated as an axiomatic system, (PDC) is superfluous, because $\mathbf{O}(A \vee \neg A)$ and $\neg\mathbf{P}(A \& \neg A)$ do not follow from (DP0)-(DP2) by his rules of inference.

Given the restrictions on well-formed formulas in von Wright's system, all well-formed formulas can be written in the form

(5.1) $F(\mathbf{P}A_1, \ldots, \mathbf{P}A_{\mathrm{i}} \ldots, \mathbf{P}A_{\mathrm{n}})$, where F is a truth-function, all occurrences of '**O**' and '**F**' have been replaced respectively by '$\neg\mathbf{P}\neg$ ' and '$\mathbf{P}\neg$', and each A_{i} is a complex act-expression. Let $a_1, \ldots, a_m$ be the simple (atomic) act-predicates in $A_1, \ldots, A_n$, let D_i be the perfect disjunctive normal form of A_i in terms of $a_1, \ldots, a_m$, and let $C^i_{1(i)}, \ldots C^i_{k(i)}$ be the conjunctive constituents of each D_i $(i = 1, ..., n)$. Each C^i_j contains for every atomic act-predicate, either a_i or its negation, but not both, and no other act-predicates. These conjunctions will be called here the C-constituents of A_i. According to the principle of intensionality (DP4), (5.1) is equivalent to

$$(5.2) \quad F(\mathbf{P}D_1, ..., \mathbf{P}D_i, ..., \mathbf{P}D_n),$$

and according to deontic distribution principle (DP1), (5.2) is equivalent to

$$(5.3) \quad F((\mathbf{P}C^1_1 \vee ... \vee \mathbf{P}C^1_{k(1)}), ...(\mathbf{P}C^n_1 \vee ... \vee \mathbf{P}C^n_{k(n)})).$$

The formulas $\mathbf{P}C^i_j$ $(i = 1, 2, ..., n;\ j = 1, 2, ..., k(n))$ are called the P-constituents of (5.1). The P-constituents of a deontic formula are logically independent of each other, except that according to the principle of permission, not all of them can be false. Given m atomic act-predicates, there are 2^m different P-constituents and 2^{2^m}-1 possible truth-value distributions over the P-constituents $\mathbf{P}C^i_j$. Every deontic formula is a truth-function of its P-constituents, and the value of any deontic formula can be determined for each value assignment to P-constituents. This gives a decision method for von Wright's system; a formula is logically true if and only if it is true under every assignment of truth-values to its P-constituents.[8]

The C-constituents C^i_j of deontic formulas are essentially maximally informative descriptions of an agent's possible actions, and as action descriptions they are analogous to Carnap's state-descriptions–complete descriptions of

[8] This allows for the computation of a semantic value for a compound *relative to any given* assignment of values to the compound's P-constituents, but it provides no representation of how the latter deontic formulas are given those assignments. This portion of the story is a bit like an empty box in a chart for a more complete semantic picture. In typical Kripke and post-Kripke style (model theoretical) semantics for deontic formulas (to be discussed in Section 7), the assignments of all deontic formulas (including analogs to C-constituents and P-constituents) is at once determined by structural elements in the models, along with an assignment of truth values to all (non-deontic) atomic formulas relative to points in the structure of the models. Thus the empty box left for von Wright's P-constituents is filled in, or "pushed down" to the lower level assignment of values to atomic descriptive sentences at points, and the relation of points to others in the structures. Of course, all this also presupposes the shift that took place from construing deontic concepts as predicates of act descriptions to construing them as connectives operating on sentences.

a possible state of the world (cf. [Carnap, 1956, pp. 9-10].)

Von Wright's "act-names" are not names in the ordinary sense, but rather general terms which can be predicated of individual acts. It might be argued that as predicative expressions they should contain an empty place which can be filled by an individual expression which refers to a particular act, an act-individual. If von Wright's "act-names" are seriously regarded as expressions which can be used for the purpose of characterizing (or "qualifying") individual acts, and deontic concepts are construed as operators by means of which such generic act-expressions are transformed into (deontic) statements, deontic operators should be syntactically analogous to quantifiers which turn open sentences into complete quantified propositions. According to this view, the infinitive clause

(5.4) Bertie to pinch the cow-creamer

in the deontic proposition

(5.5) Ought(Bertie to pinch the cow-creamer)

should be understood as an act-qualifying expression: it applies to those individual actions which consist in Bertie's pinching the cow-creamer. Thus the expression in the scope of the Ought-operator may be regarded as containing an empty place or a free variable for individual actions. Consequently the sentence (5.5) can be regarded as a complete (or closed) sentence only if the Ought-operator "binds" the free variable in question; this can be made explicit by writing (5.5) as

(5.6) $(\mathbf{O}x)$(Bertie to pinch the cow-creamer(x)),

where 'x' is an act-variable, a variable for individual acts. In this way deontic operators transform predicative expressions into complete sentences in the same way as quantifiers.

Once act-variables are introduced, it is natural to let ordinary quantifiers perform the function of binding variables and turning predicative expressions into complete (closed) sentences. Thus (5.6) should be regarded as an abbreviation of

(5.7) $\mathbf{O}(\exists x)$(Bertie to pinch the cow-creamer(x)),

where 'x' is an act-variable. In (5.7), the expression 'Bertie to pinch the cow-creamer (x)' is an action predicate, and '$(\exists x)$(Bertie to pinch the cow-creamer(x)' is the proposition that Bertie pinches the cow-creamer. The existential quantifier indicates Bertie's performance of the act of pinching. According to (5.7), (5.5) says that Bertie ought to perform an act of pinching the cow-creamer, where the word 'an' functions as an existential quantifier. This is clearly the intended meaning of (5.5).

In deontic logic, this way of treating action sentences and deontic operators was proposed by Jaakko Hintikka ([1957]; see also [Hintikka, 1971, pp. 63-5, pp. 99-101]. The view of action sentences underlying (5.7) has become familiar from Donald Davidson's [1980b] work on action sentences and the logic of action. However, the English legal philosopher John Austin argued already in the early 19th century that legal norms and rules involve quantification over individual actions and not merely over agents [Austin, 1954, pp. 19-24]. According to (5.7), the grammar of deontic operators is similar to that of other modal concepts: they are sentential operators which can be applied to (action) propositions to make deontic propositions.

(5.7) represents a "positive" obligation (or ought), and the schema

(5.8) $\mathbf{O}(\exists x)A(x)$

can be regarded as a general form of such an obligation. According to the standard interdefinability principles of deontic operators, (1.1)-(1.6), actions of type A are prohibited if and only if

(5.9) $$\begin{aligned}\mathbf{F}\exists x A(x) &\leftrightarrow \mathbf{O}\neg\exists x A(x)\\ &\leftrightarrow \mathbf{O}\forall x \neg A(x)\\ &\leftrightarrow \neg\mathbf{P}\exists x A(x),\end{aligned}$$

Actions of type A are prohibited if and only if it is obligatory that every action to be performed by the agent not be of type A, in other words, and in less contrived language, it is not permitted to perform any action of that kind. Here the word 'any' has a narrow scope, and should be translated as an existential quantifier. An action A is permitted if and only if it is permitted to perform such an action:

(5.10) $\mathbf{P}\exists x A(x)$.

In cases of simple obligation, permission and prohibition, the deontic operator has wide scope, for example, it is clear that the proposition that an agent ought to do A does not mean that some particular act ought to be an instance of A. This does not mean that quantifying in is never intelligible for act-variables in deontic contexts; for example, a promise or a contract creates an obligation to fulfill the promise or satisfy the contract, and the general obligation to keep one's promises seems to have the form

(5.11) $\forall x(Cx \rightarrow \mathbf{O}(\exists y S(y, x))$,

where 'Cx' is the predicate of making a promise, and '$S(y, x)$' means that the (individual) act y satisfies the promise x. (5.11) signifies a claim allowing for detachable, "absolute" obligation. [Hintikka, 1971, p. 100] has suggested that the wide-scope ought-proposition

(5.12) $\mathbf{O}\forall x(Cx \rightarrow (\exists y S(y, x)))$,

can serve as a representation of a prima facie commitment. (For some of difficulties related to the quantification into deontic contexts with act-variables, see [McArthur, 1981].)

The fact that deontic operators are attached to complete propositions rather than generic action terms does not mean that the logic in question is a theory of the ought-to-be rather than a theory of the ought-to-do. If the propositions and predicates in the scope of deontic operators are action propositions and predicates, the logic can also be regarded as a logic of the ought-to-do. Deontic sentences can be read in both ways.

6 The standard system of deontic logic (SDL) and close cousins

6.1 SDL

Two alterations in von Wright initial approach in 1951 easily lead to what soon came to be known as *Standard Deontic Logic*, and thus what became the dominant approach to deontic logic soon after his [von Wright, 1951]. First, if deontic concepts are represented by propositional operators, the limitations on well-formed formulas in von Wright's [1951] system can be dropped, and both mixed formulas (e.g., $p \rightarrow \mathbf{O}p$) and formulas with iterated operators (e.g., $\mathbf{O}(\mathbf{O}p \rightarrow p)$) can be accepted as well-formed. This brings the interpretation of deontic logic closer to other interpretations of modal logics, for example, the alethic and the epistemic interpretations. Such an approach offers the enormous practical convenience of just adding a new modal operator layer on top of well-known systems of propositional logic (e.g., classical truth-functional systems).[9] Von Wright adopted this

[9] The first edition of [Prior, 1955] appears to have initiated this shift after [von Wright, 1951]; of course Mally had used propositional variables two decades earlier. See Section 3.

approach himself at times.[10] However, there was also von Wright's principle of deontic contingency (PDC), whereby a logically impossible act–an act-type identified by a contradictory act-description ($A \& \neg A$), is not necessarily prohibited, and an act-type identified by a tautologous act-description ($A \vee \neg A$) is not necessarily obligatory.[11] However, in von Wright's [1951] system, any act B is permitted if $A \& \neg A$ is permitted, and no act B is obligatory unless a tautologous act $A \vee \neg A$ is obligatory. Thus accepting $\mathbf{O}(A \vee \neg A)$ and $\neg\mathbf{P}(A \& \neg A)$ as logically true excludes only empty normative systems, systems according to which everything is permitted and nothing is obligatory. Many felt this was too small a difference to matter. Moreover, the principle (PDC) is inconsistent (for example) with Leibniz's analysis of an obligatory act as an act which is necessary for a good person to perform, because any act a good person, or indeed any person, can perform, satisfies a tautologous act-description $A \vee \neg A$ of the sort von Wright embraces. In the same way, it is not possible for a good person, or any person, to perform an impossible act; thus such acts should be regarded as not permitted from the standpoint of Leibniz's analysis of obligation and permission. For the sake of theoretical simplicity, as well as continuity with classical propositional logic and modal logic, the deontic necessitation rule, applied to propositions, was routinely included in deontic systems:

(RND) If p is a theorem, $\mathbf{O}p$ is a theorem

These revisions of von Wright's [1951] system transform it into what is usually called "the standard system of deontic logic", abbreviated 'SDL' ([Føllesdal and Hilpinen, 1971, p. 13] and [Hansson, 1971, p. 122]). The propositional SDL is defined by adding to non-modal propositional logic the modal axiom schemata

(KD) $\mathbf{O}(p \to q) \to (\mathbf{O}p \to \mathbf{O}q)$

and

(DD) $\mathbf{O}p \to \neg\mathbf{O}\neg p$

and the rule of deontic necessitation (RND). Here the letters p and q can be regarded as representing arbitrary formulas. On the basis of the axioms (KD) and (DD), this system may be called the system KD (or simply D). It is a member of the family of *normal* modal logics, all of which contain

[10] For example in his key early revisions of his "old system" in [von Wright, 1964; von Wright, 1965]; see also [von Wright, 1968].

[11] Although it is a theorem of his system that all tautologous act-types are *permissible*. $\mathbf{P}(A \vee \neg A)$, or equivalently that no contradictory one is obligatory, $\neg\mathbf{O}(A \& \neg A)$.

(a counterpart of) the rule RND [Chellas, 1980, p. 114]. With its origins in the 14th century, the "Traditional Definitional Scheme" is routinely taken for granted in formulations of SDL:

(TDS) $\mathbf{P}p =_{\text{df}} \neg\mathbf{O}\neg p$ and $\mathbf{F}p =_{\text{df}} \mathbf{O}\neg p$[12]

The theorems of the system include the formulas:

(6.1) $\mathbf{O}(p \,\&\, q) \rightarrow (\mathbf{O}p \,\&\, \mathbf{O}q)$ (*Conjunctive Distributivity of* **O**)

(6.2) $\mathbf{O}p \,\&\, \mathbf{O}q \rightarrow \mathbf{O}(p \,\&\, q)$ (*Aggregation for* **O**)

(6.3) $\mathbf{O}p \rightarrow \mathbf{O}(p \vee q)$ (*Weakening*)

(6.4) $\mathbf{O}(p \rightarrow q) \rightarrow (\mathbf{P}p \rightarrow \mathbf{P}q)$

(6.5) $\mathbf{P}p \rightarrow \mathbf{P}(p \vee q)$

(6.6) $\mathbf{P}(p \vee q) \rightarrow (\mathbf{P}p \vee \mathbf{P}q)$ (*Disjunctive Distributivity of* **P**)

(6.7) $\mathbf{P}(p\&q) \rightarrow \mathbf{P}p$

(6.8) $\mathbf{O}\top$ (ON)

(6.9) $\neg\mathbf{O}\bot$ (OD)

(6.10) $\mathbf{O}p \rightarrow \mathbf{P}p$ (DD')

(6.11) $(\mathbf{O}p \,\&\, \mathbf{P}q) \rightarrow \mathbf{P}(p \,\&\, q)$

(6.12) $\mathbf{O}p \vee (\mathbf{P}p\&\mathbf{P}\neg p) \vee \mathbf{O}\neg p$ (*Exhaustion*)

(6.13) $\neg(\mathbf{O}p \,\&\, (\mathbf{P}p \,\&\, \mathbf{P}\neg p))\&\neg(\mathbf{O}\neg p \,\&\, (\mathbf{P}p \,\&\, \mathbf{P}\neg p)) \,\&\, \neg(\mathbf{O}p \,\&\, \mathbf{O}\neg p)$[13]

Two important derivable rules of inference are

(RMD) If $p \rightarrow q$ is a theorem, then $\mathbf{O}p \rightarrow \mathbf{O}q$ is a theorem,

sometimes called the "Inheritance Principle", as well as "RM", and

(RED) If $p \leftrightarrow q$ is a theorem, then $\mathbf{O}p \leftrightarrow \mathbf{O}q$ is a theorem,

a deontic "equivalence rule".

[12] Letting **O**B, **P**E, **I**M, **O**P, **O**M stand respectively for it is *obligatory* that, *permissible* that, *impermissible* that, *optional* that, and *omissible* that, the more extended scheme would be: $\mathbf{P}\mathrm{E}p =_{\text{df}} \neg\mathbf{O}\mathrm{B}\neg p$ and $\mathbf{I}\mathrm{M}p =_{\text{df}} \mathbf{O}\mathrm{B}\neg p$, $\mathbf{O}\mathrm{P}p =_{\text{df}} (\neg\mathbf{O}\mathrm{B}p\&\neg\mathbf{O}\mathrm{B}\neg p)$, and $\mathbf{O}\mathrm{M}p =_{\text{df}} \neg\mathbf{O}\mathrm{B}p$.

[13] (6.13) expresses the exclusiveness of the three classes that (6.12) says are exhaustive. The conjunction of (6.12) and (6.13) expresses the "Traditional Threefold Classification" asserting that every alternative is either obligatory, optional or impermissible, and no more than one of these. Using $\underline{\vee}$ for the *exclusive or*, this can be succinctly expressed as: $\mathbf{O}\mathrm{B}p \,\underline{\vee}\, \mathbf{I}\mathrm{M}p \,\underline{\vee}\, \mathbf{O}\mathrm{P}p$. See [McNamara, 1990; McNamara, 1996a] on the significance of this feature of the traditional scheme.

As expected, if we recast von Wright's 1951 system, now construing "**P**' and "**O**" as propositional operators (not predicates of act-types), and the variables as propositional variables (not variables for act-types), the key principles and rules mentioned earlier (DP0-DP2 and DRE) are all easily derivable in the SDL system above.

It should be noted that some principles not derivable in SDL were often deemed truths of deontic logic, especially if the "**O**" is interpreted as "It ought to be that". Perhaps the most salient example of this kind is the principle: $\mathbf{O}(\mathbf{O}p \to p)$. This can be construed as saying (roughly) that it ought to be the case that whatever ought to be is.[14]

It is important to note here that we use throughout "Standard Deontic Logic" and its abbreviation, "SDL", as proper names, not as descriptions. Many think that these are misnomers, not quite as bad as the "Holy Roman Empire", which fails on all three counts, but surely not quite the "standard" the label might suggest. It is indeed probably fair to say that most researchers think there is at least some thesis of SDL that on some prominent interpretation of "**O**" is not a logical truth at all, and furthermore there are a number of somewhat independent such complaints about SDL. So it is hardly a widely popular system of logic with only occasional outliers rejecting it as the title might suggest. Rather, it is the most widely known, well-studied system, and central in the accelerated historical development of the subject over the last 50 or so years. As such, it serves as a historical comparator, where various important developments in the subject were explicit reactions to its perceived shortcomings, and even when not, sometimes can be fruitfully framed as such.

6.2 The Leibnizian-Kangerian-Andersonian reduction

It is easy to see that Leibniz's definition of the concept of obligation (or ought),

(O.Leibniz) p is obligatory for S if and only p is necessary for S's being a good person,

can be seen as supporting the principles of SDL when conjoined with plausibly intended assumptions about the possible instantiation of goodness and about the notion of necessity involved. If the explicit reference to the agent is suppressed, (O.Leibniz) can be expressed in the form

(O.GWL) $\mathbf{O}p \leftrightarrow \mathbf{N}(g \to p)$,

where '**N**' is an alethic necessity operator and 'g' represents a proposition

[14] See [Kanger, 1957; Hintikka, 1971].

which expresses the agent's being a good person or, in the case of the ought-to-be, the goodness of the world.

The Leibnizian concept of permission (the concept of may) is defined by

(P.GWL) $\mathbf{P}p \leftrightarrow \mathbf{M}(g \,\&\, p)$.

These schemata can be regarded as partial reductions of deontic logic to alethic modal logic.

It should be observed that Leibniz's definition of 'obligation' in terms of 'good' is prima facia not subject to the same difficulty as one main definition of normative concepts in terms of a comparative concept of goodness as applying to any state of affairs that is better than its negation, as with (D.OÅqv).[15] Thus (O.GWL) leaves room for supererogatory actions.[16] If deontic logic is regarded as a theory about the ought-to-be rather than ought-to-do, the Leibnizian interpretation of $\mathbf{O}p$ may be expressed (for example) as 'for things to be best, it is necessary that p' or 'for things to be apt, it is necessary that p'.

If it is assumed that it is possible to be good or that the requirements of morality can be satisfied (if it is possible for things to be in order), that is,

(M.g) $\mathbf{M}g$,

the **O**-operator defined by (O.GWL) satisfies all the principles of SDL, provided that the **N**-operator satisfies the axioms of the modal system called T in [Chellas, 1980, p. 131], viz.

(K) $\mathbf{N}(p \to q) \to (\mathbf{N}p \to \mathbf{N}q)$

and

(T) $\mathbf{N}p \to p$

and the modal "rule of necessitation", viz.

(RN) If p is a theorem, $\mathbf{N}p$ is a theorem

[15] There are complexities here. If we interpret "**O**" as "ought", then (D.OÅqv) is more plausible, since it is plausible that "ought" is some sort of optimizing notion, unlike "must" or "obligatory". So if p is better than not p, then it plausibly does follow that it ought to be that p, even though it need not be a must that p or obligatory that p. The Leibnizian reading of **O** is also better for "must" or "obligatory". Notice however that if we assume Leibniz means a "perfectly good" man, we end up again with an optimizing notion, where it is plausible to now see "**O**" as ought" not "obligatory", and as once again in tension with supererogation. This is sometimes a problem in virtue ethical attempts to analyze permissibility and obligation while allowing for supererogation.

[16] See [McNamara, 1999] for one development along this line.

then (O.GWL) guarantees the validity of all principles of SDL.[17] Axiom (M.G) is needed for proving the consistency principle (DD).

In the 20th century deontic logic, the Leibnizian analysis of the concepts of obligation and permission was rediscovered by the Swedish philosopher Stig Kanger [1957, pp. 53-4]. Kanger interpreted the constant g as "what morality prescribes". According to this interpretation, $\mathbf{O}p$ (it is obligatory that p) means that p follows from the requirements of morality. Alan Ross Anderson [1956] put forward a reduction schema equivalent to Kanger's[18],

(O.S) $\quad \mathbf{O}p \leftrightarrow \mathbf{N}(\neg p \rightarrow S)$,

where S may be taken to mean the threat of a sanction or simply the proposition that the requirements of law or morality have been violated. If p is an action proposition and the negation sign is understood as the omission of the action expressed by p, (O.S) is equivalent to Meinong's schema (2.8).

It should be noted that in order to validate all of SDL in this "reduction", the principle T above is overkill, the normal modal system K with just an axiom saying G is possible is sufficient to generate all of SDL, and if we do add axiom T, this results in a stronger deontic fragment, namely SDL with the addition of $\mathbf{O}(\mathbf{O}p \rightarrow p)$, which we noted above is not derivable in SDL. However, it is also surely correct that the intended reductions of Leibniz, Anderson, and Kanger are ones where the notion of necessity involved is *alethic* necessity, and so ones for which the T thesis above holds. Thus from the standpoint of the Leibnizian, Kangerian, and Andersonian reductions, SDL is *too weak* a system for deontic logic, and needs to be augmented with the addition of $\mathbf{O}(\mathbf{O}p \rightarrow p)$. As noted above, others also thought this additional might be needed for independent reasons. Lastly we should note that the resulting "reduction" also allows for mixed modal-deontic formulas, and formulas involving deontic operators and the deontic constant, g. We note a few salient ones that are thesis:[19]

(6.14) $\quad \mathbf{O}g$

[17] See [Åqvist, 2002; Åqvist, 1987] for proofs of correspondences between SDL and its extensions and normal modal systems employing the Leibnizian-Kangerian-Andersonian style reduction.

[18] The arrow is interpreted as material implication here. Anderson also explored alternative interpretations using non-truth functional conditionals, such as relevant logic conditionals.

[19] Note that if the reduction is treated as offering an analysis of obligation (6.14) can have a meta-obligatory flavor, saying something like it is obligatory that all one's obligations are met; if we read g *à la* Leibniz, it says it is obligatory that what a good man would do is done.

(6.15) $\Box(p \to q) \to (\mathbf{O}p \to \mathbf{O}q)$

(6.16) $\Box p \to \mathbf{O}p$ (RND′)

(6.17) $\mathbf{O}p \to \Diamond p$ (A weak version of "Kant's Law")[20]

(6.18) $\neg\Diamond(\mathbf{O}p \& \mathbf{O}\neg p)$ (3DD′)

The supplement to this section provides some more formal details.[21]

Supplement to Section 6: some formalities

A6.1 The SDL wffs (well-formed formulas)

PV is a set of sentence letters $P_1, ..., P_i, ...$ – where "i" is 1,2,. . . . There are three primitive propositional (sentential) operators: ¬, →, **O**B (for it is **ob**ligatory that); and a pair of parentheses: (,).

Let the set of SDL-wffs be the smallest set such that (lower case "p" and "q" are metavariables ranging over formulas):

1. PV is a subset of the SDL-wff
2. For any p, p is among the SDL-wffs only if $\neg p$ and $\mathbf{O}\mathrm{B}p$ are as well.
3. For any p and q, p and q are in SDL-wffs only if $(p \to q)$ is in SDL-wffs.

We shift to two letter abbreviations here and in a few other select places to make it easier to express more operators (as well as for mixing these with other operators later on): **O**B, **P**E, **I**M, **O**M, **O**P for it is *obligatory* that, *permissible* that, *impermissible* that, *omissible* that, and *optional* that, respectively. It will be clear when this shift is in play, so no ambiguity or confusion should result.

We use the following abbreviations for formulas and subformulas of the SDL-wffs:

(DF 1) $\&, \vee, \leftrightarrow$ as usual.

[20] Kant's law is more accurately rendered as involving agential possibility (agential ability), not merely impersonal possibility, although it is possible to move closer to this by reading the modal operators in the reduction as keyed to what is predetermined and possible for a given agent. [Hansson, 1969] showed how this sort of relativization can be easily done explicitly for modal logics, though authors often leave it implicit.

[21] We will specify one other logic (VW) in tandem with its semantics, in Section 7, and in the next supplement.

(DF 2) $\mathbf{P}\mathrm{E}p =_{\mathrm{df}} \neg\mathbf{O}\mathrm{B}\neg p$ ("it is permissible that").

(DF 3) $\mathbf{I}\mathrm{M}p =_{\mathrm{df}} \mathbf{O}\mathrm{B}\neg p$ ("it is forbidden/impermissible that")

(DF 4) $\mathbf{O}\mathrm{M}p =_{\mathrm{df}} \neg\mathbf{O}\mathrm{B}p$ ("it is omissible/gratuitous that"

(DF 5) $\mathbf{O}\mathrm{P}p =_{\mathrm{df}} (\neg\mathbf{O}\mathrm{B}p \,\&\, \neg\mathbf{O}\mathrm{B}\neg p)$ ("it is optional that")

A6.2 SDL and one extension

We assume the language is that specified in 6.1. We provide this standard axiomatization of SDL, where $\vdash$ before a formula indicates it is a thesis (axiom or theorem) of the relevant system:

(A1) All tautologies of the language (TAUT)

(A2) $\mathbf{O}\mathrm{B}(p \to q) \to (\mathbf{O}\mathrm{B}p \to \mathbf{O}\mathrm{B}q)$ (KD)

(A3) $\mathbf{O}\mathrm{B}p \to \neg\mathbf{O}\mathrm{B}\neg p$ (DD)

(MP) If $\vdash p$ and $\vdash p \to q$ then $\vdash q$

(RND) If $\vdash p$ then $\vdash \mathbf{O}\mathrm{B}p$

This is essentially just the normal modal logic D, with a notational variant to indicate the deontic interpretation.

Let's also introduce one extension of SDL, called here contextually, SDL^+. SDL^+ is the system that results from adding just A4 to SDL:

(A4) $\mathbf{O}\mathrm{B}(\mathbf{O}\mathrm{B}p \to p)$

A6.3 Two Leibnizian-Kangerian-Andersonian systems

PV is a set of sentence variables $P_1, ..., P_i$, ... – where "i" is 1,2,... There are three primitive propositional (sentential) operators: $\neg$, $\to$, $\Box$. There is a distinguished propositional constant g, and a pair of parentheses, (and).

One can read g a variety of ways (e.g. as "all normative demands are met"), but we use "g" to honor Leibniz who essentially is the first to analyze the basic operators in "reductive terms", essentially a kind of virtue ethical reduction: in terms of what it is necessary for a *good* person to do. We can think of g this way as expressing the proposition that *what a good man would do is done*.

Let the set of LKA-wffs be the smallest set such that:

1. g is among the LKA-wffs.
2. PV is a subset of the LKA-wff, and
3. For any p, p is among the LKA-wffs only if $\neg p$ and $\Box p$ are among the LKA-wffs.

4. For any p and q, p and q are in LKA-wffs only if $(p \to q)$ is in LKA-wffs.

We use the following abbreviations for formulas and subformulas of the LKA-wffs:

(DF 1') $\&, \vee, \leftrightarrow$ as usual.

(DF 2') $\Box p =_{\text{df}} \neg\Diamond\neg$ ("it is possible that").

(DF 3') $\mathbf{O}\mathrm{B}p =_{\text{df}} \Box(g \to p)$

(DF 4') $\mathbf{P}\mathrm{E}p =_{\text{df}} \Diamond(g\&p)$

(DF 5') $\mathbf{I}\mathrm{M}p =_{\text{df}} \Box(p \to \neg g)$

(DF 6') $\mathbf{O}\mathrm{M}p =_{\text{df}} \Diamond(g\&\neg p)$

(DF 7') $\mathbf{O}\mathrm{P}p =_{\text{df}} (\Diamond(g\&p)\,\&\,\Diamond(g\&\neg p))$

The logic LKA_1 (essentially the Normal Modal Logic K with (A3') added):

(A1') All tautologies of the language (TAUT)

(A2') $\Box(p \to q) \to (\Box p \to \Box q)$ (K)

(A3') $\Diamond g$ $(\Diamond g)$

(R1') If $\vdash p$ and $\vdash p \to q$ then $\vdash q$ (MP)

(R2') If $\vdash p$ then $\vdash \mathbf{O}\mathrm{B}p$ (RN)

Metatheorem 6.1 *(Loosely stated) SDL is the strongest pure deontic fragment contained in LKA_1.*

The Logic LKA_2 (essentially the normal modal logic T with A3 added). LKA_2 is the system that results from adding just (A4') to LKA_1:

(A4') $\Box p \to p$ (T)

This logic just adds the characteristic T axiom to $\mathrm{LKA}_{\mathbf{1}}$, and Leibniz, Kanger, and Anderson all had alethic necessity in mind. The only reason we give the prior K version, is that (LKA_2) generates a pure deontic logic stronger than SDL, namely SDL^+.

Metatheorem 6.2 *(Loosely stated) SDL^+ is the strongest pure deontic fragment contained in LKA_2.*

See [Åqvist, 2002; Åqvist, 1987] for the exact statements of the metatheorems, as well as their proofs, which are semantic in nature.

(See also [Parent, 2008].) [Lokhorst, 2006] explores the correlation between quantified propositional variable systems reading g as "all normative demands/obligations are met", especially in the context of Anderson's original reduction using relevance logic. [McNamara, 1999] extends this approach to cover action beyond the call of duty and other concepts. The ideas of the reduction have been employed and modified in a variety of contexts, as a perusal of the DEON conference volumes indicates.

7 The semantics of Standard Deontic Logic and close cousins

7.1 The semantics for SDL

The sentences of SDL can be interpreted in terms of possible situations or world states ("possible worlds") in the same way as other normal modalities. A possible worlds model of SDL is a triple $M = \langle W, R, I \rangle$, where W is a universe of possible situations, also called points of the model, R is a binary relation on W, and I is an interpretation function which assigns to each sentence letter of the modal language a subset of W, that is, the points $u \in W$ at which the sentence letter is to be deemed true. The truth of any formula p at u under M is then expressed briefly as '$M, u \models p$', or even more briefly, where the model is left tacit, '$u \models p$', and defined recursively. If p is not true at u, it is false at u ($u \not\models p$). A sentence is called *valid* (logically true) if and only if it is true at every situation $u \in W$ for every model M, and q is a logical consequence of p if and only if there is no model M and world u such that $M, u \models p$ and not $M, u \not\models q$. Truth at a world in a model, $\models$, is defined in accord with the usual Boolean conditions which ensure that the truth-functional compounds of simple sentences receive appropriate truth-values at each possible world. The alternativeness relation R is needed for the interpretation of sentences involving the deontic operators. In the semantics of modal logic, necessary truth at a given world u is understood as truth at all points which are possible relative to u or are *alternatives* to u, and possible truth at u means truth at some alternative to u. For the concepts of deontic necessity or obligation (sometimes read as "ought") and deontic possibility or permission (may), these conditions can be formulated as follows:

(CO) $u \models \mathbf{O}p$ if and only if $v \models p$ for every $v \in W$ such Ruv,

and

(CP) $u \models \mathbf{P}p$ if and only if $v \models p$ for some $v \in W$ such Ruv.

To ensure the validity of axiom DD, it is necessary to regard R as a serial relation, in other words,

(CD) For every $u \in W, Ruv$ for some $v \in W$.

Further assumptions about the structural properties of the R-relation validate different deontic principles, and lead to different systems of deontic logic. For example, it is clear that

(7.1) $\mathbf{O}p \rightarrow p$

is not a logical truth as interpreted, and therefore R cannot be assumed to be a reflexive relation, but reading "**O**" as it ought to be that, the principle

(7.2) $\mathbf{O}(\mathbf{O}p \rightarrow p)$

seems a valid principle: It ought to be the case that whatever ought to be the case is the case. The validity of (7.2) follows from the assumption that R is secondarily reflexive, in other words,

(COO) If Ruv for some $v,$ then Rvv.

The semantics sketched above may be termed the "standard semantics" of deontic logic. [Hintikka, 1957; Hintikka, 1971; Kanger, 1957] were among the first philosophers who used an alternativeness relation between possible worlds or situations to formulate the truth-conditions of deontic sentences. It is also more generally referred to as "Kripke semantics" or "Kripke style semantics" in modal logic [Kanger, 1963].[22]

Recalling our earlier discussion of Meinong's scheme and his use of axiological preference ordering concepts, let us briefly note here that a more axiological background semantic picture for SDL, one having affinities with utilitarianism, was often endorsed. Suppose we have a set of relations, one for each world u in W, where $v \geq_u w$ is thought of as indicating that v is ranked at least as high as w relative to u. Suppose further that we assume that for each world, u in W:

1. $v \geq_u v$ (reflexivity),
2. if $v \geq_u w$ and $w \geq_u l$ then $v \geq_u l$ (transitivity),
3. either $v \geq_u w$ or $w \geq_u v$ (connectivity).[23]

[22] An excellent source on the history of the emergence of formal semantic frameworks for deontic logic is [Wolenski, 1990].

[23] So each $\geq_u$ is a total pre-ordering of W (sometimes called a complete pre-ordering or total quasi-ordering).

Now suppose we add something else people often assumed, the "Limit Assumption":

(LA) For each u, there is v such that for any $w, v \geq_u w$

That is, relative to any world u, there is always a world ranked at least as high relative to u as any worlds (i.e. there is at least one u-best world). Lastly, we now use this framework to provide a truth clause for **O** via bests:

(COB) $\mathbf{O}p$ is true at u iff p holds in all the u-best worlds

It was widely recognized that this approach will also determine SDL, but the metatheory of SDL and related systems via generalized ordering semantics has not been very widely explored compared to Kripke-style semantics.[24] Essentially, this ordering framework provides a way to *generate* the set of u-acceptable worlds out of the ordering:

v is u-acceptable iff v is a u-best world

So we get a u-acceptability relation for each world, just as is presupposed in the standard Kripke semantic structures. We need only look at what propositions hold at the u-best worlds to interpret the truth-conditions of SDL's deontic operators exactly as we did with the simpler Kripke relational structures. If the reader wonders about how seriality in the Kripke structures is captured, it is guaranteed by the Limit Assumption. For that assumption entails that for each world u, there is a u-best world. As such, the ordering semantics is overkill for SDL and most of its resources go unutilized for SDL; but as we will see a bit later, when people started thinking about how to generalize or adjust SDL to handle more complex deontic concepts, these sorts of ordering structures (and their generalizations) became quite important, as was the recognition that SDL itself could be easily subsumed under such ordering frameworks.

7.2 Semantics for the Leibnizian-Kangerian-Andersonian reduction

We now use this semantic approach to turn back to the Leibnizian-Kangerian-Andersonian reduction. Assume we have a classical propositional language with a distinguished propositional constant, g, and the modal operator, $\Box$, intended as expressing alethic necessity (with $\Diamond$ defined as $\neg\Box\neg$). Then as with the semantics above, $\langle W, R, I\rangle$ will be a model, with W interpreted as a set of points or worlds, R a binary relation on W, and I an interpretation

[24] But see [Goble, 2003; Åqvist, 1987; Spohn, 1975; Jennings, 1974] and in a slightly different setting, [Lewis, 1973; Lewis, 1974].

assigning a subset of W to each sentence letter. Truth in a model at a world is defined just as we did above, where for the necessity operator, we have:

(C$\Box$) $u \models \Box p$ if and only if $v \models p$ for every $v \in W$ such Ruv

For the moment, we place no structural constraint on R. How do we interpret g, the only element in the reduction that has a deontic or valuative flavor? As a propositional constant, let's read it as *what a good man would do is done*. This is close to Leibniz. With this in mind, we then interpret g by having I assign it a subset of W, with the intention that these will be the worlds where what a good man would do is done. We thus add one more element to the models, $\langle W, R, G, I\rangle$, and add the constraint:

(CG) $G \subseteq W$.

Then g will be true at a world u in a model if and if u is a G-world:

(Cg) $u \models g$ if and only if $u \in G$

We need one structural constraint to validate the axiom $\Diamond g$, to the effect that it is always possible in the models that what the good man would do is done. At the semantic level this amounts to adding an analog to seriality used for the semantics for SDL, namely a constraint to the effect that for every world u, u has accessible to it a world where what the good man would do is done:

$$\text{For every } u \in W, \text{ there is a } v, \text{ such that } v \in G \text{ and } Ruv.$$

With the definitions identified in the last section, this semantical system will validate all the pure deontic principle of SDL, along with other mixed principles such as $\mathbf{O}g$ and $\mathbf{O}p \to \Diamond p$. However, given the intended interpretation of the $\Box$ as expressing *alethic* necessity, and thus as supporting the T axiom,

(T) $\Box p \to p$

we would need to add reflexivity,

(Rflx) For every u $\in W, Ruu$.

thus generating a pure deontic fragment stronger than SDL$^+$, but that is clearly a consequence of the intended interpretation of the reduction.

7.3 A generalization of SDL: VW logics and their non-standard semantics

We now sketch a slight generalization of the standard framework for SDL, one that allows us to include a weakening of SDL that accords with one key aspect of von Wright's earliest work. Recall that in opening this section, we noted that von Wright interpreted "**O**" and "**P**" as act predicates, and that he also endorsed "Deontic Contingency", thereby rejecting $\mathbf{O}\top$ and $\neg\mathbf{P}\bot$ and their equivalents as logical truths, The former act predicates issue seems separately motivated, and von Wright himself later flirted with treating "**O**" and "**P**" as propositional operators. These facts raise the interesting question: What might a propositional deontic framework look like which treats "**O**" and "**P**" as propositional operators and is as close to SDL as is consistent with *Deontic Contingency*? It turns out that there is a simple such syntactic and semantic framework, one that is a conservative generalization of that for SDL.[25] The language is that of SDL. In honor of von Wright, let's call the base logic VW:

(A1)	All tautologies of the language	(TAUT)
(A2)	$\mathbf{O}(p \to q) \to (\mathbf{O}p \to \mathbf{O}q)$	(KD)
(A3)	$\neg\mathbf{O}\bot$	(OD)
(MP)	If $\vdash p$ and $\vdash p \to q$ then $\vdash q$	
(RM)	If $\vdash p \to q$ then $\vdash \mathbf{O}p \to \mathbf{O}q$	(RMD)

Although $\mathbf{O}p \to \neg\mathbf{O}\neg p$ is easily derivable from VW, neither $\mathbf{O}\top$ nor $\neg\mathbf{P}\bot$ is derivable, but VW is easily derivable from SDL. If we add $\mathbf{O}\top$ to VW, we get a system equipollent to SDL. But how will the semantics for **O** work? It can't be standard, else $\mathbf{O}\top$ would be validated, and thus it would need to be a theorem for the logic to match the semantics.

The basic idea is simple. The model structures are those for normal modal logics like K and SDL above, with no structural constraints on the accessibility relation. The key difference at the semantic level is that the clause for **O** is non-standard:

(CO') $u \models \mathbf{O}p$ iff there is a v such that Ruv and for every v if Ruv, then $v \models p$

Thus something is obligatory at u iff there is a u-acceptable world to begin with, and all such worlds are p-worlds. Plainly, if there is no u-acceptable

[25] This basic orientation appeared in [McNamara, 1990], but the elementary metatheory was done in [McNamara, 1988]. Max Cresswell pointed out in conversation that there are affinities to the [Kripke, 1965] treatment of some non-normal modal logics using "non-normal worlds". See also [Cresswell, 1967].

world, $u \models \mathbf{O}p$ for all p (including $\top$), so $\mathbf{O}\top$ is *not* valid. With $\mathbf{P}p$ defined as $\neg\mathbf{O}\neg p$, we get this non-standard clause for $\mathbf{P}$:

(CP') $u \models \mathbf{P}p$ iff either there is no v such that Ruv or $v \models p$ for some v such that Ruv.[26]

But then if there is no v such that Ruv, $u \models \mathbf{P}p$, for any p (including $\bot$), and so $\neg\mathbf{P}\bot$ is *not* valid.

Notice however that $\mathbf{O}p \to \neg\mathbf{O}\neg p$, which is derivable in VW, is also valid in all models per the clause above. For suppose the antecedent is true at u. Then there is a p-world that is a u-alternative and all u-alternatives are p-worlds, but if so, $\mathbf{O}\neg p$ must be false at u for otherwise there would have to be at least one u-alternative that was both a $\neg p$-world and a p-world.

What then is the role of seriality if $\mathbf{O}p \to \neg\mathbf{O}\neg p$ is already validated without any constraints? Given the non-standard clause for $\mathbf{O}$, if we add seriality as a constraint, "$\mathbf{O}\top$" is then validated, SDL is determined, and all is "back to normal". Adding seriality essentially assures that the first clause in the non-standard truth definition of "$\mathbf{O}$" above is automatically met, and so the conjunctive clause is then equivalent to the standard one (the right conjunct above). Similarly, for the clause for $\mathbf{P}$: the first clause is excluded by seriality, so the disjunctive clause is equivalent to the familiar one (the right disjunct above). Thus the framework is a conservative generalization of the standard one for SDL, but one where von Wright's contingency intuitions can be modeled, and so nothing need be guaranteed obligatory or impermissible in the base logic, since there need not be, as it were, any normative standard at all, though in keeping with von Wright's intuitions, if anything is obligatory at all, then so too will $\top$ be; and semantically, that anything at all is obligatory at u amounts to saying u has some standard, namely a non-empty set of u-acceptable ways things might be. It also seems to us more fitting to frame things this way given the place of von Wright's work in stimulating the emergence of SDL, and the fact that his principle of contingency is really separate from his initial conception of "$\mathbf{O}$" and "$\mathbf{P}$" as action. We also note in passing that had propositional deontic logic originally been conceived this way, the base deontic logic playing the role SDL now plays would not have been a normal modal logic (although obviously a close cousin).

[26]Similarly for the remaining three operators defined above: $u \models \mathbf{I}\mathrm{M}p$ iff there is a v such that Ruv and for every v, if Ruv then $v \models \neg p$; $u \models \mathbf{O}\mathrm{M}p$ iff either there is no v such that Ruv or $v \models \neg p$ for some v such that Ruv; $u \models \mathbf{O}\mathrm{P}p$ iff either there is no v such that Ruv or both $v \models \neg p$ for some v such that Ruv and $v \models p$ for some v such that Ruv.

7.4 Classical quantification and SDL semantics

Most presentations of deontic logic are restricted to propositional logic. This is a serious and unnecessary limitation; as was observed earlier (in Section 5), some normative propositions and relations can be formalized in a plausible way by combining deontic operators and quantifiers. The semantics outlined above can be extended in an obvious way to quantified deontic logic by adding to our formal language quantifiers, predicative expressions, individual variables, individual parameters (arbitrary names), and functional expressions which can be used for generating complex individual terms from simple terms. The models of quantified SDL are structures $\langle W,\ R,\ U, D, I\rangle$, where W is a universe of possible situations ("worlds"), R is a binary alternativeness (accessibility) relation between situations (as with propositional SDL), U is a set of individuals, D is function which assigns a subset of U to each $v \in W$ —the individuals existing in v, $D(v)$, and the interpretation function I assigns to each non-logical expression (individual term, predicative expression, or functional expression) the intension of the expression, that is, a function from possible situations to extensions or referents:

(7.3) (i) For a simple individual term (parameter) c, $I(c, u) \in D(u)$.

(ii) For each n-place predicate G, $I(G, u)$ is a set of ordered sets of n individuals (n-tuples) $\langle i_1, i_2, ..., i_n\rangle$, where each $i_j \in D(u)$, that is, $I(G, u)$ is a subset of $D(u)^n$.

(iii) For each function symbol f, $I(f, u)$ is an operation on $D(u)$, that is a function which has $D(u)$ as its domain as well as its range of values.

For example, if G is the relation of loving, $I(G, u)$ is the set of all ordered pairs of individuals $c \in D(u)$, $d \in D(u)$ such that c loves d in the situation (world) u.

The truth-conditions of quantified sentences may be defined in some standard way, for example, in terms of variant interpretations (cf. [Bostock, 1997, pp. 85-6] and [Mates, 1965, pp. 54-6]). If M is a model with an interpretation function I, let M/c be a model with an interpretation function I/c, called the *c-variant of* I, which is like I except that it may assign a different individual to the singular term c; thus I/c and $M/$c differ from I and M at most with respect to the value of c. As long as c does not appear in the formula (open sentence) Φ, c can be regarded as denoting any arbitrary individual in the relevant domain under some variant of I, I/c, and the sentence $\forall x\Phi$ is then true under I if and only if the sentence $\Phi(c/x)$ obtained by substituting c for x in Φ is true under every c-variant of I. Thus the truth-conditions of quantified sentences can be expressed as follows:

(7.4) $M, I, u \models \forall x\Phi$ if and only if $M/c, I/c, u \models \Phi(c/x)$ for every c-variant I/c of I such that $\mathrm{I}/c(c) \in D(u)$, where the parameter c does not appear in Φ.

(7.5) $M, I, u \models \exists x\Phi$ if and only if $M/c, I/c, u \models \Phi(c/x)$ for some c-variant I/c of I such that $\mathrm{I}/c(c) \in D(u)$, where the parameter c does not appear in Φ.

The truth-conditions of other complex sentences and atomic sentences are defined in the standard way, except that they are relativized to possible situations, and the meanings of deontic operators are defined in the same way as in propositional deontic logic.

The individual variables of quantified deontic logic may be interpreted as variables for individual actions, as in the examples discussed in Section 5, or as variables for agents, and the domains $D(u)$ may be interpreted in the similar ways. (See [Åqvist, 1987, pp. 84-5].)[27] The function $D(u)$ may be a constant function, in which case all situations $u \in W$ have the same domain of individual objects, or the situations may involve different domains. The deontic counterpart of the Barcan formula,

(7.6) $\forall x\mathbf{O}Gx \rightarrow \mathbf{O}\forall xGx$,

states that if G is obligatory (or required) for everyone, then it ought to be so that everyone satisfies G. It is clear that this inference is not valid in all applications, but it is valid in all models with constant domains. For suppose the antecedent is true at u, then for each individual i at u, at each deontic alternative to u, i satisfies G. But if the domains at each world are the same, then the individuals from u satisfying G at the alternatives are all the individuals there are at the alternatives. So at each u-alternative, everyone individual satisfies G, so the consequent must be true at u. (In ordinary idiomatic English, a sentence like, 'Everyone ought to be happy' may be understood as expressing either a wide-scope (de dicto) proposition to the effect that it ought to be the case that everyone is happy or a narrow scope (de re) proposition to the effect that for each person that exists, it ought to be the s/he is happy.) (7.6) does not hold in models in which the domain of individuals may expand when we move from a situation to one of its deontic alternatives. For example, suppose at u there are just 10 people left and that that is perfectly evident to each of those 10 people, so that each ought to believe there are just ten people left; if there are u-alternatives with expanded domains, although it will follow that each of these ten will there believe there are just ten people left, it will not follow that the additional

[27] Although see [McArthur, 1981] for problems with interpreting the variables as ranging over concrete actions.

other people there will share that belief. It is easy to formally verify that the Barcan formula does not hold in all variable domains models. On the other hand, the converse of (7.6) holds both in constant domain and in variable domain models in which the domains of all deontic alternatives to a world u must contain every individual that exists at u (and perhaps more), but the conditional

(7.7) $\forall x\mathbf{P}Gx \rightarrow \mathbf{P}\forall xGx$,

that is,

(7.8) $\mathbf{O}\exists xGx \rightarrow \exists x\mathbf{O}Gx$,

is invalid in both kinds of models. Everyone is permitted to have a dinner in Casa Paco, a public restaurant, but no situation in which everyone is having dinner in Casa Paco is permitted (normatively acceptable), because the legal seating capacity of the restaurant is 40 customers.[28] In the same way, the sentence 'Someone ought to rescue the cat Gussie from the shelter for abandoned pets' is ambiguous: It can be understood as having the form of the antecedent of (7.8) (a wide-scope ought) or the form of its consequent, and the former interpretation does not mean that some specific person has a (personal) obligation to rescue Gussie. In these respects deontic modalities are logically similar to alethic and epistemic modalities. It should be observed that both interpretations of (7.8) can be regarded as ought-to-do propositions in the sense that the predicate in the scope of the deontic operator may be an action predicate. The failure of (7.7) and (7.8) to hold makes possible the tragedy (or paradox) of the commons and other similar problems. An attempt by everyone to perform in the same situation or at the same time what they take to be a permitted action can have normatively unacceptable consequences. (See [Hardin, 1968] and [McConnell and Brue, 2002, p. 596].) Different assumptions about deontic alternativeness relation and about the domains of individuals lead to different systems of quantified deontic logic. (For quantifiers in modal logic, see [Garson, 2001], [Girle, 2009, pp. 106-125], [Priest, 2008, pp. 308-48] and [Bell *et al.*, 2001, pp. 171-183]; see also [Hintikka, 1957; McArthur, 1981; Goble, 1994; Goble, 1996] on quantifiers in deontic logic.)

The supplement to this section provides some more formal details (for the propositional systems).

[28] Or per (7.8), it may be obligatory that someone leave the lifeboat (else no one will be saved), but not that there is some one person such that she is obligated to leave, else there would be no need to draw straws to transform the first situation into one like the second, and it would also mean that at least someone in the boat could not go beyond the call by going overboard voluntarily, since s/he would be obligated to do so by (7.8).

Supplement to Section 7: some formalities

A7.1 Semantics for SDL and SDL$^+$:

We first define the frames (structures) for modeling SDL.

F is an SDL (or KD) frame: $F = \langle W, A\rangle$ where:

1. W is a non-empty set (the points or worlds)
2. A is a subset of $W \times W$ (the acceptability relation)
3. A is serial: $\forall u \exists v Auv$.

A model is such a frame paired with an assignment function from the sentence letters of the language of SDL to the subsets of W:

M is an SDL model: $M = \langle F, I\rangle$, where F is an SDL Frame and I is a function from the propositional letters to subsets of W in the frame (the "truth sets" for the letters).

We now define truth in a model for all sentences of the language of SDL, where "$M, u \models p$" stands for "p is true at u in model M":

Basic truth-conditions: (Here and occasionally in proofs we will use "PC" as short for truth-functional propositional calculus.)

(PC) $M, u \models p$ iff $u \in I(p)$

$M, u \models \neg p$ iff $M, u \not\models p$ where p is a sentence letter

$M, u \models p \to q$ iff either $M, u \not\models p$ or $M, u \models q$

(**O**B) $M, u \models \mathbf{O}\mathrm{B}p$ iff $\forall v$ (if Auv then $M, v \models p$)

Derivative truth-conditions:

(Truth functional operators as usual.)

(**P**E) $M, u \models \mathbf{P}\mathrm{E}p$ iff $\exists v$ (Auv and $M, v \models p$)

(**I**M) $M, u \models \mathbf{I}\mathrm{M}p$ iff $\neg\exists v$ (Auv and $M, v \models p$)

(**O**M) $M, u \models \mathbf{O}\mathrm{M}p$ iff $\exists v$ (Auv and $M, v \models \neg p$)

(**O**P) $M, u \models \mathbf{O}\mathrm{P}p$ iff $\exists v$ (Auv and $M, v \models p$) & $\exists v$ (Auv and $M, v \models \neg p$)

Truth in a model: $M \models p$ iff p is true at every world in M.
Validity in a class C of models: $C \models p$ iff $M \models p$, for every M in C.

Recall that SDL$^+$ was used for convenience to denote the result of adding A4, $\mathbf{O}$B($\mathbf{O}$B$p \to p$), to SDL. We also noted that the semantic constraint associated with this was secondary reflexivity:

(COO) $\forall u \forall v\,(Auv \to Avv)$

We now note two well-known metatheorems:

Metatheorem 7.1 *SDL is determined by the class of all SDL models. 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).*

Metatheorem 7.2 *SDL$^+$ is determined by the class of all secondary reflexive SDL models.*

For key elements see [Åqvist, 2002]; but for some additional metatheory see [Åqvist, 1987].

A7.2 Semantics for LKA$_1$ and LKA$_2$:

F is an LKA$_1$-frame: $F = \langle W, R, G \rangle$ where:

1. W is a non-empty set
2. R is a subset of $W \times W$ (accessibility relation)
3. G is a subset of W (deontically acceptable worlds)
4. $\forall u \exists v\,(Ruv \,\&\, v \in G)$.

Note that here the acceptable worlds do not vary relative to a world as in the SDL frames.

We then add an assignment function to get a model:

M is an LKA$_1$-model: $M = \langle F, I \rangle$, where F is an LKA$_1$ frame and I is a function from the propositional letters to various subsets of W in the frame.

The truth conditions for formulas can now be given.

Basic truth-conditions at a world, u, in a model, M:

(PC) same as for SDL

($\Box$) $M, u \models \Box p$ iff $\forall v$ (if Ruv then $M, v \models p$

(g) $M, u \models g$ iff $u \in G$

Derivative truth-conditions:

(Truth functional operators as usual.)

($\Diamond$) $M, u \models \Diamond p$ iff $\exists v$ (Ruv and $M, v \models p$)
(**O**B) $M, u \models \mathbf{O}\mathrm{B}p$ iff $\forall v$ (if $Ruv \,\&\, v \in G$ then $M, v \models p$)
(**P**E) $M, u \models \mathbf{P}\mathrm{E}p$ iff $\exists v$ ($Ruv \,\&\, v \in G \,\&\, M, v \models p$)
(**I**M) $M, u \models \mathbf{I}\mathrm{M}p$ iff $\forall v$ ((if $Ruv \,\&\, v \in G$ then $M, v \models \neg p$)
(**O**M) $M, u \models \mathbf{O}\mathrm{M}p$ iff $\exists v$ ($Ruv \,\&\, v \in G \,\&\, M, v \models \neg p$)
(**P**E) $M, u \models \mathbf{O}\mathrm{P}p$ iff $\exists v$ ($Ruv \,\&\, v \in G \,\&\, M, v \models p$) & $\exists v$ ($Ruv \,\&\, v \in G \,\&\, M, v \models \neg p$)

Truth in a model and validity in a class of models is defined as above for SDL.

Recall that LKA_2 was used for convenience to denote the result of adding axiom T, $\Box p \to p$, to LKA_1. The semantic constraint associated with T is reflexivity: $\forall u Ruu$.

Metatheorem 7.3 *LKA_1 is determined by the class of all LKA_1 models.*

Metatheorem 7.4 *LKA_2 is determined by the class of all reflexive LKA_1 models.*

Metatheorem 7.5 *The pure deontic fragment of LKA_1 is SDL.*

Metatheorem 7.6 *The pure deontic fragment of LKA_2 is SDL^+.*

A7.3 Semantics for VW:

We first define the frames (structures) for modeling VW.

F is an VW Frame: $F = \langle W, A \rangle$ where:

1. W is a non-empty set (the points or worlds)
2. A is a subset of $W \times W$ (the acceptability relation)

This is the same as the definition of the SDL frames except that seriality is dropped.

An assignment function is added to get a model:

M is an VW Model: $M = \langle F, I \rangle$, where F is an SDL Frame and I is a function from the propositional letters to subsets of W in the frame.

We now define truth in a model for all sentences of the language of VW:

Basic truth-conditions:

(PC) same as for SDL, including $M, u \not\models \bot$

(**O**B) $M, u \models \mathbf{O}\mathrm{B}p$ iff $\exists v Auv \,\&\, \forall v$ (if Auv then $M, v \models p$)

Derivative truth-conditions:
(Any remaining truth-functional operators as usual, including for $\top$: $M, u \models \top$.)

(**P**E) $M, u \models \mathbf{P}\mathrm{E}p$ iff $\neg\exists v Auv \vee \exists v\, (Auv \,\&\, M, v \models p)$

(**I**M) $M, u \models \mathbf{I}\mathrm{M}p$ iff $\exists v Auv \;\&\; \neg\exists v\, (Auv \,\&\, M, v \models p)$

(**O**M) $M, u \models \mathbf{O}\mathrm{M}p$ iff $\neg\exists v Auv \vee \exists v\, (Auv \,\&\, M, v \models \neg p)$

(**O**P) $M, u \models \mathbf{O}\mathrm{P}p$ iff $\neg\exists v Auv \vee [\, \exists v\, (Auv \,\&\, M, v \models p)$
$\&\, \exists v\, (Auv \,\&\, M, v \models \neg p)\,]$

Truth in a model: $M \models p$ iff p is true at every world in M. Validity in a class C of models: $C \models p$ iff $M \models p$, for every M in C.

Recall that SDL is VW+$\mathbf{O}\mathrm{B}\top$, and that for the semantic clause for **O**B above, the semantic constraint associated with $\mathbf{O}\mathrm{B}\top$ is serialty itself: $\forall u \exists v Auv$.

Recall that DD in this framework is valid without any semantic constraints, but not $\mathbf{O}\top$, thus reversing the situation in the standard SDL semantics.

We now state two metatheorems:

Metatheorem 7.7 ([McNamara, 1988]) *VW is determined by the class of all VW models.*

Metatheorem 7.8 ([McNamara, 1988]) *SDL is determined by the class of all serial VW models.*

8 Problems and paradoxes regarding the standard systems

In this section[29] we will consider some of the "paradoxes" associated with the "standard systems": SDL and suitably similar systems, here to include the two expressively stronger LKA systems that generate SDL, and the logically weaker VW system, which shares the same language as SDL. The use of "paradox" is widespread in discussions of deontic logic, and is consistent with a broad use of the term elsewhere, but "puzzles", "challenges", "problems", "dilemmas" will often be used, and seems less loaded. The number of problems attributed to standard systems is large, and these have often served to fuel new work after the classic period of the 1950s. We will list and briefly describe many of them, grouped under various associated headings (e.g. principles that are often thought to figure centrally in the associated puzzles). There will be both continuity with some of the earlier historical material presented, and occasional repetition of coverage, since we wish this section and its subsections to be something a reader might at times consult without the preceding material in focus; there will also be more detailed coverage of puzzles that have received the most attention or seemed the most challenging.

8.1 A puzzle with the very idea of deontic logic

Jørgensen's dilemma - truth and normative language[30]

(This was discussed in considerable detail in Section 4 above. Here we merely give a very compressed sketch.)

The view that evaluative sentences (e.g. "That is beautiful/ugly", "That is good/bad", "That is wrong/right") are not the sort of sentences that can be either true or false was held by many researchers in the first half of the twentieth century, especially during the heyday of positivism. This leads to a dilemma. Deductive logic involves the study of what follows from what. Truth is essential to deductive consequence, as well as to notions of consistency, entailment, contradiction, etc. But then deontic logic is impossible, since its sentences are among the evaluative ones, and thus neither true nor false. Yet normative sentences of the sort studied in deontic logic do seem to stand in familiar logical relationships to one another, so deontic logic must be possible after all.

A widespread distinction was made between *norms* and *normative propositions*[31]. The idea is that a normative sentences such as "You may enter

[29] This section benefits from [McNamara, 2006; McNamara, 2010].

[30] Cf. [Jørgensen, 1937 and 1938].

[31] [Hedenius, 1941; von Wright, 1963; Alchourrón and Bulygin, 1981; Alchourrón and Bulygin, 1971; Makinson, 1999; Stenius, 1963]. [von Wright, 1963] credits Hedenius for

freely" 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 for free entrance to a museum). 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 of the sentence by the authority. In contrast, the use by the passerby is deemed descriptive: it is used to report that permission to do so is a standing state, not to grant permission. Often the two uses are deemed mutually exclusive, with only the latter use allowing for truth or falsity. Some have challenged the exclusiveness, by appealing to speech-act theory along with semantics. The idea is that the one in authority not only grants the permission by performing the speech act of uttering the relevant sentence "You may enter freely", but also thereby makes what it said true (that you may enter freely).[32] Still, many believe that norms are nonetheless distinct from normative propositions, and that a logic of norms is also needed.

"Input-Output Logic" is a recent robust program to provide a logical framework for norms as non-truth-evaluable items. (See e.g. [Makinson, 1999; Makinson and van der Torre, 2000; Makinson and van der Torre, 2003], and the chapter by Parent and van der Torre in this volume.) In a more general vein, there is the older tradition of developing logics for imperatives, and the debate about whether there can even be such. See Hansen's chapter in this volume on imperatival logic. [Vranas, 2010] provides a recent defense of the possibility of imperatival logic, and Vranas [Vranas, 2011] offers a new theory of validity for imperatival inference. Let us also mention another tradition, the imperative-based (or norm-based) approach to deontic logic, which focuses not on a logic of imperatives (perhaps even denying that possibility), but on the use of imperatives as a key foundational component in a semantics for logics for truth-evaluable deontic sentences. See for example, [Hruschka, 2004; Hansen, 2008; Horty, 1994; Horty, 1997; Horty, 2003] and the seminal [van Fraassen, 1973].

8.2 A problem centering around $\mathbf{O}\top$

The logical necessity of obligations problem[33]

the distinction.

[32]Cf. [Lemmon, 1962b; Kamp, 1974; Kamp, 1979]. [Kempson, 1977] argues that performative utterances often work this way. For example, if a legitimate authority in the right context pronounces two people married, it may not only be the case that the speech act performed renders them married, but the sentence "You are now married" may be a true description at its moment of utterance, the dual character perhaps captured by "You are, hereby, married".

[33]We are unaware of any standard name for this problem.

Consider the apparent possibility that

(8.1) Nothing is obligatory.

A natural representation of this in the language of the standard systems (with quantifiers added) is:

(8.2) $\neg\exists q\mathbf{O}q$.

But RND of SDL entails ON, $\vdash \mathbf{O}\top$, so supplemented with propositional quantification, we would get $\vdash \exists q\mathbf{O}q$. SDL thus seems to imply that it is a truth of logic that something is obligatory—that there could not be a situation with no obligations; yet (8.1) appears to express something not only possible, but plausibly thought to be true at times in the past in our universe, so SDL appears to be too strong [Chellas, 1974]. This holds for all the standard systems except VW. [von Wright, 1951] argues that neither $\mathbf{O}\top$ nor $\mathbf{P}\bot$ nor their denials, are logical truths, so we should opt for their absence as logical truths as a "principle of contingency" for deontic logic. The weakening of SDL described in Section 7.3 above, VW, captures this absence.

Later, in [von Wright, 1963, pp.152-4], von Wright argues that $\mathbf{O}\top$ does not express a real prescription. [Føllesdal and Hilpinen, 1971, p. 13] argue that ON of SDL at best excludes "*empty* normative systems" with no obligations, and that furthermore, since no one can fail to fulfill $\mathbf{O}\top$ anyway, it is not a pressing concern.[34] However, no one can bring it about that $\top$, so it would seem that no one can fulfill $\mathbf{O}\top$, although no one can violate it either. [al Hibri, 1978] discusses various early takes on this problem, rejects ON, and later develops a deontic logic without it. [Jones and Pörn, 1985] explicitly reject ON for "ought" in the system developed there, where the concern is with what people ought to do.

Note that reading $\mathbf{O}$ as "it ought to be the case that", as it often was read, makes it less clear that ON is problematic. "It ought to be that contradictions are false" does not sound jarring, but here there is no longer a clear link to "$\mathbf{O}$" and what agents ought to do or bring about.

8.3 Puzzles centering around the rule RMD

The violability of obligations problem[35]

It is often thought that a central distinguishing mark of obligations is that they are *violable* in principle, unlike purely factual claims. It seems hard to swallow that it is obligatory that the sun will set today, much less that

[34] See also [Prior, 1958].

[35] This objection is suggested by remarks in [von Wright, 1963, p. 154].

it is obligatory that either it does set today or it is not the case that it does, something that couldn't be otherwise on purely logical grounds. This suggests the following as a conceptual truth about obligations:

(8.3) If p is logically necessary, it is logically impossible that it is obligatory that p

In SDL, a natural expression of such a violability constraint would seem to be this rule:

(8.4) If $\vdash p$ then $\vdash \neg \mathbf{O}p$ (Violability)

But Violability immediately yields $\neg\mathbf{O}\top$ directly contradicting theorem ON of SDL. So no SDL (or stronger) system can consistently rule out inviolable obligations.[36]

For systems with the expressive resources of the LKA systems, a stronger violability condition is expressible, to the effect that nothing obligatory is necessary:

(8.5) $\Box p \to \neg\mathbf{O}p$ (Violability')

But this is inconsistent with all LKA systems, since $\Box p \to \mathbf{O}p$ and $\Box\top$ are theses.

Note that even in a system weaker than SDL, one that lacked RND and ON, as long as the rule RMD (if $\vdash p \to q$ then $\mathbf{O}p \to \mathbf{O}q$) is derivable, then adding the Violability rule above would render the system useless. For by PC, $\vdash p \to \top$, so it follows by RMD that $\vdash \mathbf{O}p \to \mathbf{O}\top$; but since by PC, $\vdash \top$, by Violability, it follows that $\vdash \neg\mathbf{O}\top$, and hence $\vdash \neg\mathbf{O}p$. Thus with Violability added to such a system, we get as a thesis that nothing is obligatory. And this means that although the weaker VW can consistently rule out inviolable obligations, it can do so only at the expense of ruling out all obligations.[37]

Free choice permissions puzzle[38]

Consider:

(8.6) You may either have <u>c</u>ake or <u>i</u>ce cream.[39]
(8.7) You may have cake and you may have ice cream.

[36] VW can, but only at the expense of ruling out violable obligations, and thus all obligations, as well.

[37] [Jones and Pörn, 1985] design a system explicitly intended to countenance a violability condition for one of their operators, and this constraint is endorsed in [Carmo and Jones, 2002] as well.

[38] Cf. [Ross, 1941].

[39] Underlined letters suggest our intended symbolization schemes.

Natural symbolizations of (8.6) and (8.7) in the language of the standard systems are:

(8.8) $\mathbf{P}(c \vee i)$

(8.9) $\mathbf{P}c \,\&\, \mathbf{P}i$

Furthermore, it is also natural to see (8.7) as following from (8.6): if I am permitted to have either, then having each is permissible (though perhaps not both). But (8.9) does not follow from (8.8) in standard systems. This is not a theorem:

(P*) $\mathbf{P}(p \vee q) \rightarrow (\mathbf{P}p \,\&\, \mathbf{P}q)$.

Furthermore, if (P*) were added to a system that contained VW (and thus to any that contained SDL), disaster would result. For from RMD and the definition of **P**, we get $\mathbf{P}p \rightarrow \mathbf{P}(p \vee q)$.[40] But then with (P*), it would follow that $\mathbf{P}p \rightarrow (\mathbf{P}p \& \mathbf{P}q)$, for all q, so we easily generate a theorem to the effect that everything is permissible if anything is:

(P**) $\mathbf{P}p \rightarrow \mathbf{P}q$,

This is absurd on its face, and in any system containing VW, it would in turn yield as a theorem that nothing is obligatory:

(8.10) $\vdash \neg \mathbf{O}p$.

For reductio, assume $\mathbf{O}p$, for some p. Then by DD (No Conflicts), it follows that $\mathbf{P}p$. But one instance of (P**) is $\mathbf{P}p \rightarrow \mathbf{P}\neg p$, so we would then have $\mathbf{P}\neg p$, which by RED (Substitution of Provable Equivalents), PC and the definition of **P** generates $\neg \mathbf{O}p$, contradicting our assumption.

This puzzle has led many to conclude that there are two senses of "permissibility" that need to be separated out, and that the language of SDL (and thus VW) represents only one of those. One sense might be the simple absence of a prohibition, and thus expressible in standard systems. The other might be a stronger sense of permission that would support (P*) but without supporting (P**), as suggested in [von Wright, 1968]. [Føllesdal and Hilpinen, 1971] (and [Carmo and Jones, 2002]) wonder if this problem might not be a pseudo-problem calling for no more than SDL's expressive resources, since the conjunctive sense of an "or-permission" can simply be expressed as a conjunction of permitting conjuncts, $\mathbf{P}p \& \mathbf{P}q$, and the weaker sense as a permitted disjunction, $\mathbf{P}(p \vee q)$. [Kamp, 1974;

[40]Since $\vdash \neg(p \vee q) \rightarrow \neg p$, $\vdash \mathbf{O}\neg(p \vee q) \rightarrow \mathbf{O}\neg p$, and thus $\vdash \neg\mathbf{O}\neg p \rightarrow \neg\mathbf{O}\neg(p \vee q)$.

Kamp, 1979] provide a nuanced analyses of the semantics and pragmatics of permission statements, including disjunctive ones. Here we have only skimmed the surface. For more on the rich topic of various concepts and analyses of permission, see the chapter in this volume by Hansson on the topic.

Ross's paradox[41]

Consider:

(8.11) It is obligatory that you mail the letter.
(8.12) It is obligatory that you mail the letter or you burn the letter.

Natural renderings in the standard systems are:

(8.13) $\mathbf{O}m$

(8.14) $\mathbf{O}(m \vee b)$

However, $\vdash \mathbf{O}m \rightarrow \mathbf{O}(m \vee b)$ follows by RMD from $\vdash m \rightarrow m \vee b$, so (8.14) follows from (8.13) in any VW system, but arguing from (8.11) to (8.12) seems rather odd. Among other things, it seems to suggest that the obligation expressed in (8.11) to mail the letter automatically generates a distinct obligation that I am able to fulfill by burning the letter. Of course, the latter is presumably forbidden, but it remains odd to think I could plead partial mitigation in failing to mail the letter by burning it instead with "Well, at least I fulfilled my obligation to mail or burn it".

The good samaritan paradox[42]

Consider:

(8.15) It is obligatory that Jones help Smith who is being mugged.
(8.16) It is obligatory that Smith is being mugged.

Now the following equivalence appears to be logically true:

(8.17) Jones helps Smith who is being mugged if and only if Jones helps Smith and Smith is being mugged.

But relying on this equivalence, if we then symbolize (8.15) and (8.16) in the language of VW in the most natural way, we get:

(8.18) $\mathbf{O}(h \& m)$

(8.19) $\mathbf{O}m$

[41] Cf. [Ross, 1941].
[42] Cf. [Prior, 1958].

But $(h\&m) \rightarrow m$ follows by truth-functional logic, so by RMD, it follows that $\mathbf{O}(h\&m) \rightarrow \mathbf{O}m$, and then we can derive (8.19) from (8.18). But does it really follow from its being obligatory that Jones come to the aid of Smith who is being mugged that Smith's mugging it itself obligatory?

Note that Prior casts this paradox in a prohibition form, using this trivial variant of RMD given the definition of $\mathbf{F}$: If $\vdash p \rightarrow q$ then $\vdash \mathbf{F}q \rightarrow \mathbf{F}p$, suggesting that the impermissibility of Smith being robbed implies the impermissibility of helping him who is being robbed).

It is also doubtful that the paradox is due to the fact that there are two people involved. It can be recast with just one agent via $\mathbf{F}$(i.e. $\mathbf{O}\neg$) as "The victim's paradox": the victim of a crime can help herself only if there is a crime, but then, it will follow under similar symbolization that it is impermissible for the victim of the crime to help herself, since the crime is impermissible. Similarly for "The repenter's paradox": the robber repents for his crime only if there is a crime, and so by similar reasoning we might get a symbolization suggesting that repenting is wrong since the crime is wrong. These early variations were used to argue against certain early proposed solutions to the good samaritan paradox (e.g. [Nowell Smith and Lemmon, 1960]).

Åqvist's paradox of epistemic obligation[43]

Here is a much discussed variant of the preceding paradox. Consider:

(8.20) The bank is being robbed.
(8.21) It is obligatory that Jones (the guard) knows that the bank is being robbed.
(8.22) It is obligatory that the bank is being robbed.[44]

Let us imagine that we have added a logic for a propositional knowledge operator, K. We can then let "Kr" symbolize "Jones knows that the bank is being robbed", and then a natural way to symbolize (8.20)-(8.22) in a VW system so augmented is:

(8.23) r
(8.24) $\mathbf{O}Kr$
(8.25) $\mathbf{O}r$

But a logic for propositional knowledge will presumably support knowledge's entailment of truth (e.g. that if Jones knows that the bank is being robbed then the bank is being robbed). So $Kr \rightarrow r$ would be a thesis of any such

[43] Cf. [Åqvist, 1967].
[44] (8.20) is inessential but is listed to suggest part of the natural context for (8.21).

augmented VW system. But then it would follow by RMD that $\mathbf{O}Kr \rightarrow \mathbf{O}r$ is also a thesis of any such system, and we can then use that thesis to derive (8.25) from (8.24) by MP. Applied to the case above, the logic seems to suggest that (8.22) follows from (8.21), which seems absurd. It seems that it is obligatory for the guard to be in the know about any bank robberies taking place, and so this one, but surely it does not follow that it is thereby also obligatory that the robbery take place. Thus it appears to be in the spirit of the standard systems, and any weaker systems endorsing RMD, to generate fictitious consequences from cases of obligatory knowledge.

It is also worth noting that we should not be misled by the typical examples into thinking of this problem as "such logics can't handle obligatory knowledge of *contrary-to-duty facts*", for the same problem extends to cases of facts that are optional in normative status. If a reformed but known bank robber, Jones, enters the bank, it may also be that it is obligatory that the guard knows Jones is in the bank, and then RMD appears to wrongly entail that it is obligatory that Jones is in the bank, even though it is completely optional that Jones is there.

There have been a variety of responses to these RMD-related paradoxes. One response to these has been to try to explain them away. For example, Ross's paradox is often quickly dispensed with as based on confusion or as not really being a problem when the system in question is properly understood (e.g. [Føllesdal and Hilpinen, 1971; Brown, 1996a]). Others deflect it arguing that it is semantically correct and only pragmatically odd, and reflects features that any adequate theory of the pragmatics of deontic language must predict, so no special problem for deontic logic [Castañeda, 1981]. It is also often suggested that regarding the good samaritan paradox, RMD is not the real culprit because, viewed rightly, it is really a conditional obligation paradox [Castañeda, 1981; Tomberlin, 1981]. Others however suggest that things are as they seem and the above paradoxes are of a piece in all genuinely invoking RMD and reflecting RMD's problematic character for genuine deontic reasoning. [Jackson, 1985; Goble, 1990a] are closely related examples of approaches to deontic logic rejecting RMD from a principled philosophical perspective. [Jackson, 1985] links an "ought to be" operator to counterfactuals and informally explores its semantics and logic; whereas [Goble, 1990a] takes a similar approach but generalizes the idea to cover "good" and "bad" as well, with [Goble, 1990b] providing characterization results for the identified logics. Interestingly, their approaches also intersect with the philosophical issue of "actualism" and "possibilism" in ethical theory.[45] [Loewer and Belzer,

[45] Possibilists assert that an agent is obligated to bring about any p that is part of the the optimal overall outcome she could achieve by her actions, even when the goodness of

1986] provides an interesting discussion of the traditional puzzle, as well as Forrester's puzzle (below), in terms of their system 3D. [Hansson, 1990; Hansson, 2001] systematically explore systems of deontic logic in terms of general attributes of different preference orderings, using these to classify types of normative predicates or operators as *prohibitive* and *prescriptive* (e.g. a prohibitive status as one where anything worse than it has that status also). His work is predicated on the assumption that RMD is a key source of the main paradoxes of the standard systems, and so he devises non-standard systems intended to not countenance principles such as RMD. In this vein, see Hansson's chapter on alternative semantics for deontic logics in this volume. [Hansson, 2001] is also important in its own right 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).[46] Opinions about these puzzles we have grouped together need not be monolithic of course. For example, [Carmo and Jones, 2002] take Ross's puzzle seriously, the free choice permission puzzle to be a pseudo problem, and the good samaritan puzzle to be resolvable using the resources needed for resolving puzzles with deontic conditionals.

8.4 Puzzles centering around DD and OD

Sartre's dilemma - conflicting obligations[47]

A conflict or *dilemma* is a situation where there are one or more obligations not jointly realizable. The typical case involves a conflict of two obligations. For example, suppose I promised Mary to meet her, and that I promised another friend that I would not meet Mary. It would then seem that I have, by my promises, made the following true:

---

p depends on all sorts of other things that she would not in fact bring about were she to bring about p. Actualists assert that an agent is obligated to bring about any p if that would in fact be better than not doing so, and this of course can crucially depend on what else I would do (optimal or not) were I to bring about p ([Jones and Pörn, 1986; Jackson, 1988; Greenspan, 1975; Goldman, 1976; Thomason, 1981a] provide early discussions.)

[46] [van der Torre, 1997] is a nice general source covering issues surrounding RMD as well, along with much else.

[47] Cf. [Lemmon, 1962a]. In the original example, a young man is obliged to avenge his brother's death (by leaving home and fighting the Nazi occupation) and he is also obligated to stay home and aid his mother (devastated by the loss of the brother). [von Wright, 1968] talks of "predicaments" and cites the Book of Judges, where Jephthah promises God he will sacrifice the first being he meets on his way home from war, if God gives him victory. God does, and the first being he meets upon his return is his beloved daughter. Note that both of these are plausibly thought of as dual-sourced obligations, but our example, following [Marcus, 1980], reflects the possibility of conflicts generated by a single normative principle (e.g. it is obligatory to keep one's promises), which in turn reinforces the idea that conflicts of obligation can be circumstantial and needn't be generated by normative systems that are supposedly inconsistent (cf. [Williams, 1965]).

(8.26) It is obligatory that I meet Mary (now).
(8.27) It is obligatory that it is not the case that I meet Mary (now).

If so, then I have an explicit conflict of obligations. People generate conflicting appointments easily enough under pressure to please, in forgetful moments, due to errors in our calendar entries, etc. It also appears that they result in conflicting obligations in a perfectly ordinary sense of the term.[48] But a natural first blush representation of these in the language of VW and SDL is:

(8.28) $\mathbf{O}m$

(8.29) $\mathbf{O}\neg m$

But given DD, $\mathbf{O}p \to \neg\mathbf{O}\neg p$, is a theorem of VW, we are quickly led from the conflict expressed by (8.28) and (8.29), to the contradiction expressed by $O\neg m \& \neg O\neg m$. So we must conclude that (8.28) and (8.29) make an inconsistent pair per VW. Yet, the original seems not only logically coherent but all too familiar.[49]

At the end of this section, there is a supplement where we consider some challenges faced once we decide to develop conflict tolerant logics.

A puzzle surrounding Kant's law

Kant's law typically involves a notion of possibility stronger than that of mere logical or metaphysical possibility. In discussions in ethical theory, where "Kant's law" arose, it is agential:

(KL) Anything morally obligatory for an agent must be *within the agent's ability.*[50]

This principle has been widely advocated in ethical theory, one thought being that, at least for all things considered obligations, the fact that something is not even in an agent's power to do is itself a sufficient consideration

[48] These obligations need not be all-things-considered-non-overridden obligations, but that does not entail that these are not obligations (any more than "it's not a brown dog" entails "it's not a dog") nor does it mean that we needn't model them. For a simple framework that allows for such obligations, as well as comparing them, see [Brown, 1996b].

[49] [Lemmon, 1962a] argues early on that a conflict of obligations may involve no contradiction. [Williams, 1965] stresses the contingency of conflicting obligations and briefly contrasts this with inconsistency as unrealizability in any world. [Marcus, 1980] argues explicitly for the standard world-theoretic conception of consistency as joint realizability in some world in some model (not in all worlds in all models, as with say the set of tautologies). See also [McNamara, 1996].

[50] It is sometimes used more broadly in deontic logic for a weakened version, one that follows from Kant's stronger version. See next puzzle.

to eliminate it from further consideration in a determination of what is to be all in all required. In an optimizing framework like utilitarianism, (KL) is strongly supported by the standard maxim that one is morally obligated to do the best she *can*. It is also often endorsed in various deontic-agential frameworks (e.g. [Horty, 2001; McNamara, 2000]).

But now consider:

(8.30) I'm obligated to pay you back $100 by tonight.
(8.31) I can't pay you back $100 by tonight (e.g. I just spent it on something shopping).

Let us represent the above sentences in the language of our LKA systems where we have a possibility operator. Although agency is not itself represented in the LKA systems, we can still interpret the possibility operator therein as "what is *consistent with the abilities of* some background agent, Jane Doe", and likewise for the deontic operators we might interpret them as indicating what is obligatory for such a Jane Doe [Brown, 1992; McNamara, 2000].[51] (8.30 and (8.31) might then be naturally symbolized as follows:

(8.32) $\mathbf{O}p$

(8.33) $\neg\Diamond p$

(8.30) and (8.31) appear to be consistent. Alas, people often wind up with financial obligations they cannot fulfill, be it from neglect, unforeseen circumstances, or whatever. So it seems that the notion of an unfulfillable obligation is no contradiction in terms. But in the LKA systems, it is a theorem that $\mathbf{O}p \rightarrow \Diamond p$. So from (8.32) and (8.33) we get $\Diamond p \& \neg\Diamond p$, a contradiction, and so (8.32) and (8.33) are inconsistent. Yet (8.30) and (8.31) seem consistent. I can clearly *owe* money I'm unable to pay back, but doesn't that ordinarily entail that I have a financial *obligation* I cannot meet?

One strategy here might be to posit ambiguity or context shift and employ a distinction between deliberative contexts of evaluation and judgmental contexts as suggested by [Thomason, 1981a] and [Thomason, 1981b], bolstered by arguing that we need the distinction elsewhere anyway.[52] In judgmental contexts (or the judgmental sense), evaluations such as (8.30)

[51] The puzzle would remain even if agency and agential ability were explicitly represented.

[52] [Thomason, 1981a] credits [Greenspan, 1975; Powers, 1967] for stressing the contextuality of oughts. In the distinct context of the contrary-to-duty paradox (discussed below), others have endorsed the contextuality or ambiguity of *oughts* (e.g. [Jones and Pörn, 1985; Prakken and Sergot, 1996; Carmo and Jones, 2002]).

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 then but need no longer be; whereas in deliberative contexts, where we are focused on what to do now, we deny that there is an obligation to do what is now undoable. Whether this interesting distinction provides a truly satisfactory solution to the above problem is beyond the scope of this essay, but the puzzle appears to be underexplored in deontic logic.

Conflation of impossible obligations with conflicting obligations[53]

Weaker than Kant's Law is the claim that nothing *logically* impossible is obligatory. In the standard systems, this weaker claim can be expressed as a rule:

(8.34) $$\text{If } \vdash \neg p \text{ then } \vdash \neg \mathbf{O}p,$$

(8.34) is a derived rule in any VW system. For suppose $\vdash \neg p$. Then by PC, $\vdash p \leftrightarrow \bot$, and then from OD, $\vdash \neg\mathbf{O}\bot$, we get $\vdash \neg\mathbf{Op}$. This in itself is not necessarily a problem for standard systems. For claiming that, say, I'm obligated to both be home and not be home because I for some reason promised you just this logically impossible thing is less convincing then saying that two separate promises might yield two distinct obligations to keep conflicting appointments, each executable, though not jointly. For it might be maintained that the concept of obligation is such that obligation claims to do the logically impossible are *logically* self-defeating. In either event, the standard systems are better insulated from this sort of objection than from the objection that conflicts of obligation are possible. Assuming we are dealing with a system that has RED (and thus any of the standard systems), the rule above is equivalent to OD, and so we can put the point more simply by saying that the following is a thesis of all standard systems, and is plausible:

(OD) $\mathbf{O}\neg\bot$.[54]

However all this suggests that there is a clear difference between conflicting obligations and a singular obligation regarding something logically impossible, and this in turn means there is a serious expressive limit in the standard systems. For within them, from a conflict of obligations such as $\mathbf{O}p \& \mathbf{O}\neg p$, we can derive an obligatory logical contradiction, $\mathbf{O}\bot$, and vice versa. Any

[53]Cf. [Chellas, 1974]. See also [Chellas, 1980; Schotch and Jennings, 1981].

[54]However [Da Costa and Carnielli, 1986] develops a paraconsistent deontic logic that would at least allow for some contradictory obligations.

logic with both KD and RMD (and thus RED), the following is a theorem:

(8.35) $(\mathbf{O}p \& \mathbf{O}\neg p) \leftrightarrow \mathbf{O}\bot$ (Collapse)

For by KD, $\vdash \mathbf{O}(\neg p \to \bot) \to (\mathbf{O}\neg p \to \mathbf{O}\bot)$, and by RED, $\vdash \mathbf{O}(\neg p \to \bot) \leftrightarrow \mathbf{O}p$, so $\vdash \mathbf{O}p \to (\mathbf{O}\neg p \to \mathbf{O}\bot)$. For the right to left direction, by RMD, $\vdash \mathbf{O}\bot \to \mathbf{O}p$ and $\vdash \mathbf{O}\bot \to \mathbf{O}\neg p$, since $\vdash \bot \to q$, for any q. Yet 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 DD from OD is now quite routine in conflict-allowing deontic logics, and OD is assumed in most deontic logics. [Chellas, 1974; Chellas, 1980; Schotch and Jennings, 1981] contain early discussions of this expressive limit and advocate different non-normal modal logics to handle this problem (among others).[55]

The limit assumption dilemma[56]

Recall our sketch of an alternative ordering semantics for SDL, and the use there of the Limit Assumption:

(LA) For each u, there is v such that for any $w, v \geq_u w$

Although the limit assumption has often been assumed true in the use of ordering semantics for deontic logic, it is a controversial assumption to make, especially as a matter of logic. It seems at least conceivable that there might be a scenario in which the ordering of worlds in the purview of some world u has no upper limit 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 can be set anywhere from 0 to 1, where both 0 and 1 yield disaster, but all the numbers, n, between 0 and 1 not only avoid disaster, but yield increasingly more overall good as n grows (cf. [Merrill, 1978]). If we countenance the possibility of such scenarios, and thus drop (LA) in our semantics, we must alter the standard truth clause for **O** via bests:

(COB) $\mathbf{O}p$ is true at u iff p holds in all the u-best worlds.

For in models with no u-best worlds, nothing is obligatory and everything becomes permissible by this clause, but this seems too strong a result. For example, in the deontic dial case, it seems clearly obligatory to not turn the dial to 1 or 0, however otherwise perplexing the scenario is.

[55] Chellas employs minimal models (or neighborhood semantics) and Schotch and Jennings generalize Kripke models using multiple accessibility relations.

[56] Cf. [Lewis, 1973].

[Lewis, 1973; Lewis, 1974] argued that the limit assumption's use in deontic logics (and for counterfactual logics) is unjustified, and thus that our clauses for deontic (and counterfactual) operators must be adjusted. Most logicians accept this in principle, and often employ a more complex clause such as:

(COB') $\mathbf{O}p$ is true at u iff p is true from some point on up in the u-ordered worlds.[57]

In models where (LA) holds, (COB') is provably equivalent to the simpler (COB) which assumes there are u-best worlds, so the new clause is conservative. But in models where (LA) fails, it will not follow from (COB') that nothing is obligatory and everything is permissible. For example, in models intended to represent Barley's deontic dial scenario, the new clause does get the result that it is obligatory to not turn the dial to 0 and obligatory to not turn the dial to 1, and thus the dial endpoints are impermissible.

However, the new clause also has some perplexing results. For example, in the case of the dial, it seems for each setting, n, between and including 0 but before 1, it ought to be set *past* n, since there will be an accessible u-world where all such worlds ranked as high as u are worlds where the dial was turned past n. Not only does this raise the question about where to turn it positively (specificity), but the truth of the set consisting of all recommended settings of the form *the dial is turned past* n, for $0 < n \leq 1$, entails the truth of "the dial is turned to 1", which is something you ought *not* do by the same clause. Thus the set of things you ought to do is an inconsistent set. Although no syntactic conflict will show up in the system (no finite set of formulas of the form $\mathbf{O}p_1, ..., \mathbf{O}p_k$ are all true in the model while $p_1 \& \ldots \& p_k$ is false in all models, we nonetheless can have an infinite set of obligations which cannot be jointly fulfilled. This seems to be a case of *conflicting obligations* - a situation where one's obligations are not jointly realizable, and thus belongs under that heading.[58] So although we are given some clear directions - don't place the dial at either extreme and do place it somewhere in between 0 and 1, it is also the case that anywhere in between that we do place it, we will be wrong for not having placed it closer to 1 than that. [Merrill, 1978] argues that a related problem (called "The Confinement Problem" in [McNamara, 1995]) is a problem for Lewis semantics, but [Lenk, 1978] argues that it is a problem for utilitarianism, not deontic logic. However, [Fehige, 1994] suggests that logicians must still

[57] [Lewis, 1973; Lewis, 1974].

[58] As it turns out, SDL is also characterized by COB and COB', whether or not the limit assumption holds, as long as the preference relation is a total preordering of W (connected, reflexive and transitive). Things however become more complex once connectivity is dropped. See [Goble, 2003].

make choices here and that, ironically, there is no clear best choice for them either: "...When the best options are lacking, then so are flawless accounts of the lack"[Fehige, 1994, p. 42]. Endorsing (LA) as a matter of logic seems unjustified, yet accommodating its denial seems to lead to its own challenges and puzzles.

Plato's dilemma - deontic defeasibility[59]

Suppose I promised to meet you for dinner, and thereby incurred an obligation to do so, but suppose also that as I am about to leave, my child begins to have an asthmatic attack, and it is clear that he needs me to rush him to the hospital. It would then seem that both of these claims are true:

(8.36) I'm obligated to meet you for dinner (now)
(8.37) I'm obligated to rush my child to the hospital (now).

Here we seem to have an indirect non-explicit conflict of obligations, where it is not practically possible to satisfy both obligations, but neither thing required is logically inconsistent with the other. Note that unlike the earlier case of two conflicting appointment obligations that appeared to be on a par for all we said, here we are immediately inclined to judge that the obligation to help my child *overrides* my obligation to meet you for dinner—the former takes clear *precedence over the latter*. Shifting focus to the weaker obligation in (8.36), we might say that it is *defeated* by that of (8.37). Furthermore, except in extra-ordinary circumstances, we would also judge that no other obligation overrides the obligation to help my child, and thus that this obligation is an *all things considered obligation (or a undefeated obligation)*, unlike the obligation to meet for dinner. Lastly, we would ordinarily think that my obligation to rescue my child is not only not overridden by any other obligation, but that it strictly overrides any obligations I might have that conflict with it, and thus that it is not only not defeated or overridden, but is *overridingly obligatory* or a *strict obligation*. We are also prone to speak more abstractly and say that there is an *exception* here to the *general obligation* to keep one's appointments (or promises), for the circumstances are extenuating.

It should be noted there is no uniform use of terms such as "dilemma" in deontic logic (or ethical theory); some define a "dilemma" as an *unresolvable conflict*: a conflict of obligations where neither of the conflicting obligations defeats the other (cf. [Sinnott-Armstrong, 1988]). On this use of "dilemma", although the earlier case with two appointments, as well as the above case, can be construed as *conflicts* of obligation, the current example

[59] Cf. [Lemmon, 1962a]. Plato's dilemma involves returning a weapon when the owner is in a rage and intending to (unjustly) kill someone with it. Our interpretation of the issues raised by Plato's dilemma is a bit different than Lemmon's.

is not construed as a *dilemma*, since one of the two obligations does defeat and override the other. Sometimes "predicament" is also used, again either for a conflict or for a conflict that is unresolvable.

We have already indicated that standard systems have no mechanism for representing a conflict of obligations as a logical possibility. So clearly the issues here go beyond their capacity, but it is also important to note that once we set out to represent conflicts of obligation, there is the further issue of representing the logic of relationships between conflicting obligations and statuses of obligations deriving from these relationships, such as one overriding another, one defeating another, one being undefeated by any others and so being an all things considered obligation, one being a general one (e.g. it is obligatory to keep one's promises) that holds by default but not unexceptionally, etc. The issue of conflicting obligations *of different weight*[60] and the *defeasability* of obligations by other obligations (or even by circumstances–an obligation to meet a friend for dinner who now himself can't make it because ill) clearly requires much more than just having a logic that allows for conflicts, although that is a necessary condition.

There have been a variety of approaches to this domain and the associated issues, with considerable intensification in the 1990s. [von Wright, 1968] *informally* proposed minimization of evil as a natural tool for resolving conflicts of obligation, thereby suggesting the aptness of reliance on an ordering relation. [Alchourrón and Makinson, 1981] gives an early formal system for conflict resolution using 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 (see also [Horty, 1997; Horty, 2003]). [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 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 [Bulygin, 1992; Moreau, 1996], both of which attempt to represent W. D. Ross-like notions of prima facia obligation, etc. Earlier related works of interest that were ahead of the curve

[60] Cf. [Brown, 1996b].

on some aspects of defeasibility are the influential conceptual framework of [Chisholm, 1964], and in a similar but more formal vein, that of [Loewer and Belzer, 1983; Belzer, 1986; Loewer and Belzer, 1991]. Also notable are the discussions of defeasibility and conditionality in [Alchourrón, 1993; Alchourrón, 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. [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.[61] See also the chapter by Goble in this volume, which covers many of the topics in this section, including those in the following supplement.

Supplement to 8.4 on some challenges for conflict tolerant logics

A minimal conflict tolerant logic

Two early conflict-tolerant logics are [van Fraassen, 1973; Chellas, 1974]. ([Lewis, 1974] contains a note suggesting Chellas may have circulated his system in 1970, but of course this may be true of van Fraassen as well for all we know. We list both, since so proximate.)

Suppose we want a conflict tolerant logic (and we are not yet concerned with representing the further notions associated with defeat among conflicting obligations)? What should we keep from the standard systems and what should we reject? Answers to this question are not easy nor uncontroversial. Here, we cannot possibly consider all the options, much less their comparative merits, so instead we will consider one natural and simple pathway to an elementary conflict tolerant logic, one much like the earliest ones to emerge, and then describe some of the challenges it faces. We hope this will give the reader some flavor for issues and complications that arise in developing conflict tolerant logics. (See [Goble, 2009] for a more elaborate discussion of various issues and options, as well as his chapter in this handbook.)

[61] [Alchourrón, 1996; Asher and Bonevac, 1996; Prakken, 1996] are all found in *Studia Logica* 57.1, 1996. [Nute, 1997] is dedicated to defeasibility in deontic logic (both CTDs and defeasible deontic consequence) and is an excellent single source with articles by many recent key players.

If we will allow conflicts, then minimally, we want a conjunction like the following to be consistent of course:

(EC) $\mathbf{O}p \,\&\, \mathbf{O}\neg p$ (Explicit conflict)

(We ignore here that a general definition of conflicts of obligation does not say anything about explicit conflicts merely that there is a set of two or more obligations not all jointly realizable.)

What else? First and foremost, we want to make sure that unlike with all the standard systems, we cannot generate *Deontic Explosion*—the indiscriminate derivability of all formulas in the face of conflicts:

(DEX) $\mathbf{O}p \,\&\, \mathbf{O}\neg p \to \mathbf{O}q$ (Deontic explosion)

This would render any system that recognized the possibility of conflicts utterly useless in the face of one. Suppose also that we want the logic to reject the possibility of obligatory contradictions—that is, suppose we want to retain (6.9)/(OD) as a thesis:

(OD) $\neg\mathbf{O}\bot$

This is not unreasonable, since we might say that the prospect of an obligatory logical contradiction is immediately logically self-defeating. Assuming so, we will then have to avoid our previously mentioned (8.35) "Collapse":

(OD) $\mathbf{O}p \,\&\, \mathbf{O}\neg p \leftrightarrow \mathbf{O}\bot$ (Collapse)

For otherwise this will immediately rule out conflicts when conjoined with (OD); and even if we wanted to allow for some special cases where there were obligatory contradictions, we don't want every conflict to generate one. Collapse seems undesirable for any reasonable conflict tolerant logic. What about the consequence principle:

(RMD) If $\vdash p \to q$ then $\vdash \mathbf{O}p \to \mathbf{O}q$?

Well, we have seen that there are certainly considerations that can be raised against this principle, especially without any restriction (e.g. so that even tautologies are obligatory if anything is); but on the other hand, it is certainly attractive to be able to draw conclusions about what else is obligatory from some of the logical consequences of things that are obligatory. So let's here retain RMD in our exploration of conflict tolerance. However, let's set RND, that any theorem is obligatory, aside as a distraction. What of (6.2) - Aggregation?

(6.2) $\mathbf{O}p \,\&\, \mathbf{O}q \rightarrow \mathbf{O}(p\&q)$ (**O**-aggregation)

Clearly this must be rejected given what we have said already. For consider our explicit conflict above, $\mathbf{O}p\&\mathbf{O}\neg p$. From any such explicit conflict, if we granted (6.2), an obligatory contradiction, $\mathbf{O}(p\&\neg p)$ would be derivable, and we have already said this is not plausible; furthermore, from RMD, since $(p\&\neg p) \rightarrow q$, for any q, we would get deontic explosion.

What of SDL's KD:

(KD) $\mathbf{O}(p \rightarrow q) \rightarrow (\mathbf{O}p \rightarrow \mathbf{O}q)$?

We must reject this as well, for even without RM, and with the very plausible and widely endorsed

(RED) If $\vdash p \leftrightarrow q$ then $\vdash \mathbf{O}p \leftrightarrow \mathbf{O}q$?

as the only deontic principle, KD would generate the left to right portion of Collapse, $\mathbf{O}p\&\mathbf{O}\neg p \rightarrow \mathbf{O}\bot$, which is surely unacceptable in its own right. (From RED, $\mathbf{O}\neg p \leftrightarrow \mathbf{O}(p \rightarrow \bot)$, then along with K, we would get $\mathbf{O}\neg p \rightarrow \mathbf{O}(p \rightarrow \bot)$, viz. $(\mathbf{O}\neg p\&\mathbf{O}p) \rightarrow \bot$ [McNamara, 2004].) Let's take just what we have so far to be our minimal conflict tolerant logic, the logic EMD [Chellas, 1974; Chellas, 1980]. Assume it has the same language as SDL, but it has just one deontic axiom, OD, and one deontic rule, RMD, along with the power of truth-functional logic for the language:

(PL) Propositional logic

(OD) $\neg\mathbf{O}\bot$

(RMD) If $\vdash p \rightarrow q$ then $\vdash \mathbf{O}p \rightarrow \mathbf{O}q$

(Regarding "EMD": "M" is Chellas label for the thesis $\Box(p\&q) \rightarrow (\Box p\&\Box q)$, "E" for the rule "If $\vdash p \leftrightarrow q$ then $\vdash \Box p \leftrightarrow \Box q$", and "D" for $\neg\Box\bot$. Given truth-functional logic, RMD is interderivable with the combination of E and M [Chellas, 1980], so EM plus D is equivalent to RM plus D above.)

A simple semantics for this can be easily given in terms of what are called "minimal models" or "neighborhood semantics" [Chellas,

The original footnote read (but using "!" and "~" for the falsum, and negation here):
"From RED, **O**~p ↔ **O**(p → !), then along with K, we'd get **O**~p → (**O**p → **O!**), i.e., (**O**p & **O**~p) → **O!** [McNamara 2004].
The "O" operators in front of the two falsums were somehow dropped, and the parenthesis is in the wrong place giving the antecedent of K not the needed consequent. This probably happened somehow in moving all the footnotes into a parenthetical part of the boxed off text. Sorry I missed this in final copy.

1980] (also called Montague-Scott semantics [Sloman, 1970; Montague, 1970]). As in Kripke models, we have a set of worlds, W, and a valuation function, v, assigning sets of worlds to the atomic sentences, but now we replace the Kripke accessibility relation with a function that maps worlds to sets of sets of worlds (often thought of as sets of propositions).

$$(OB) \qquad OB : W \to \mathrm{Pow}(\mathrm{Pow}(W)) \quad \text{i.e. } OB(u) \subseteq \mathrm{Pow}(W)$$

So the value of the obligation function for any given world, u, is a set of subsets of W—the propositions the obligation function assigns to u as mandated. The truth conditions (relative to a model) for obligation statements are as follows:

$$(\text{CO'}) \qquad u \models \mathbf{O}p \text{ iff } ||p|| \in OB(u)$$

It is obligatory that p at u (in a model) iff the proposition expressed by p (the set of p-worlds) is among those mandated by OB for u. We then validate OD by stipulating that the empty set (representing the contradictory proposition true at no worlds) is never mandated at a world:

$$(OB\text{-D}) \qquad \emptyset \notin OB(u), \text{ for every } u$$

Notice that in some models, $OB(u)$ will not contain W, so $\mathbf{O}\top$ is not validated. Similarly, nothing has been said to indicate that if $OB(u)$ contains a set α and a set β, that it thereby must contain their ~~union~~ "intersection", so Aggregation is invalid, as desired.

With this as our brief framing, we now turn to some puzzles/challenges such a conflict tolerant logic faces.

Van Fraassen's general challenge

In [van Fraassen, 1973], van Fraassen published perhaps the first logical-semantic framework for conflicts of obligation. (Compare also [Chellas, 1974] for a different early conflict tolerant logical and semantic framework.) His approach is to layer it on top of a framework for imperatives, the interesting details of which will not concern us here. (See [Horty, 1994; Horty, 1997] for very influential expositions and explorations of van Fraassen's framework in tandem with developing new conflict tolerant systems inspired by developments in AI. Horty's work has helped to bring the importance of van Fraassen's challenge to the attention of deontic logicians.) The first key point is that he has an initial conflict tolerant system for **O** much like the one above. Van Fraassen then gives a simple example of a prima facia desirable inference that the simple conflict tolerant logic he has endorsed cannot ratify. The example (attributed by van Fraassen to Robert Stalnaker), in its **O**-version (ignoring the underlying imperatives) is :

(vFI) You ought to either honor your father or your mother. You ought not honor your father. So you ought to honor your mother.

Formalized, the inference pattern looks like this:

(vFI') $\mathbf{O}(f \vee m), \mathbf{O}\neg f, \text{ so } \mathbf{O}m$

With **O**-aggregation it is easy to generate the conclusion using RMD: from $\mathbf{O}(f \vee m)$ and $\mathbf{O}\neg f$, we get $\mathbf{O}((f \vee m)\&\neg f)$, and from the latter along with RMD (which generates RED), we easily get $\mathbf{O}m$. However, we have cast **O**-aggregation aside above so as to avoid being able to derive an obligatory contradiction from every explicit conflict of obligations. So we cannot reason like this here. Van Fraassen raises a technical question, explored by Horty in the aforementioned papers, which we also pass over here, and asks a more general question, that Horty also articulates more fully, and we will call it "van Fraassen's challenge":

(vFC) Having accepted the possibility of conflicting obligations, how do we develop a conflict tolerant logic that avoids the two extremes of a logic so anemic that there is virtually no conclusions at all we can draw from joint premises (that don't follow from each premise alone), and a logic that is so strong that it generate deontic explosion (or an equally unacceptable variant thereof)?

(van Fraassen also offers the first proposed solution to this until-recently neglected problem that he identified, which we set aside for the moment other than to say vaguely that it is a sort of two-level generalized consistent aggregation approach. See [van Fraassen, 1973] and/or the aforementioned references to Horty for details.)

Van der Torre's van Fraassen-inspired puzzle

We provide a reconstruction based on [van der Torre, 1997].

Recall that above we pointed out that we needed to reject KD in a conflict tolerant context because it would otherwise generate this part of deontic collapse: $(\mathbf{O}p\&\mathbf{O}q) \to \mathbf{O}\bot$. But now notice that (vFI') above is really a barely disguised instance of KD: $\mathbf{O}(\neg f \to m) \to (\mathbf{O}\neg f \to \mathbf{O}m)$. So the *pattern* of inference (vFI'), despite any initially plausible ring, is unacceptable on reflection, and then so is (vFI), since *logically* invalid—not valid in virtue of its form. What makes the pattern sound plausible, as with the instance (vFI) itself, is that we naturally think of the wffs, $(f \vee m)$ and $\neg m$ *as mutually consistent* [McNamara, 2004]. This suggests endorsing the following principle modestly restricting **O**-aggregation as a natural solution to van Fraassen's challenge above:

"Oq" should be O~p

(CA) If $\nvdash p \to \neg q$ then $\mathbf{O}p\&\mathbf{O}q \to \mathbf{O}(p\&q)$ (Consistent aggregation)

missing turnstile before "**O**"

[van der Torre, 1997] attributes "consistent aggregation" to van Fraassen, but it appears on reflection (and in conversation with van der Torre) that this was more likely adopted from [Horty, 1994], who uses "consistent agglomeration" there, attributing the notion to [Brink, 1994].

[Brink, 1994] mentions this principle and gives it a qualified endorsement, and it was endorsed in an earlier draft of [McNamara, 2004] for DEON 2002. It is certainly a natural first amendment to consider in developing a conflict tolerant logic, and plausible at first blush. After all, the main reason for rejecting aggregation is the possibility of aggregating incompatible obligations, so if we know that two obligatory things are mutually compatible, what reason can there be to not go ahead and aggregate them. Van der Torre provides a decisive answer in the context we are exploring, for he shows us that however plausible this may sound, as long as we have (CA) and (RMD), we will quite easily generate, a trivial variant of the very result we said had to be avoided first and foremost in a conflict tolerant logic: Deontic Explosion. Here is the trivial and surely unacceptable variant of (DEX):

(DEX') $\mathbf{O}p\&\mathbf{O}\neg p \to \mathbf{O}q$ for any q such that $\nvdash \neg q$

In other words, if there is any conflict of obligation, every non-contradiction will be obligatory.

Now call this *van der Torre's thesis*:

(vdTT) From (RMD) and (CA), (DEX') follows

Suppose we have an explicit conflict, $\mathbf{O}p\&\mathbf{O}\neg p$, and suppose we have some consistent q. Then, either $\neg p \,\&\, q$ or $p \,\&\, q$ is consistent. First suppose $\neg p \,\&\, q$ is consistent. Then its equivalent, $(p \vee q) \,\&\, \neg p$ is consistent. So by (CA), we have $\mathbf{O}(p \vee q) \,\&\, \mathbf{O}\neg p \rightarrow \mathbf{O}((p \vee q) \,\&\, \neg p)$. Now, from $\mathbf{O}p$, by (RMD) we get $\mathbf{O}(p \vee q)$, and along with $\mathbf{O}\neg p$, we get $\mathbf{O}((p \vee q) \,\&\, \neg p)$, and then from this by (RMD) again, $\mathbf{O}q$ follows. Secondly, suppose $p \,\&\, q$ is consistent. Then by precisely parallel reasoning, we get $\mathbf{O}q$. So either way, $\mathbf{O}q$ follows.

So, here we have a prima facia well-motivated restriction on O-aggregation that along with RMD generates a version of explosion no more palatable than the original (especially considering that we have already ruled on the only exception to (DEX')'s explosive scope: $\neg\mathbf{O}\bot$ and its equivalents, so that, in effect, we get that everything that is not already logically ruled out as impossibly obligatory becomes obligatory in the face of any conflict).

So meeting *van Fraassen's challenge* is not as easy as it might at first seem. We are thus left puzzling about what form of aggregation specifically, if any, can be endorsed in allowing us to meet van Fraassen's challenge, given that this natural one fails; or must we whittle away instead at RMD, or follow yet some other path to meet the challenge?

Van der Torre, and others, have referred to this puzzle as "van Fraassen's puzzle", but this appears to be a misnomer. Although no doubt derived by van der Torre's reflection on van Fraassen's rich, compact, and sometimes cryptic remarks at the end of [van Fraassen, 1973], there is no mention of the deontic explosion problem that van der Torre articulates (although explosion is mentioned much earlier by van Fraassen in the article), nor is there more than, at best, a suggestion of the simple consistent aggregation principle above, and no mention of it. However, it is noteworthy that van Fraassen offers a solution to what we have called here "van Fraassen's challenge" that certainly involves a semantic version of a different restricted aggregation principle, and so it is certainly possible that he entertained the simpler consistent aggregation principle.

The idle aggregation puzzle (van Fraassen-Hansen)

Let us now point out that the problem may not be so readily solved merely by restricting RMD either, suggesting the challenge and puzzle of how to solve it is robust. For even if RMD is too strong, we must surely adopt some principles governing practical reasoning allowing us to reason about consequences of what is obligatory for us. Consider the following example adapted from one communicated by Jörg Hansen in 2002:

p: Jones keeps an appointment this morning in New York.
p': Jones travels to New York this morning.
q: Jones keeps an appointment in London this afternoon.
q': Jones travels to London this morning.

([Horty, 2003] provides a similar example where there is an obligation to both attend an event and pre-notify, and likewise for another event, in a different place at the same time. Here we have a conflict between two obligatory conjunction, but from RMD, it follows that each notificational conjunct is obligatory, and as these are mutually consistent, by consistent aggregation, their conjunction would be obligatory too.)

Imagine, not implausibly, that p practically necessitates p', and q practically necessitates q', and add that given the times and distances, Jones is unable to keep both appointments. Nonetheless it might be that traveling to both places this morning is open to Jones, for example, by driving to JFK airport in New York early this morning (long before his New York meeting) and flying directly from there to London. It seems implausible to conclude that Jones is obligated to both travel to New York and travel to London, which in turn is only achievable through the mad frenzied dash just sketched. The travel obligations derive exclusively from the appointment obligations that conflict and can't be jointly realized, so it is not plausible that a singular conjunctive obligation to travel to both New York and to London follows. Now notice that this problem is not easily solved by just saying RMD is implausible, for p' and q' above are actions in my power that are *practical prerequisites* of p and q respectively. Reasonable restrictions of RMD need to allow us to make inferences like these from premises about practical prerequisite of obligations we have, since this seems to be nothing short of central to practical reason itself.

So it appears that it is more plausible to see the key issue as being about how to properly restrict aggregation beyond (CA), even if there are other independent reasons to want to restrict RMD. A faithful representation of obligations must allow us to derive from our obligations further obligations to realize their practical prerequisites; but at the same time, it seems we must disallow the derivation of idle conjunctive obligations. (This presentation draws on [McNamara, 2004].)

For more on the last three interrelated [van Fraassen, 1973]-inspired issues which have only received their due attention more recently, see for example, [Horty, 1994; van der Torre, 1997; van der Torre and Tan, 2000; Horty, 2003; Hruschka, 2004; McNamara, 2004; Goble, 2005;

Goble, 2009], as well as Goble's chapter in this volume.

(The above problem was conveyed by Jörg Hansen to McNamara at DEON 2002 and discussed in McNamara 2004; it appears that van Fraassen recognized a very similar problem of potential over-generation of conclusions that served as the inspiration for Hansen's articulation of the problem as it applied to McNamara's earlier system. We benefited from discussion with Lou Goble, Jorg Hansen, John Horty, and Leon van der Torre on these van Fraassen-inspired puzzles.)

8.5 Puzzles centering around deontic conditionals

The paradox of derived obligations[62]

Consider this statement:

(8.38) Bob's promising to meet you commits him to meeting you

Two very natural attempted representations of claims like that in (8.38) in standard systems were suggested:

(8.39) $\mathbf{O}(p \to m)$ (cf. [von Wright, 1951])

(8.40) $p \to \mathbf{O}m$ (cf. [Prior, 1955])[63]

Consider (8.39) first, which was how von Wright first interpreted statements like (8.38). The following are theorems by RMD, and thus in all standard systems: $\mathbf{O}\neg r \to \mathbf{O}(r \to s)$ and $\mathbf{O}s \to \mathbf{O}(r \to s)$. Thus if the logic of (8.38) were correctly realized in SDL by representing it as (8.39), it would follow that anything impermissible commits us to everything, and that for anything obligatory, everything commits us to it. Does (8.40) fair better? No. The following are simply tautologies: $\neg r \to (r \to \mathbf{O}s)$ and $\mathbf{O}s \to (r \to \mathbf{O}s)$. So if the logic of (8.38) were correctly realized in standard systems by representing it as (8.40), it would follow that, anything false would commit us to anything whatsoever (e.g. since I did not promise you to meet, it would follow that my promising to meet you commits me to not meeting you) and again, for anything obligatory, everything commits us to it (e.g. if I'm obligated to phone you, then my living in a time with no phones commits me to phoning you). As Prior notes, the problems are reminiscent of the paradoxes of strict implication (reading (8.39) and material implication (reading (8.40), respectively. This raises the question: is it simply beyond the resources of standard systems to properly represent notions of commitment or conditional obligations? The next paradox convinced logicians that indeed it is.

[62] Cf. [Prior, 1954].

[63] There, Prior credits G. E. Hughes for this alternative symbolization.

Chisholm's contrary-to-duty paradox[64]

Here is Chisholm's famous quartet:[65]

(8.41) It ought to be that Jones goes to the assistance of his neighbors
(8.42) It ought to be that if Jones goes to the assistance of his neighbors, then he tells them he is coming
(8.43) If Jones doesn't go to the assistance of his neighbors, then he ought not tell them he is coming
(8.44) Jones does not go to their assistance

It is widely thought that (8.41)-(8.44) constitute a *mutually consistent* and *logically independent* set of sentences: all four might be true at once, and none is a deductive consequence of the others. We will treat these as central desiderata: a correct representation of the logic of (8.41)-(8.44) must be consistent with these two constraints.[66] The problem, in a nutshell and from a high altitude, is that it is not at all as easy as it might seem to faithfully represent scenarios like those in the quartet and still meet the above two constraints, and it proved to be a real shortcoming of the standard systems as people quickly came to realize that they could not be represented there. On the positive side, it has been a catalyst for distinctive and expansive work in deontic logic. It is perhaps the most important puzzle in the history of 20$^{\text{th}}$ century deontic logic, and so we will spend some more time on it. Here we will briefly characterize the problem for the standard systems. The supplement to 8.5 provides more detail about some attempted solutions to this puzzle.

First we provide some terminology that has emerged regarding the taxonomy of the puzzle ingredients. Since (8.41) tells us what Jones ought to do unconditionally, it is a *primary obligation*, the only one in this context.[67] (8.43) is a *contrary-to-duty obligation* (a CTD), an instance of the type of claim after which the puzzle is named. In the context of (8.41), (8.43) says what Jones ought to do on the condition that he *violates* (or at least does not fulfill) his primary obligation in (8.41). In contrast, (8.42) says what

[64] Cf. [Chisholm, 1963].

[65] We give Chisholm's original example since the piece is so seminal in deontic logic, but this also means that (8.42) and (8.43) have a different form with *ought* having a different surface scope. The difference between (8.42) and (8.43) in Chisholm's original formulation is largely seen as a distracting artifact, and in many presentations of the puzzle, (8.42) is adjusted to follow the form of (8.43), which form is thought to be the more challenging one to represent, and the most central to contrary-to-duty conditionals.

[66] Others will be alluded to in passing below. Carmo and Jones 2002 [Carmo and Jones, 2002] argues for no less than seven desiderata for any solution.

[67] We will follow tradition here in sloughing over the differences between an obligation and what ought to be and what one ought to do, since we believe the puzzle reappears as we shift across these three distinct notions.

else Jones ought to do on the condition that Jones *fulfills* his primary obligation, and so (8.42) could be called a "compliant-with-duty obligation" in this context. Finally, (8.44) is just a factual claim, which conjoined with (8.41), implies that Jones violates (or at least does not fulfill) his primary obligation. The relativization to context for both labels is crucial, since in another context, (8.42) could be a contrary-to-duty instead (e.g. advanced notice is not important, and Jones agreed with a friend that they would both surprise the neighbors with their help) and (8.42) might be compliant with duty in the right context (e.g. due to character defects of the neighbors in question, if they knew Jones was coming to their assistance they would not make the vital efforts now essential to Jones not being too late to help at all). Thus the taxonomy involves tracking the relationships between the normative and factual claims across a piece of discourse.

How might we represent the Chisholm quarter in the standard systems?[68] The most natural first stab appears to be:

(8.41')	$\mathbf{O}g$
(8.42')	$\mathbf{O}(g \to t)$
(8.43')	$\neg g \to \mathbf{O}\neg t$
(8.44')	$\neg g$

Here we read (8.42) with **O** having wide scope, and (8.41) with **O** having narrow scope, following the surface of the original. Chisholm noted that by principle KD, we get $\mathbf{O}g \to \mathbf{O}t$ from (8.42'), and then $\mathbf{O}t$ from (8.41') by MP. In turn, from (8.43') and (8.44'), we get $\mathbf{O}\neg t$ by MP alone. But the combination of $\mathbf{O}t$ and $\mathbf{O}\neg t$ contradicts DD ($\mathbf{O}t \to \mathbf{O}\neg t$). Thus (8.41')-(8.44') is an inconsistent quartet in any of the standard systems, unlike the original whose logical form they are alleged to represent. Various other representations in the standard systems have similar shortcomings. For example, we might try reading the second and third premises uniformly either on the model of (8.42') or on the model of (8.43'). After all, it is not clear what motivates framing them differently in the original quartet (8.42)-(8.43), and this oddity is often dropped in contemporary discussions. If we use

(8.43") $\mathbf{O}(\neg g \to \neg t)$

instead of (8.43'), we lose independence since (8.43") is derivable from (8.41') in the standard systems by RMD (as we saw in discussing the *paradox*

[68] We use primes here to make it easier to keep track of the various correlated statements in the different quartets.

of derived obligation). Likewise if we use

(8.42") $\quad g \to \mathbf{O}\neg t$

instead of (8.42'), independence is again lost, for this is derivable from (8.44') by PC alone. So again, we end up with unfaithful representations of the logic of the original quartet.

We can sum up the problems with these three ways to interpret (8.41)-(8.44) in the standard systems in the following table:

	First	Second	Third
	$\mathbf{O}g$	$\mathbf{O}g$	$\mathbf{O}g$
	$\mathbf{O}(g \to t)$	$\mathbf{O}(g \to t)$	$g \to \mathbf{O}t$
	$\neg g \to \mathbf{O}\neg t$	$\mathbf{O}(\neg g \to \neg t)$	$\neg g \to \mathbf{O}\neg t$
	$\neg g$	$\neg g$	$\neg g$
Problem	$\therefore \bot$	$\mathbf{O}g \vdash \mathbf{O}(\neg g \to \neg t)$	$g \vdash g \to \mathbf{O}\neg t$

Our first attempt yields a contradiction—the set is rendered inconsistent. The second and third attempts lose independence, since one of the four follows from the others in each case (in fact form just one premise, as indicated). The only remaining apparent combination would replace (8.42') with (8.42") $g \to \mathbf{O}\neg t$ and (8.43') with (8.43") $\mathbf{O}(\neg g \to \neg t)$, but that just combines the loss of independence in the second and third attempts, so it is rarely mentioned.

Given the extreme simplicity of the semantics offered for the standard systems, it is not surprising that it is not capable of representing complex normative situations in a satisfactory way, and it is not difficult to see why Chisholm's example in particular cannot be represented in a satisfactory way. As was observed above, the semantics of SDL is based on a division of worlds (situations) into normatively acceptable and unacceptable ones, with the O-sentences defined so that they describe how things are in the deontically acceptable situations. But CTD sentences do not describe how things are in deontically acceptable worlds; instead they tell what is to be done or how things ought to be under deontically *unacceptable* conditions, and specific ones at that (e.g. in worlds where Jones does not go to the assistance of his neighbors). For that, we need a way to pick out not only worlds where things have gone wrong, but where things have gone wrong in some specific way indicated by the clause of the CDT that expresses the violation of the primary obligation; and then we must go on to select propositions that are *relatively-acceptable*—acceptable relative to the assumption that the worlds will be those unacceptable ones where the specific violation conditions hold.[69] So it is no wonder that SDL and kin cannot express CTDs.

One difference with these conditional obligations or ought-statements that was noted early on was that they are *defeasible* in the sense that they do not satisfy the principle of *strengthening the antecedent*, which of course does hold for material implication:

(SA) $(p \to q) \to ((p\&r) \to q)$ (Strengthening the Antecedent)

The corresponding thesis for the deontic conditionals in focus in Chisholm's puzzle was virtually universally recognized to be invalid:

(8.45) $\mathbf{O}(q/p) \to \mathbf{O}(q/(p\&r))$

Even if Jones ought to tell if he will help, it will not follow that it is also true that he ought to tell if he will help and telling will cause some disaster.

It is now virtually universally acknowledged that the Chisholm Paradox shows that the sort of deontic conditional expressed in (8.43) above can't be faithfully represented in SDL, or even as a composite of some sort of a unary deontic operator and a *material* conditional. Here is one of the key places where deontic logicians are in full agreement.

By giving pride of place to contrary-to-duty requirements, the puzzle also brings into relief a crucial feature of (most if not all) normative requirements: their violability.[70] Since we are quite imperfect creatures, the possibility of violation is hardly idle. It is crucial for us to know not only what is to be done, but also what to do in turn when what ought to be done in the first place is not done. Consider the role of apologies in repairing the torn social fabric, or statements in contract law about what is owed in reparation if one party fails to provide what is owed in the primary clauses of the contract (e.g. amazon.com owes you a refund if the wrong item arrives). We would have nuclear meltdowns without emergency clauses about what a crew is to do at a nuclear power plant when the crew has failed to do something required and things have started to go wrong. When things

[69] Cf. [Lewis, 1974; Jones and Pörn, 1985], although note that Lewis, followed by many others, concludes that more is needed, namely a preference ordering of worlds allowing (for a selection of the "best of the bad" compatible with any particular given violation) whereas Jones and Porn argue that this is not necessary, a representation of what holds in the non-acceptable worlds (along with what holds in the acceptable worlds, and thus in both classes) will suffice.

[70] [Jones, 1990] argues for the importance of violability to legal knowledge representation and consequentially for the importance of deontic logic for such knowledge representation, stressing particularly the issue of representing contrary-to-duty contexts; See also [Jones and Sergot, 1993] which argues that violability (and the possibility thereof) is what gives deontic logic much of its importance.

do go wrong, thankfully, they often have not gone as wrong as they can go; we can take adjusting actions mitigating the harm. Damage control is vital, and in turn, the ability to reason accordingly is vital. The puzzle also obviously places deontic *conditional* constructions at center stage, inviting us to ponder: What is the correct logic behind reasoning with deontic conditionals generally, and particularly, in contexts where the conditionals appear to tell us what we are obligated to do if we violate some other obligation? Lastly, the puzzle also raises the question of how to track fulfillment, nonfulfillment, and particularly violation of obligations, along with conditional obligations, across a set of statements (a piece of extended discourse), as indicated by our initial context-relative taxonomy of the Chisholm quartet.

Among the things contested regarding this paradox are whether or not what is needed is some special *primitive dyadic deontic conditional* operator or just some *non-material* conditional conjoined to a monadic deontic operator, as well as the more general question of what essentially is needed to faithfully represent the logic of deontic conditionals like those in the puzzle. For the interested reader, we explore these further in the *Appendix on Chisholm's puzzle & conditional norms.* Next, we introduce a closely related puzzle.

Forrester's paradox[71]

Here is a version very close to the original:[72]

(8.46) Smith ought not kill Jones
(8.47) If Smith will <u>k</u>ill Jones, then Smith ought to kill Jones <u>g</u>ently
(8.48) Smith will kill Jones

As with the kindred Chisholm puzzle, this triplet appears to express a mutually consistent set of claims, with each claim independent of the remainder.

Here is a natural way to symbolize (8.46)-(8.48) in our standard systems:

(8.46') $\mathbf{O}\neg k$

(8.47') $k \to \mathbf{O}g$

[71] Cf. [Forrester, 1984]. This is also called the "gentle murder paradox".

[72] As the title, "gentle murder, or the adverbial samaritan" indicates, Forrester introduced the puzzle as "the most powerful version yet" of the *good samaritan puzzle*, one intended to rule out prior scope solutions targeting the original, and as might then be expected, he opts to drop RMD. However, Forrester's puzzle was cast in terms of deontic conditionals much like those above (with (8.47) above as a key auxiliary premise, and its wide scope analogue as main premise), and it is thus often construed instead as a variant of *Chisholm's paradox*, pointing again to the challenge of modeling deontic conditionals that seem to be telling us what to do if wrong will be done. We thus place it here, although RMD will also be invoked in showing that SDL and ilk are inadequate. It has features of both puzzles.

(8.48') k

Now from (8.47') and (8.48') by MP, it follows that

(8.49) $\mathbf{O}q$

But the following seems to be a natural language logical truth:

(8.50) Smith kills Jones gently only if Smith kills Jones

Assuming so, let's imagine an augmented standard deontic system where (8.50) is a formal logical truth, which we will symbolize here for simplicity as

(8.51) $g \to k$

From this, by RMD, it will follow in such an augmented system that

(8.52) $\mathbf{O}g \to \mathbf{O}k$

and so by MP again, we get

(8.53) $\mathbf{O}k$

But now with (8.53) added to (8.46'), we have conflicting obligations, in contradiction with DD of the standard systems. So it looks like the standard systems cannot coherently represent Forrester's triplet.

Although (8.46)-(8.48) seem like they could all be true, it seems difficult to swallow that (8.47) and (8.48) entail that Smith is obligated simpliciter to kill Jones gently (e.g. it seems we can consistently add that Jones has not the slightest justification for harming Smith at all).[73] On the other hand, if we side with those favoring interpreting the deontic conditional in (8.47) as non-material but still subject to a version of modus ponens, we must then accept the inference from (8.47) and (8.48) to the informal analog of $\mathbf{O}g$.[74] So it appears that unless we reject some principle of SDL such as RMD, we will still generate a contradiction.[75] Note also that simply opting

[73]In the Appendix on Chisholm's paradox at the end of this chapter, the reasoning here will be discussed in in the context of motivating the position of the friends of "*deontic detachment*".

[74]This corresponds to the position favored by friends of "*factual detachment*", also discussed in the Appendix on Chisholm's paradox.

[75]Some have suggested this is still a problem stemming from scope difficulties, others have argued that the problem is that RMD 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 RMD. We have listed this

to reject DD as a response is not a natural avenue, since it does not seem very plausible to say that the problem is simply that we have a conflict of obligations, the obligation to kill Jones and the obligation to not kill Jones (or the obligation to kill Jones gently and to not kill Jones gently). If this is right, then it supports the contention that what is at issue ultimately in this puzzle, and with the Chisholm puzzle, pertains to these particularly troubling CTDs.

There is a vast literature on this subject, and here we can only sketch a fragment. Some of this material is briefly discussed in the Appendix to 8.5. In a brief note responding to Prior's paradox of derived obligation, [von Wright, 1956] introduced the often-used undefined dyadic operator approach to the syntax of conditional obligations, $O(q/p)$. See also [von Wright, 1964; von Wright, 1965] for further developments of his approach, and an explicit recognition of the importance of Chisholm's paradox. Von Wright's approach is primarily syntactic and axiomatic. [Danielsson, 1968; Hansson, 1971], followed a bit later by [Lewis, 1973; Lewis, 1974], provide formal semantics for conditional obligation, using preference orderings of worlds to model CTDs construed via an undefined dyadic operator of von Wright's sort. [Åqvist, 2002; Åqvist, 1987] provide systematic presentations of this sort of approach, as well as analogue systems for deontic conditionals in the Leibniz-Kangerian-Andersonian vein (and discussions of other paradoxes). [van Fraassen, 1972; Loewer and Belzer, 1983; Jones and Pörn, 1985] give important and influential alternative models for CTDs, each offering some interesting variants of the former more standard picture. [al Hibri, 1978] contains an early important survey of a number of these approaches to CTDs (and other puzzles), as well as a defense and development of her own system. In addition to [Lewis, 1973; Lewis, 1974], an important recent presentation and development of the metatheory of standard and near-standard monadic and dyadic deontic logics via classic and near-classic ordering structures is [Goble, 2003]. [Moore, 1973; Chellas, 1974; Chellas, 1980; Goble, 1990a] offer influential alternative approaches that do not use an undefined dyadic deontic operator. Instead they opt for representing deontic conditionals using a non-material

---

puzzle here rather than under the *good samaritan puzzle* (and thus under puzzles associated with RMD) 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 RMD, we would appear obligated to kill her, which has no plausibility by anyone's lights, and thus the puzzle calls for rejecting RMD. However, this would still include a stance on contrary-to-duty conditionals and detachment.

conditional,$\Rightarrow$, along with a unary deontic operator to generate a genuine compound sentence form, $p \Rightarrow \mathbf{O}q$. [Dayton, 1981] contains an important early critical discussion of this sort of approach. [Tomberlin, 1981] is a very influential informal discussion of various approaches to deontic conditionals. [Bonevac, 1998] is a recent argument against the dyadic approach to conditional obligation, suggesting that defeasible reasoning techniques developed in AI (see [Brewka, 1989]) can handle the problems with CTDs. In contrast, [Smith, 1993; Smith, 1994] stress the difference between violability and defeasibility, and the relevance of the former rather than the latter to CTDs. [Åqvist and Hoepelman, 1981; Vorobej, 1982; Loewer and Belzer, 1983] are early approaches to solving the puzzle (or versions thereof) by incorporating temporal notions. [Jones, 1990], as well as [Prakken and Sergot, 1994; Prakken and Sergot, 1996], are influential for their arguments that temporal notions are not essential to the Chisholm's paradox, so the solution cannot lie there. [Castañeda, 1981] argued that by distinguishing between propositions and "practitions" (roughly actions), most puzzles for deontic logic could be solved, including Chisholm's paradox; [Meyer, 1988] takes a similar approach but employing techniques from dynamic logic to represent actions, their combinations, and deontic notions. [Prakken and Sergot, 1996] is influential for arguing that action is inessential to the Chisholm paradox, so that the solution cannot lie there. For some work on CTDs in a branching time framework see [Herrestad, 1996; Horty, 2001; Bartha, 1999] and Bartha's chapter 11 in [Belnap *et al.*, 2001]. [Carmo and Jones, 2002] is an important recent handbook chapter reviewing various approaches to deontic conditionals in detail, as well as proposing and defending a solution of their own. [Nute, 1997] is a collection dedicated to defeasible deontic logic with a number of essays on Chisholm's puzzle (see especially [van der Torre and Tan, 1997; Prakken and Sergot, 1997]).

We should also mention the important and influential topic of *counts as conditionals*, which we regret not being able to discuss but briefly here. In short, a conditional is introduced to represent the idea of one proposition's realization counting as or constituting another's realization. For example, as mentioned in Section 8.1 and earlier sections, a performative act may constitute the realizing conditions of some other act. For example, in the right institutional settings, my raising my hand counts as my voting on the measure, the proposition that "You are hereby married", uttered by a magistrate counts as a truth-maker for the proposition that you are married, and a delegated person in a business signing an agreement for an order counts as a truth-maker for the business itself being obligated to meet the terms of the agreement. Since such constructions can be regimented into relations between propositions as we did in the middle example just above, they can be treated as a form of conditional, and logics devised accordingly. Such logics can then be integrated with various agential notions, to in turn represent various institutional transactional phenomena like delegation, authorization, and the trigger of changes in the normative position of various agents in institutions, and even of the institutions themselves, as in our last example. The locus classicus on this topic is [Jones and Sergot, 1996]. See also the chapter in this volume by Jones and Grossi covering this topic.

Let us also note here the not altogether unrelated accumulating work of Lindahl and Odelstad on formal representations of the role of "intermediate concepts" in law. If you are accused of a crime such a burglary, a variety of legally (and stipulatively) defined concepts will be invoked, such as "forced entry", "property", "theft", "person", etc. Although these terms are familiar, appropriated for a legal system, they are defined, and do not always track their normal use precisely. For example judges and lawyers need to be familiar with the exact legal definitions of these terms in order to be competent in adjudicating a legal accusation. In turn, such terms are often defined explicitly (eventually) in terms of extra-legal or "natural" terms that are not encoded in any legal definitions, such as "entering", "object", "transport", "human being", "building", etc. These serve as the "grounds" for the applicability of the intermediate legal concept. The intermediate concepts in turn might be associated with various normative consequences such as sanctions, loss of rights, etc. There is a sense in which, roughly, the stipulatively defined concepts function as intermediaries between the extra-legal grounds they are defined in terms of and the higher level legal-normative consequences that they are linked to (e.g. that burglary is punishable by imprisonment for up to ten years, that the convicted burglar can be held accountable for damages with loss of some of his property for reparations, etc.). See [Lindahl and Odelstad, 2000] for a concise overview, as well as the chapter by Lindahl and Odelstad in this volume for a more comprehensive account.

8.6 Some further normative expressive inadequacies of the standard systems

We have already noted the apparent expressive inadequacies in standard systems regarding Chisholm's paradox and deontic conditionals. In this section we turn instead to some monadic normative notions that appear to be inexpressible in the languages for VW, SDL and LKA_1 and LKA_2. In a number of cases, it appears that these notions were at least tacitly targeted for representation in standard systems although not actually expressible in them.

Urmson's puzzle - Indifference versus optionality[76]

Consider:

(8.68) It is optional that Jones helps Smith, but not a matter of indifference

We routinely assume that optional matters are not thereby matters of indifference. Yet deontic logicians and ethicists routinely read the condition "$\neg\mathbf{O}p \,\&\, \neg\mathbf{O}\neg p$", as "It is *indifferent* that p" ($\mathbf{I}\mathrm{N}p$) rather than as "It is optional that p" ($\mathbf{O}\mathrm{P}p$).[77] But then it would seem to follow trivially that $(\neg\mathbf{O}p \,\&\, \neg\mathbf{O}\neg p) \rightarrow \mathbf{I}\mathrm{N}p$: if anything is neither obligatory nor prohibited then it is indifferent, that is, neither obligatory nor prohibited. So in the standard systems, the best we can do in symbolizing (8.68) is by way of a tautological contradiction:

(8.68') $(\neg\mathbf{O}p \,\&\, \neg\mathbf{O}\neg p) \,\&\, \neg(\neg\mathbf{O}p \,\&\, \neg\mathbf{O}\neg p)$

Many alternative actions, including heroic ones, are neither obligatory, prohibited, nor matters of indifference. Urmson implores us to not conflate these two concepts and thereby indirectly rule out many cases of moral heroism that are often morally exemplary and optional, and to instead develop logical schemes that ratify the following constraint (Urmson's Constraint):[78]

(UC) $\mathbf{I}\mathrm{N}p \rightarrow \mathbf{O}\mathrm{P}p, \text{but not } \mathbf{O}\mathrm{P}p \rightarrow \mathbf{I}\mathrm{N}p$

But the standard systems can only represent optionality at best; they lack the expressive resources to carve up the optional zone into the *indifferent* and the *optional but non-indifferent*. Yet indifference was tacitly an early target for representation in as much as it was thought that this was aptly represented in the standard systems.

The problem of action beyond the call of duty[79]

Some alternatives are beyond the call of duty (**B**C) or supererogatory (e.g. volunteering to take on a challenging project for your department having already "paid your dues"). The standard systems have no resources to represent this notion, since they can say nothing more fine-grained about

[76]See [Urmson, 1958].

[77]Beginning with [von Wright, 1951], and recurring pervasively. Note that because we will begin to discuss distinct notions often conflated with one another in ethical theory and deontic logic, we will employ two letter abbreviations for operators to represent additional concepts more transparently.

[78]See [McNamara, 1996a] for further discussion.

[79]See [Urmson, 1958]. It may be that supererogation and action beyond the call are subtly distinct [McNamara, 2011a; McNamara, 2011b]. We slough over this issue here.

them than that they are optional (neither obligatory nor impermissible), but although the optionality of what is beyond the call.

(BC-OP) $\mathbf{B}\mathrm{C}p \rightarrow \mathbf{O}\mathrm{P}p$

is desirable, its converse, $\mathbf{O}\mathrm{P}p \rightarrow \mathbf{B}\mathrm{C}p$, surely is not. As Urmson's Constraint above (UC) indicates, matters of indifference are optional.[80] So representations of this asymmetry in standard systems will end up being trivial or incoherent (.e.g. $\vdash \mathbf{O}\mathrm{P}p \rightarrow \mathbf{O}\mathrm{P}p$, $\nvdash \mathbf{O}\mathrm{P}p \rightarrow \mathbf{O}\mathrm{P}p$, respectively). Also, note that $\mathbf{B}\mathrm{C}p \rightarrow \neg\mathbf{I}\mathrm{N}p$ is also desirable, so to represent this notion fully, we need distinct representations of optionality and indifference.

The must versus ought dilemma[81]

Consider:

(8.69) Although you can skip this meeting, you ought to attend (and you must attend either this one or the next).

We routinely make such distinctions in situations where no conflicting obligations are present. (8.69) appears to properly entail that it is optional that you attend - that you *can* attend and that you *can* also not attend, although preferable to attend. In context, the latter two uses of "can" paradigmatically express permissibility. Yet "ought" is routinely the reading authors give for deontic necessity in deontic logic and in ethical theory, and "permissibility" is routinely presented as its dual. But this suggests the following symbolization of the first two conjuncts of (8.69):

(8.70) $\mathbf{P}\mathrm{E}\neg p \,\&\, \mathbf{O}\mathrm{B}p$

But (8.70) is equivalent to $\neg\mathbf{O}\mathrm{B}p\&\mathbf{O}\mathrm{B}p$ (by RED and TDS), which contradicts DD. It is much more plausible to construe the "can" of permissibility ($\mathbf{P}$E) as the dual of "must" ($\mathbf{M}$U) than as the dual of "ought" ($\mathbf{O}$U). It appears that $\mathbf{M}\mathrm{U}p \rightarrow \mathbf{O}\mathrm{U}p$ is desirable but not $\mathbf{O}\mathrm{U}p \rightarrow \mathbf{M}\mathrm{U}p$, and $\mathbf{M}\mathrm{U}p \rightarrow \neg\mathbf{P}\mathrm{E}\neg p$ is desirable, but not $\mathbf{O}\mathrm{U}p \rightarrow \neg\mathbf{P}\mathrm{E}\neg p$. This suggests a dilemma for the standard systems (and most work in deontic logic):

> Either permissibility is represented in the standard systems, but "ought" is inexpressible in it (despite the widespread assumption otherwise) or "ought" is represented in those systems, but permissibility and impermissibility are inexpressible in them despite the widespread assumption otherwise. You can't have it both ways at the same time.

[80][Chisholm, 1963] is a landmark here, and as we've seen, these issues clearly overlap with things Meinong was beginning to explore much earlier (as Chisholm notes as well).

[81]From [McNamara, 1990].

That the dual of permissibility is expressed by "ought" is a problematic but pervasive "Bipartisan Presupposition" in both deontic logic and ethical theory.[82] [McNamara, 1996c], and in more detail [McNamara, 1990], argues that there is also very strong pressure from the use of the modal auxiliaries "must" and "ought" in non-deontic contexts to a) distinguish them, b) to take the former to properly entail the latter, and c) to not posit that there is an ambiguity in "ought" for the purpose of saying that there is one sense in which it means the same as "must".[83]

The least you can do problem[84]

Consider

(8.71) You ought to have been on time; the least you could have done was called, and you didn't do even that

Although there has been lots of attention to constructions like that in the first clause of (8.71), the construction in the second clause has been almost totally ignored in deontic logic and ethical theory. Yet it is familiar and widely used, and it appears to entail that there was some minimally acceptable alternative that included calling (to say you would be late), whereas the first clause suggest there was also an acceptable but preferred alternative, which was to just be on time (and so not call to say you would be late). The third clause suggests the criticism that even though you had permissible options of different ranks, you did less than even the minimally acceptable option, and thus you comported yourself impermissibly. Presumably, we want things such as $\mathbf{O}\mathrm{B}p \to \mathbf{L}\mathrm{E}p$, but not $\mathbf{L}\mathrm{E}p \to \mathbf{O}\mathrm{B}p$, $\mathbf{L}\mathrm{E}p \to \mathbf{P}\mathrm{E}p$, etc. This rich notion of what is minimally acceptable among the permissible options is plainly not expressible in the standard systems.

As the reader might surmise, the set of notions above appear to be part of an underexplored interlocking family of normative notions.[85]

---

[82] It is a merit of [Jones and Pörn, 1986] that it recognizes there is a clear difference between deontic uses of "ought" and "must", and it provides an early attempt to distinguish the two (in a logical system with a formal semantics). However, "must" ends up being modeled as something akin to *practical* necessity in their system (whatever obtains in all scenarios - permissible or not) rather than deontic necessity (whatever holds in all *permissible* scenarios). For a cumulative case argument that "must" is the dual of permissibility, not "ought", and thus that it is "must", not "ought", that tracks the traditional concern in ethical theory and deontic logic with what is permissible, impermissible, and obligatory, see [McNamara, 1990; McNamara, 1996c].

[83] Note that c) is of limited interest as a reply anyway, since even granting it for sake of argument, surely the important task then is to analyze the sense of "ought" that is *not* equivalent to the sense of "must", and to integrate these two in one logic.

[84] See [McNamara, 1990].

[85] For attempts to begin to address the last four problems by the simplifying ploy of extending familiar standard or near standard systems, see [McNamara, 1996c; Mares and

The challenge of normative gaps[86]

As we say in discussing Jørgensen above, in some normative contexts, explicit permissions, prohibitions and requirements are issued by some normative authority. But then we must allow for a type of gap: cases where p is neither *explicitly* obligatory, impermissible, nor permissible in a normative system because it is not *explicitly* commanded, prohibited or permitted. Yet in all the standard systems,

$$\text{(6.12)} \qquad \mathbf{O}p \vee (\mathbf{P}p \,\&\, \mathbf{P}\neg p) \vee \mathbf{F}p \qquad \text{(Exhaustion)}$$

is a thesis. In fact, it is nearly *tautological* given the TDS (Traditional Definitional Scheme). For in primitive notation, it amounts to just this: $\mathbf{O}p \vee (\neg\mathbf{O}\neg p \& \neg\mathbf{O}\neg\neg p) \vee \mathbf{O}\neg p$, and so only RED is needed to replace $\neg\mathbf{O}\neg\neg p$ with $\neg\mathbf{O}p$ with the result saying essentially "it is this, that, or neither" (where "neither" always gets cashed out as entailing permissibility). So any system endorsing just the language and definitional scheme of SDL that includes truth-functional logic and just the one deontic rule, RED, will ratify Exhaustion. Then, for any proposition p, p will either have the status of being impermissible or obligatory or permissible (since optional). This precludes **P**, **O**, and **F** in the standard systems from being normative notions that allow for gaps.

The problem of the directionality of obligations[87]

In the standard systems, the bearers of obligations, if any are intended, go unrepresented. Furthermore very often obligations are obligations to a specific person or institution: *Jones* is obligated *to Smith* that p be the case. For example, I am obligated to you (by contract say) to paint your fence, and you in turn (upon completion) are obligated to me to pay me. Directed obligations are also related to rights. If I am obligated to you to paint your fence, then you have a claim on me to do so, and if you are obligated to pay me for the paint job, then (upon completion), I have a claim on you to pay me. Notice also that, typically, no one else, other than perhaps representatives of the law, have any claim on me to paint your fence, or on you to pay me for having done so. [Herrestad and Krogh, 1995] argues that an explicit representation of the directionality of obligations (and prohibitions and permissions) is not only needed to represent one important aspect of many (if not most) obligations, but that it also facilitates a better representation of relations between claims and obligations in the tradition

McNamara, 1997; McNamara, 1999], which provide a cumulative case argument for the broad outlines of a solution to these representational problems.

[86] Cf. [von Wright, 1968]. See also [Alchourrón and Bulygin, 1971] for another key early source.

[87] Cf. [Herrestad and Krogh, 1995].

of [Hohfeld, 1919], and in the logical work on normative positions inspired by Hohfeld's work, beginning with Kanger's seminal work (e.g. [Kanger, 1957]). See also Sergot's chapter in this volume on normative positions.

8.7 A problem calling for attention to action and agency in deontic contexts

The jurisdictional problem and the need for the representation of agency [88]

Consider the following claims:

(8.72) Jeeves is obligated to not bring it about that Bertie's teeth are brushed

(8.73) Jeeves is obligated to not bring it about that Bertie's teeth are not brushed

There are limits to Jeeves duties and his rights as valet for Bertie, and Bertie's teeth-brushing is out of his jurisdiction - he is required to not interfere in that area, and thus to neither bring it about that Bertie's teeth are brushed (e.g. by forcibly doing so), nor to bring it about that they are not brushed (e.g. by pinching Bertie's tooth brush). Can we represent these in the standard systems? Many, following [von Wright, 1964; von Wright, 1965], have freely read "**O**" as "Smith is obligated to bring it about that__" (for some mock agent, Smith) or as "Smith is obligated to see to it that __". Even ignoring the complex integration of agential and deontic notions in this reading, and letting advocates of this reading have it, can we represent (8.72) and (8.73)? It does not seem we can do any better than this:

(8.74) $\neg \mathbf{O}p$

(8.75) $\neg \mathbf{O}\neg p$

Together, (8.74) and (8.75) simply provide the conditions for optionality - Jeeves is not obligated to bring it about that Bertie's teeth are brushed and not obligated to bring it about that they are not brushed. But that is not what (8.72) and (8.73) are saying. They decidedly entail that it is *not* an optional matter whether or not Jeeves brings it about that Bertie's teeth are brushed. Put another way, (8.74) and (8.75) could be true even if (8.72) and (8.73) are both false. For example, that is the situation if Jeeves is permitted to assure that Bertie's teeth are brushed and permitted to

[88] The first reference we have found coming close to explicitly formulating this problem is [Lindahl, 1977, p. 94], where the "none of your business" terminology is invoked, but it was recognized by Kanger, since essentially presupposed in his analysis of rights-related notions in his seminal [Kanger, 1957]. See also [von Wright, 1968].

assure that they are not brushed, since Bertie has had recent gum surgery, and Jeeves gets to decide what is apt. Shifting the negation signs inward, thereby creating conflicting obligations, is no help either.

If we want to represent scenarios like the one above, it seems we must allow for the negations to be able to operate on the agency itself, so reading "**O**" with agency built in will not serve. To adequately represent these situations, we need to represent agency separately from the deontic operators, and then explore their interactions. Action, agency, and deontic operators will be taken up in the next section.

9 Actions and agency in deontic logic

Philosophers have made a distinction between two kinds of ought, the ought-to-be (Seinsollen) and the ought-to-do (Tunsollen) [Castañeda, 1970], and it has been suggested that since the deontic operators of SDL are propositional operators, the standard deontic logic and the extensions and revisions discussed above should be regarded as theories of the ought-to-be rather than theories of the ought-to-do. However, as was observed earlier, the propositions in question may be action propositions, propositions to the effect that an agent does something or that an action of a certain kind is performed or not performed (omitted). In this approach, deontic concepts are not applied to generic actions or act-types, as in von Wright's 1951 system (see Section 5), but to propositions about individual actions. Another alternative that resides between the impersonal reading and the personal and agential reading of **O** is reading **O** as specifying that it is obligatory *for Smith* that it be the case that p (personal, but not agential), and then the agential form is a special case - it is obligatory for Smith that it be the case that Smith brings it about that q. (See [Krogh and Herrestad, 1996; McNamara, 2004].)

G. H. von Wright has observed that actions (or acts) usually involve changes in the world:

> "Many acts may ... be described as the bringing about or effecting ('at will') of a change. To act is, in a sense, to interfere with the 'course of nature'."[von Wright, 1963, p. 36]

Von Wright analyzes actions in terms of three world-states or occasions: (i) the *initial state* or *origin* which the agent changes or which would have changed if the agent had not been active (had not interfered with the course of nature), (ii) the *end-state* or the *result-state* which results from the action [von Wright, 1963, p. 28], and (iii) the *counter-state* which would have resulted from the initial state without the agent's interference, in other words, the state which would have resulted from the agent's passivity. The counter-state is needed for expressing the "counterfactual element" [von Wright, 1963, pp. 43-4] or *sine qua non* condition of action (cf. [Hart and Honoré, 1959, pp. 103-122]).

The characterization of acts by means of three states or occasions makes it possible to distinguish $2^3 = 8$ different modes of action with respect to a single state of affairs p. These modes of action may be defined as follows: Let $W = \{u, v, w...\}$ be a set of possible world-states or occasions, and let us assume that the agent can be either active or passive in a given state. Let d be a function which assigns to each $u \in W$ a state which results from the agent's activity at u, and let e be a function which assigns to each $u \in W$ the corresponding counter-state. The truth-value of p at u is denoted by '$V(p, u)$', and as usual, '$V(p, u) = 1$' (where '1' means the value *true*) will be abbreviated '$u \models p$'. For example, if $u \models \neg p$, $d(u) \models p$ and $e(u) \models \neg p$, we can say that the agent brings it about that p or produces the state of affairs that p. In this case p becomes true as a result of the agent's action: without the agent's action it would have remained false that p. The falsity of p at the initial state and at the counter-state constitute an opportunity for the agent to bring it about that p. On the other hand, if p is false at $d(u)$ (the end-state) under otherwise similar circumstances, we can say that the agent omits to bring it about that p. In this way we obtain the action possibilities presented in Table 2. Here '**BA**' abbreviates 'bring it about that' and '**SS**' stands for 'sustain (the state that)'. For the sake of brevity,

	u	$d(u)$	$e(u)$	Mode of action	Rendering
Act 1	$\neg p$	p	$\neg p$	Bringing it about that p	**BA**p
Act 2	p	p	$\neg p$	Sustaining the state that p	**SS**p
Act 3	$\neg p$	$\neg p$	$\neg p$	Letting it remain the case that $\neg p$	**omBA**p
Act 4	p	$\neg p$	$\neg p$	Letting it become the case that $\neg p$	**omSS**p
Act 5	p	$\neg p$	p	Bringing it about that $\neg p$	**BA**$\neg p$
Act 6	$\neg p$	$\neg p$	p	Sustaining the state that $\neg p$	**SS**$\neg p$
Act 7	p	p	p	Letting it remain the case that p	**omBA**$\neg p$
Act 8	$\neg p$	p	p	Letting it become the case that p	**omSS**$\neg p$

Table 2: The main action-types according to von Wright

we shall use below the expression 'the agent brings about (or produces) p', instead of saying that an agent brings it about that p. In this simplified terminology, we can say that Act1 is an act of *producing* p, Act2 is an act of *sustaining* (*preserving*) p, Act5 is an act of *destroying* p, and A6 is an act of *preventing* p.

If $V(p, d(u)) \neq V(p, e(u))$, the truth-value of p depends on the agent's activity; in this case the agent is active with respect to p; otherwise the agent may be said to be passive with respect to p. The action-types in which $V(p, d(u)) = V(p, e(u))$ are omissions (abbreviated '**om**'). As was observed earlier, an omission in the proper sense should be distinguished from the non-performance of an act: an agent can omit an act only in a situation in which he has an opportunity to perform the act in question; thus an omission entails non-performance, but not conversely. If $V(p, d(u)) \neq V(p, e(u))$ and $V(p, d(u)) \neq V(p, u)$, the action in question is a productive or a destructive act, but if $V(p, d(u)) \neq V(p, e(u))$ and $V(p, d(u)) = V(p, u)$, the action is an act of sustaining or preserving some state of affairs.

In von Wright's analysis, actions are characterized by means of propositional expressions which refer to the result-state, the initial state, and the counter-state of the action, and the propositions which describe the states are transformed into action propositions by means of the praxeological operators **BA**, **SS**, and **om**. In many recent systems of the ought-to-do, action propositions are formed in this way, and simple action descriptions are given the form '$\mathbf{Do}(a, p)$', where '**Do**' is a modal (praxeological) operator for action or agency and p is a propositional expression. The **Do**-operator is usually read 'a brings it about that' or 'a sees to it that'. This analysis of action sentences goes back to the 11th century philosopher St. Anselm, who investigated the meaning of the Latin phrases 'facere esse' (to bring it about that'), 'facere non esse', 'non facere esse', and 'non facere non esse'. (Cf. [Henry, 1967, pp. 123-9] and [Segerberg, 1992, p. 348-51].) The logical relations among these concepts can be represented as a square analogous to the square of modalities, and this suggests that they can be treated for logical purposes as modal concepts, as praxeological modalities.

[Kanger, 2001] has presented an analysis of action and agency in terms of the concept of *seeing to it that* p. He regarded a statement of the form 'a sees to it that p', '$\mathbf{Do}(a, p)$', as a conjunction

(CDo) $\mathbf{Do}(a, p) \leftrightarrow \mathbf{Ds}(a, p) \,\&\, \mathbf{Dn}(a, p)$,

where '**Ds**' may be said to represent the *sufficient condition aspect* of agency and '**Dn**' stands for the *necessary condition aspect* of agency. (Cf. [Hilpinen, 1974, p. 170] Kanger reads '$\mathbf{Ds}(a, p)$' as

> p is necessary for something a does

and '$\mathbf{Dn}(a, p)$' as

> p is sufficient for something a does

These readings are equivalent to

(9.1) $\mathbf{Ds}(a, p)$: Something a does is sufficient for p

and

(9.2) $\mathbf{Dn}(a,p)$: Something a does is necessary for p

Kanger interpreted the agency operators **Ds** and **Dn** in terms of two alternativeness relations on possible "universes" [Kanger, 2001, p. 152, 159]. In a simplified form, Kanger's conditions may be expressed as follows:

(CDoS) $u \models \mathbf{Ds}(a,p)$ iff $w \models p$ for every w such that $S^s(u,w)$

and

(CDoN) $u \models \mathbf{Dn}(a,p)$ iff $w \models \neg p$ for every w such that $S^n(u,w)$

Kanger's phrase "something a does" may be paraphrased as 'some action D performed by a"; thus (9.1) and (9.2) may be rewritten as

(9.3) $\mathbf{Ds}(a,p)$: Some action D performed by a is sufficient for p

and

(9.4) $\mathbf{Dn}(a,p)$: Some action D performed by a is necessary for p

where 'D' is a variable for action types. Reformulated in this way, it is clear that strictly speaking, Kanger's theory is not an analysis of action, but an analysis of the concept of seeing to it that. The concept of action ("something a does") is part of *analysans*.

The praxeological action (or agency) operator is sometimes read 'see to it that', sometimes 'bring it about that'. In so far as these expressions are used in ordinary discourse, they do not have the same meaning. An agent a can "see to that p" either by bringing it about that p or sustaining the state that p, that is, by making sure that p is not "destroyed"; thus seeing to it that p does not entail bringing it about that p. According to this interpretation, $\mathbf{Do}(a,p)$ is equivalent to $\mathbf{BA}p \vee \mathbf{SS}p$ in von Wright's schema. The modality of the action does not depend on the initial state (situation), and we get the four action modalities distinguished by St. Anselm:

(9.5) (i) $\mathbf{Do}(a,p)$: a sees to it that p
(ii) $\neg\mathbf{Do}(a,p)$: a does not see to it that p
(iii) $\mathbf{Do}(a,\neg p)$: a sees to it that $\neg p$
(iv) $\neg\mathbf{Do}(a,\neg p)$: a does not see to it that $\neg p$

A common feature of von Wright's and Kanger's analyses is that both analyze action/agency in terms of two conditions. Kanger's first condition, the **Ds**-condition, may be termed the *positive* condition, and the second condition, the **Dn**-condition, may be termed the *negative* condition of agency.

(Cf. [Belnap, 1991, p.792]) The latter condition corresponds to von Wright's counterfactual condition of agency. It states that if the agent had not acted the way he did, p would not have been the case. Some philosophers have disagreed about the formulation of the negative condition. [Pörn, 1977] has argued that we should accept instead of Kanger's **Dn**-condition only a weaker negative requirement, viz. '$\neg\mathbf{Dn}(a, \neg p)$', abbreviated here '$\mathbf{Cn}(a, p)$':

(ACN) $u \models \mathbf{Cn}(a, p)$ iff $w \models \neg p$ for *some* w such that $S^n(u, w)$

This condition can be read: but for a's action it might not have been the case that p [Pörn, 1977, p. 7]; that is, it was not unavoidable for a that p. [Åqvist, 1974, p. 81] has accepted a similar weak form of the counterfactual condition. According to Pörn and Åqvist, the negative condition should be formulated as a might-conditional, not as a would-conditional. Other versions of the analysis of agency by means of a positive and a negative condition have been [Lindahl, 1977; Åqvist and Mullock, 1989], and Nuel Belnap, John Horty, Michael Perloff, and others. (For discussion of such approaches, see [Belnap *et al.*, 2001; Horty, 2001]; for different forms of the positive and the negative condition, see [Hilpinen, 1997, pp. 11-20].)

There is also a morally and legally relevant concept of bringing it about with a might-conditional as a positive condition and a would-conditional as a negative condition:

(9.6) $\mathbf{BA}^{\star}(a, p) \rightarrow \mathbf{Cs}(a, p) \,\&\, \mathbf{Dn}(a, p)$

where '$\mathbf{Cs}(a, p)$' means that something a does makes p possible or enables (contributes to) p. In cases of this kind, a's actions are a sine qua non-condition of p, and a may be regarded as a contributing agent of the state of affairs p, and held at be least partly responsible for it.

According to von Wright's, formulation of the counterfactual (*sine qua non*) aspect of action, the agent's "passivity" at any given world-state or occasion (situation) u would lead to a single world-state (counter-state) $e(u)$. The values of the functions d and e are assumed to be world-states or situations, not sets of world-states. This means that the counterfactuals underlying von Wright's analysis satisfy the principle of Conditional Excluded Middle:

(9.7) Either: if the agent had been passive, it would have been the case that q, or: if the agent had been passive, it would have been the case that not-q

more generally,

(CEM) $(p \Rightarrow q) \vee (p \Rightarrow \neg q)$, where $\Rightarrow$ is a sign for a counterfactual or subjunctive conditional. (CEM) dos not always hold because sometimes q might or might not be the case if it were the case that p. (Cf. [Lewis, 1973, p. 79]) Thus we should revise von Wright's analysis by assuming that the agent's passivity in a situation u might lead to various alternative world-states, depending on how u might change without the agent's interference, for example, as a result of the actions of other agents. This can be represented by means of a function which has as its value the set of those world-states which could result from the agent's passivity. Such a representation agrees with the analysis of counterfactuals based on set selection functions given above in the Appendix to 8.5. In the same way, an action whose initial state or origin is u is representable by a function which assigns to u the set of possible world-states which could result from the action.

Von Wright formulates the counterfactual condition in terms of the agent's passivity, or what might be called the *zero action*. Such an account is inapplicable to many action situations which do not include a clear alternative of passivity. If D is the action of bringing it about that p or seeing to it that p in a certain way, we may define the counterfactual aspect of D in terms of the omission of D, or not doing D, or doing something else instead of D. Von Wright's analysis can be enriched in the same way as Kanger's theory, by assuming that the agent can change the initial situation u in different ways by undertaking different actions or by performing some action in different ways, in other words, we may assume that the agent can perform in a given situation various actions $A_1, ..., A_n$, each of which is represented by means a function which assigns to each situation u the set of world-states to which the action might lead the agent from u. In this way von Wright's analysis, applied to the concept of seeing to it that, assumes the form

(9.8) $\mathbf{Do}(a, p)$ if and only if a performs some action D such that
(i) if a were to do D, p would be the case, and
(ii) if a did not do D, it would not be the case that p

According to (9.8), an agent a may be said to see to it that p if and only if p's being the case is counterfactually dependent on something a does.

According to von Wright, the truth-values of sentences, including those of action sentences, are relative to occasions or world-states [von Wright, 1963, p. 23]: occasions are the points of evaluation of sentences (or propositions). As we have seen, an action proposition involves three occasions, the initial state, the end-state, and a possible counter-state. Is an action sentence regarded as true or false in the initial state or in the end-state; in other words, on which occasion does the agent perform the action? This question is closely related to the question about the time of an action (cf. [Thomson, 1971]). In his [1983] paper von Wright argues that the sentence

(9.9) $\mathbf{BA}p \to p$

is not a logical truth on the ground that

> "[$\mathbf{BA}p \to p$] would say that if a state is produced on some occasion then it is (already) there on this occasion. But this is logically false." [von Wright, 1983, pp.195-6]

This suggests that if action sentences are evaluated with respect to occasions or world-states, we should regard the initial occasion as the point of evaluation. (If an agent brings it about that p, p is false on the initial occasion.) Thus we should define (for example) the truth of '$\mathbf{BA}p$' as follows:

(9.10) $u \models \mathbf{BA}p$ iff $u \models \neg p, d(u) \models p$ and $e(u) \models \neg p$

According to (9.10), sentence (9.9) is logically false, whereas

(9.11) $\mathbf{BA}p \to \neg p$

is logically true. (Cf. [Segerberg, 1992, p. 358])

Condition (9.10) is problematic if '$\mathbf{BA}p$' is read 'the agent brings it about that p', that is, if '$\mathbf{BA}p$' is regarded as a genuine action proposition which says that the agent *does* something. According to von Wright, an action involves changing a situation or a state in some respect or keeping it unchanged, and the state (or 'world') u is understood here as the situation which either is or is not changed by the agent's action. We cannot assume that '$\mathbf{BA}p$' is part of the description of the very situation which is changed (or kept unchanged) by that action. It is natural to say that the agent chooses to perform an action at the initial state u: u is the state from which the action 'originates', but the sentences '$\mathbf{BA}p$', '$\mathbf{SS}p$', '$\mathbf{omBA}p$' and '$\mathbf{omSS}p$' cannot be regarded as true or false at u if they are understood as genuine action sentences. It would be better to say that the agent does something *to* the initial state, that is, changes it or keeps it unchanged, than to say that the action is performed *at* the initial state. von Wright's view seems to be supported by Nero Wolfe, who has remarked:

> "The average murder, I would guess, consumes ten or fifteen seconds at the outside. In cases of slow poison and similar ingenuities death of course is lingering, but the act of murder is commonly quite brief." [Stout, 1980, p. 16]

According to Wolfe (and Donald Davidson, see [Davidson, 1980a]), a poisoner kills the victim, that is, brings it about that the victim is dead, in a situation in which the victim is not dead; the death may occur much later. However, according to Wolfe (and Davidson), the act of bringing about the death of the victim consists in pouring the poison in his drink, and the initial situation changed by that action is a situation in which the poison is still safely in the little bottle in the poisoner's hand. The act of pouring the poison cannot be said to be performed in such a situation.

Many other authors who have analyzed the concept of action as a praxeological modality have accepted the success principle analogous to (9.9),

(9.12) $\mathbf{Do}(a, p) \rightarrow p$

as a valid principle for the concept of seeing to it that p. For example, Brian Chellas, who uses '$\Delta_a p$' for 'a sees to it that p', says about (9.12):

> "This is perhaps the most minimal substantive axiom for Δ. One can see to it that such-and-such is, or be responsible for such-and such's being, the case only if such-and-such is the case." [Chellas, 1969, p. 66] (See also [Kanger, 2001, p. 149-150].)

Many subsequent theories of action and agency have followed Chellas's example in this respect. (See [Belnap, 1991; Belnap and Perloff, 1988; Belnap and Perloff, 1992; Elgesem, 1993; Elgesem, 1997; Sandu and Tuomela, 1996; Belnap *et al.*, 2001; Horty, 2001].) It is clear that one can be responsible only for what is in fact the case or what has actually happened, but it is not equally clear that one can "see to it that p" only if it is the case that p. This does not hold in von Wright's theory of action. It is misleading to say that one can see to it that p only if it is the case that p: as von Wright has pointed out, a person can bring it about that p only if it is *not* the case that p, and bringing it about that p may be a case of seeing to it that p. We can say, of course, that an agent *has seen* to it (or *has brought* it about) that p, and is held responsible for p, only if it is the case that p. Statements about (causal) responsibility are evaluated only at the end-states of actions. Thus we have to distinguish here between (present tense) action sentences and statements about agency. A person is an agent of a certain result only if he *has done* something which has caused (or will cause) the result.[89] (In Nero Wolfe's example of killing by poisoning, and in other similar cases in which the outcome can be known beforehand with certainty, we may hold

[89] This fits the intention, if not the reading, of what Belnap calls an "*achievement* stit" operator ("stit" for "sees to it that"): Smith achievement-sees to it that p just in case p *now* holds and *was* guaranteed by a *prior* choice of Smith's. Thus p must now hold for this compound sentence to be true, but as a result of some past action that was instrumental in p's now being the case (there is a negative might condition as well as a positive condition).

the agent (the poisoner) responsible for what will happen, even though a court of law would not find him guilty of murder before the victim is dead.)

[Segerberg, 1992, p. 373] has observed that Chellas's action semantics provides no picture of action itself and suggested that this failure may be related to the validity of the T-principle mentioned above. But von Wright's rejection of the T-principle of modal logic does not make his theory superior to Chellas's theory in this respect; on the contrary, as we have seen, Chellas's theory can be given a reasonable (re)interpretation as a theory of agency statements, but von Wright's choice of the initial states as the circumstances of evaluation of action sentences excludes such an interpretation. If Chellas's theory is understood in this way, the lack of a counterfactual condition seems to be a weakness, but such a condition can of course be added to his analysis. ([Hilpinen, 1997, p. 17].)

One potential source of confusion here is the possibility of understanding the expression 'possible world' in two different ways. It can mean either temporary world-state (a moment) or a world-history, that is, a sequence of world-states. In von Wright's approach, a possible world is understood in the former way; it is a possible state of the world at a given moment, a world-state. If events are regarded as changes (or world-state transformations) and an action is regarded as the bringing about of a change, we obviously cannot assume that action propositions are interpreted as sets of possible worlds: actions do not take place *within* possible worlds. On the other hand, if possible worlds are understood as histories or courses of events, we can say that an agent performs an action in a possible world.

Von Wright's analysis of action in terms of alternative successions of world-states suggests an integration of these two conceptions into a semantic representation based on a branching frame of moments (states of the world, situations) and transitions between moments, that is, a structure $(W, <)$, where the elements of W represent moments (situations, world-states), and $<$ is a treelike partial ordering such that for any u, v, and $w \in$ W, if $u < w$ and $v < w$, then either $u < v$ or $v < u$ or $u = v$. (Cf. [von Wright, 1968, pp. 38-57].) The moments $u \in W$ can be interpreted as possible choice situations or the initial world-states which the agent may change by his actions, and some of the successors of u in the ordering are the situations which may result from his action, that is, the possible end-states of the action. In this model, an action A can be represented by a set of ordered pairs (u, w), with u as the initial state and w as a possible end-state or result-state of A, in other words, actions are regarded as binary relations on W. (See [Åqvist, 1974, p. 77] and [Czelakowski, 1997, p. 50].) Von Wright suggests this model of action when he observes that when a state of affairs either begins (or ceases) to exist as a result of an agent's action, the "occasion" on which the action takes place should be regarded as consisting of two "phases", one in which the state of affairs is absent (present), and another phase in which the state of affairs is present (absent). ([von Wright, 1983, p. 174, pp. 195-6]; see also [von Wright, 1968, p. 65].) Many philosophers have characterized actions in ways which fit this model. For example, [Apostel, 1982, p. 104] has observed that "an action is a transformation of nature in order to realize a purpose", and in his "action-state semantics" for imperatives C. L. Hamblin has analyzed actions or deeds in terms successive world-states [Hamblin, 1987, pp. 137-166]. According to [Weinberger, 1985, p. 314], "an action is a transformation of states within the flow of time" involving a subject (an agent), who "has at his disposal a range for action, i.e., at least two states of affairs which are possible continuations of a given trajectory in the system of states."

This way of representing actions and world-states requires two kinds of predicates and propositional expressions, expressions which describe possible states of the world (for example, 'the door is closed'), and action terms (predicates) and propositions which describe the way in which an agent changes the world (for example, 'to open the door', 'Bertie opens the door'). The former are true or false at the states $u \in W$, and the latter characterize the transitions (u, w) in W. An action term becomes a propositional sign when it is completed by an indexical sign which indicates an agent (or agents). Let $p, r, s, ...$ be propositional symbols, and let F, G, H, $\ldots$ be action terms or action descriptions. Action terms can be simple or complex: the latter are formed from simple action terms by act-connectives, some of which are analogous to propositional connectives. For example, if F and G are action terms, the following expressions are also action terms:

(ActT1) $F + G:$ doing A or B

(ActT2) $F \wedge G:$ doing A and B together

$F+G$ represents a choice between the actions F and G. It is also convenient to have an expression for the omission of an act, in the sense of doing something instead of F:

(ActT3) $\sim F:$ omitting F

'$\sim F$' is applicable to all individual actions (world state transitions) which fail to exemplify F. Systems of dynamic deontic logic usually also contain act-connectives which have no counterparts in propositional logic, for example [Segerberg, 1990, pp. 205-6] and [Segerberg, 1992, p. 376]:

(ActT4) $F; G : F$ followed by G

(ActT5) $F^{\star}$: doing F a finite number of times

The ordered pairs of states assigned to an action sentence A may be called the possible *performances* of A. A world-state w is said to be possible relative to u or accessible from u if and only if it is possible for some action or sequence of actions to lead from u to w. Let us denote this accessibility relation by POS, and let POS/u be the set of transitions which originate from u, briefly expressed, 'u-transitions'. In the following, the expression 'c does A at u', where c is an agent, is used to refer to an action which has u as its initial state, that is, that is, a set of transitions from u to various possible outcome states. Normative concepts can be defined in this framework by dividing world state transitions into normatively acceptable (legal, permitted, right) and deontically unacceptable (illegal, forbidden, wrong) transitions (cf. [Stenius, 1982, pp. 270-1, 276-80]). Let LEG/u be the set of legal transitions which originate from u, and let ILL/u be set of illegal u-transitions. The following conditions express the assumptions that any possible transition from u is either legal or illegal, and no transition is both legal and illegal:

(DDet) $\mathrm{LEG}/u \cup \mathrm{ILL}/u = \mathrm{POS}/u$

and

(DCons) $\mathrm{LEG}/u \cap \mathrm{ILL}/u = \emptyset$

(DDet) may be called the principle of deontic determinacy. If it holds for any situation u, the normative system in question has no gaps. (Cf. [von Wright, 1996, p. 47].) According to (DCons), no transition can be both legal and illegal. The assumption that there is some normatively acceptable way out of every situation, in other words,

(DactD) For every $u \in W, \mathrm{LEG}/u \neq \emptyset$

corresponds to principle (D) of SDL, that is, the postulate that every world (situation) has some deontic alternative.

Let I be an interpretation function which assigns to each action A its possible performances (a subset of $W \times W$), and let I/$u(A)$ be the performances of A which originate from u; thus $\mathrm{I}/u(A) \subseteq \mathrm{POS}/u$. The basic normative concepts of prohibition, permission (may), and obligation (ought) - deontic action modalities - can be defined by the following truth-conditions:

(CF.act) $u \models \mathbf{F}A$ iff $\mathrm{I}/u(A) \subseteq \mathrm{ILL}/u$

(CP.act) $u \models \mathbf{P}A$ iff $\mathrm{I}/u(A) \cap \mathrm{LEG}/u \neq \emptyset$

and

(CO.act) $u \models \mathbf{O}A$ iff $\mathrm{I}/u(\sim A) \subseteq \mathrm{ILL}/u$

where '$\sim A$' means that the agent does at u something incompatible with A, i.e., does not do A. These definitions are variants of the truth-conditions of normative propositions in SDL. According to (CF.act), an act A is prohibited in a given situation if every possible performance of A at that situation is illegal, and A is permitted if and only if it can be performed in a legal way. (cf. [Czelakowski, 1997, p. 60].) According to (CO.act), A is obligatory at u if only if the failure to do A would be illegal.

According to (CP.act), the permissibility of an action A means that some possible performances of A (at a given moment u) are deontically acceptable. For example, A may be permitted in this sense if it can be only performed together with some other acts, or performed in a legal way. This is a "weak" concept of permission which corresponds to that defined in SDL. In the present framework it is possible to define another concept of permission which may be termed a "strong permission". When we say that an act A is permitted in a given situation, we often mean that A itself is not illegal, in other words, that no sanction is attached A, and not only that some (possible) performances of A would be deontically acceptable in the situation. This sense of 'permission' can also be expressed in the form of a conditional: If the agent were to do A, he would not do anything illegal. The truth-conditions of such a conditional can be formulated by means of a selection function f which selects from $\mathrm{I}/u(A)$ the transitions which exemplify A but change the original situation u in other respects in a "minimal" way. Such transitions may often be described by saying that the agent does *only* A. This concept of "strong" permission is defined as follows:

(CP$^{\mathrm{s}}$.act) $u \models \mathbf{P}^s A$ iff $f(\mathrm{I}/u(A), u) \subseteq \mathrm{LEG}/u$

We might say that the f-function selects from $\mathrm{I}/u(A)$ the *minimal* performances of A. For example, if Bertie's Aunt Agatha gives him permission to take one scone, it means that the action of taking one scone is acceptable, in other words, that Aunt Agatha would not reprimand Bertie if he were to take one scone and do nothing else. On the other hand, it is permitted for a driver to flash her right turn signal - but only if she is going to make a right turn as well. The latter action is an example of a weakly permitted action (assuming that making a right turn is permitted), whereas the former action (taking a cookie) is strongly permitted. The formulation (CP$^{\mathrm{s}}$.act) is analogous to one of the standard ways of expressing the truth-conditions of conditionals by means of a selection function $f(I(p), u)$ which selects, for each proposition $I(p)$ and a situation u, the p-situations closest to u (as close to u as the truth of p permits): a conditional $p \Rightarrow q$ is true at u if and only if the consequent q is true at all selected p-worlds (i.e., the worlds in which p is true). Thus ($\mathrm{CP^s}$.act) fits the most natural reading of a strong permission to do A: if you were to do only A, you would not be doing anything illegal. (Cf. [Dignum *et al.*, 1996, pp. 200-3].) The selection function f used in ($\mathrm{CP^s}$.act) selects the "minimal" performances of A from the set of all possible performances of A, just as the truth of a conditional $p \Rightarrow q$ is determined by the selection of the p-worlds minimally different from the actual situation (or the situation where the conditional is being evaluated) [Hilpinen, 1993, p. 309].

If the disjunctive permission 'You may do F or G' is interpreted as a strong permission in the sense defined by ($\mathrm{CP^s}$.act),

$$\text{(9.13)} \qquad \mathbf{P}^s(F+G) \rightarrow \mathbf{P}^s F \,\&\, \mathbf{P}^s G$$

if and only if

$$\text{(9.14)} \qquad f(\mathrm{I}/u(F), u) \cup f(\mathrm{I}/u(G), u) \subseteq f(\mathrm{I}/u(F+G), u)$$

i.e., if the minimal performances of a disjunctive act includes the minimal performances of both disjuncts. This need not always be the case; for example, assume that Aunt Dahlia has ordered Bertie to wear black socks, and then gives the following permission:

> Bertie, you may also wear grey socks or purple socks, but you should consult Jeeves before wearing purple socks.

(See [Kamp, 1979, p. 271].) If this sentence is used normatively (performatively), Aunt Dahlia makes a disjunctive action permitted for Bertie, but refers to Jeeves's authority for the determination of the permissibility of one of the disjuncts. Therefore (9.13) is not a logical truth, but it may hold in many situations, and for pragmatic reasons it may be assumed to hold in situations in which a permission sentence is used performatively, if it would not be otherwise clear what has been permitted, that is, which performances of $F+G$ have been made normatively acceptable.

In this conceptualization of action and action propositions, the expression 'a sees to it that p', where p is an "ordinary" proposition which describes a state of the world, can be taken to mean that (i) a performs some action F which is sufficient to transform the initial state into one in which p holds, or if p is already the case, is sufficient to sustain p, and (ii) there is an alternative action G such that if a had performed G instead of F, p might have been false in the result state. This notion of 'seeing to it that' can be formally expressed by means of a modal operator $[F]$ which we define as follows first:

(9.15) $u \models [F]p$ iff $f(u, F) \subseteq I(p)$,

i.e., $[F]p$ means that any possible performance of F at u would lead to a situation in which p is true, and f here is a selection function mapping a world and an action to a set of p-worlds. The [ ]-operator is a necessity operator relativized to the action F. In general, a necessity operator relativized to the antecedent of a conditional can be used to express the meaning of a subjunctive conditional; thus the left-hand side of (9.15) may be read: if an agent were to do F, p would be true. The corresponding possibility operator is defined by

(9.16) $u \models \langle F \rangle$ iff $f(u, F) \cap I(p) \neq \emptyset$

Now in turn, 'a sees to it that p', as $\mathbf{Do}(a, p)$, can be defined as follows:

(9.17) $(u, w) \models \mathbf{Do}(a, p)$ iff there is an action F (with a as the agent) such that
 (i) $u \models [F]p$, and $(u, w) \in \mathrm{I}/u(F)$, and
 (ii) F has in u an alternative G such that $u \models \langle G \rangle \neg p$

(i) expresses here the sufficient condition aspect of 'seeing to it that', and (ii) is a weak form of the necessary condition aspect.

In many systems of the logic of the ought-to-do developed in the 1980's and 1990's, simple action descriptions are not regarded as primitive terms, as in the approach outlined above, but are obtained from propositional expressions by means of an action operator similar to the Do-operators considered earlier, which turns propositional expressions into action propositions, usually read 'a sees to it that' or 'a brings it about that'. As was noted earlier, such representations do not give as good an analysis of the concept of action, but can be regarded as representations of different forms of agency. An analysis of that kind has become widely employed in the recent work on the logic of agency and deontic logic. (See [Belnap and Perloff, 1988; Xu, 1995; Brown, 1996a; Bartha, 1999; Belnap *et al.*, 2001; Horty, 2001].)

The combination of different modes of agency with deontic concepts makes it possible to represent several types of obligation and permission and different legal or deontic relations between individuals and groups. For example, consider a state of affairs involving two persons, $F(a, b)$. According to [Kanger, 1957; Kanger and Kanger, 1966], a suitable agency operator $\mathbf{Do}(x, p)$ can be combined with deontic operators to distinguish four basic types of right (or different basic senses of the expression 'right'):

(R1) $\mathbf{ODo}(b, F(a, b))$

(R2) $\neg\mathbf{ODo}(a, \neg F(a,b)) \leftrightarrow \mathbf{P}\neg\mathbf{Do}(a, \neg F(a,b))$

(R3) $\neg\mathbf{O}\neg\mathbf{Do}(a, \neg F(a,b)) \leftrightarrow \mathbf{PDo}(a, F(a,b))$

(R4) $\mathbf{O}\neg\mathbf{Do}(b, F(a,b))$

(R1)-(R4) define four basic normative relations between a and b which from a's perspective can be regarded as different relational concepts of right. In (R1), b has a duty to see to it that $F(a,b)$; this is equivalent to a's *claim* in relation to b that $F(a,b)$. (R2) can be described as a's freedom (or *privilege*) in relation to b that $F(a,b)$; this means that a has no obligation to see to it that $\neg F(a,b)$. Kanger called (R3) a's *power* in relation to b that $F(a,b)$, and (R4) a's *immunity* in relation to b that $F(a,b)$. The replacement of the state of affairs $F(a,b)$ by its opposite $\neg F(a,b)$ yields four additional concepts of right which [Kanger and Kanger, 1966, pp. 121-2] called counter-claim (R1'), counter-freedom (R2'), counter-power (R3'), and counter-immunity (R4'). Kanger and Kanger called the 8 relations defined by (R1)-(R4) and their negative analogs *simple* types of right. The normative relationship between any two individuals with respect to a state of affairs p can be characterized completely by means of the conjunctions of the eight simple types of right or their negations. There are $2^8 = 256$ such conjunctions, but the simple types of right are not logically independent of each other: according to the logic of the deontic **O**-operator and the agency operator **Do**, only 26 combinations of the simple types of right or their negations are logically consistent. [Kanger and Kanger, 1966, pp. 126-7] called these 26 relations the "atomic types of right". The atomic types provide a complete characterization of the possible legal relationships between two persons with respect to a single state of affairs. It is perhaps misleading to call these 26 relations "types of right", because they include as their constituents duties as well as claims and freedoms. Thus Kanger's theory of normative relations can be regarded as a theory of duties as well as rights [Lindahl, 1994].

Kanger's concepts (R1)-(R4) seem to correspond to the four ways of using the word 'right' (or four concepts of a right) distinguished by W. N. Hohfeld (1919), and he adopted the expressions 'privilege', 'power' and 'immunity' from Hohfeld. Kanger apparently intended (R1)-(R4) as approximate explications of Hohfeld's notions. However, Kanger's concepts of power and immunity differ from Hohfeld's concepts. According to Kanger, both power and freedom are permissions: a power consists in the permissibility of actively seeing to it that something is the case, whereas freedom means that there is no obligation to see to it that the opposite state of affairs should be the case. [Lindahl, 1977, pp. 193-211] and others have argued that Hohfeld's concept of power should be analyzed as a legal *ability* rather than a permission (a *can* rather than *may*). (See [Lindahl, 1994; Bulygin, 1992;

Makinson, 1986].)

An agency operator such as the **Do**-operator considered above can be iterated, and it can therefore be used to form sentences which contain several nested occurrences of various modal operators: deontic operators, praxeological operators (for various forms of action and agency), and epistemic operators, which can be relativized to different agents. This feature makes it possible to apply deontic logic and the logic of agency to the analysis of complex social and normative phenomena, for example, the analysis of different concepts of right and other normative relations [Kanger, 1984; Makinson, 1986; Lindahl, 1994], governmental structures and the concept of parliamentarism [Kanger and Kanger, 1966], normative positions and normative change [Lindahl, 1977; Jones and Sergot, 1993; Sergot, 1999], and the study of social control, influence, and responsibility [Pörn, 1989; Santos and Carmo, 1996].

As even this brief exposition of Kanger's analysis of legal relations might suggest, the specification of such relations lends itself rather well to computational techniques, as demonstrated rather explicitly by the work of Sergot in extending the theory of normative positions. (See [Sergot, 1999], as well as his chapter in this volume.) This is only one among the many rich ways in which computer science and deontic logic have developed a fruitful relationship. For a locus classicus on this, see most of the chapters (including the chapter two overview) in [Meyer and Weiringer, 1993], the first volume of papers drawn from the inauguration of the *Deontic Logic in Computer Science* series of binannual conferences (DEONs), the preeminant conference forums (with associated publications) for work in deontic logic.

Acknowledgements

We would like to thank the editors, Dov Gabbay, Jeff Horty, Ron van der Meyden, Xavier Parent, and Leendert van der Torre for overseeing the handbook project, and for their patience, with special thanks going to Xavier Parent for his careful reading and helpful comments on final drafts of this chapter.

Appendix to 8.5 on Chisholm's puzzle and conditional norms

Consider the key inferences generating the Chisholm paradox: the inference from (8.41') and (8.42') to $\mathbf{O}t$, and the inference from (8.43') and (8.44') to $\mathbf{O}\neg t$. Each involves "detachment" of an **O**-statement from a pair of premises, one being a deontic conditional. Let us explore this by introducing the following symbolism,

(NS) $\mathbf{O}(q/p)$,

taken here merely as a neutral shorthand for a natural language conditional obligation or ought statement such as (8.43) above.[90] $\mathbf{O}(q/p)$ is then to be read as "if p, then it ought to be that q". We will also assume that monadic obligations are necessarily equivalent to special dyadic obligations, per the following fairly standard analysis of unconditional obligations:

(UCO) $\mathbf{O}p =_{\text{df}} \mathbf{O}(p/\top)$

That is, it is obligatory (simpliciter) that p if and only if it is obligatory that p if tautological conditions hold (which they always do of course).[91]

Two types of "detachment principles" [Greenspan, 1975] emerged quickly in the literature on Chisholm's paradox:

(FDt) $(p \,\&\, \mathbf{O}(q/p)) \rightarrow \mathbf{O}q$ (Factual detachment)

(DDt) $(\mathbf{O}p \,\&\, \mathbf{O}(q/p)) \rightarrow \mathbf{O}q$ (Deontic detachment)[92]

(FDt) says it is a logical truth that given both if p then it ought to be that q and p *itself*, then it ought to be that q. (DDt) says that given both if p then it ought to be that q and *it ought to be that* p, then it ought to be that q. As the principles' names indicate, given the same deontic conditional ($\mathbf{O}(q/p)$), the main difference is that per (FDt), it is the factual claim (p) that allows us to detach the deontic conclusion ($\mathbf{O}q$); whereas per (DDt), it is the deontic claim ($\mathbf{O}p$) that allows us to detach that conclusion.[93]

[90]The logical differences between "obligation" and "ought" will not matter here, so we will use them interchangeably here.

[91]This definition has been widely endorsed and employed, but not universally so (e.g. [Alchourrón, 1993; Carmo and Jones, 2002] reject it). But see also [Parent, 2012] on some difficulties with some alternatives to UCO.

[92]We add "t"'s so that references to deontic detachment will not be confused with those to DD (SDL's no conflicts principle).

[93]Those who followed [von Wright, 1956] in viewing deontic conditionals as sui generis and not definable via a monadic operator and any non-evaluative conditional notion rejected (FDt), even if we shift to a non-material conditional. (8.42) above follows Chisholm's original example in having the conditional explicitly in the scope of the English "it ought to be that" construction, so it is not a "deontic conditional" as just characterized. For that, we would have to add that (8.42) is logically equivalent to the non-wide-scope construction: "if Jones does go, then he ought to tell them he is coming". Although this is hardly obvious, as mentioned above, the difference between (8.42) and (8.43) in Chisholm's original formulation is largely seen as inessential, so that "purified" presentations of premises in the role of (8.42) and (8.43) would both match each other in superficial form, and usually that of "if ..., then it ought to be that..." as in (8.43).

Regarding the standard systems, if we were to interpret a deontic conditional with "**O**" having narrow scope, that is, as a material conditional with an obligatory consequent (i.e. $(p \to \mathbf{O}q)$, as in (8.43') above), (FDt) would be derivable by MP, but (DDt) would not be derivable (e.g. the T axiom is not a thesis). Conversely, if we interpret a deontic conditional with "**O**" having wide scope, that is, as an obligatory material conditional (i.e. $\mathbf{O}(p \to q)$, as in (8.42') above), (DDt), but not (FDt), is derivable by principle KD.[94]

Earlier we saw that neither of these two interpretations of natural language deontic conditionals via *material* conditionals is at all tenable, but the fact that the two interpretations require an acceptance of one form of detachment and a rejection of the other reflects an important fact: endorsement of both types of detachment (without some restriction) is only plausible if it is plausible to conclude that the Chisholm scenario involves an outright conflict of obligations. For if both detachment forms are endorsed, we end up both obligated to tell (the neighbor we are coming) and also obligated to not tell them we are coming. Most have thought that this is not a case of conflicting obligations, and that something else generates the puzzle. As a result, researchers tended to divide up into two camps according to which principle they took to be deductively valid [Loewer and Belzer, 1983]. We can thus think of the two emerging positions as refinements on the failed narrow scope and wide scope readings of deontic conditionals via material conditionals. For as we can see above, the second premise needed in addition to the reinterpreted deontic conditional, in each case, parallels the narrow scope and wide scope readings via material conditionals: p itself is needed for (FDt); $\mathbf{O}p$ is needed for (DDt).

Deontic logicians who favored (FDt) typically held that deontic conditionals like those in (8.43) involve a *non-material* conditional, such as a subjunctive conditional, but otherwise things are just as they appear. The logical form matches the surface grammatical form: the main operator is deemed to be a conditional that has a consequent in the scope of a monadic

Either way, the inference from (8.41) and (8.42) - or the relevant analog to (8.42) - to "it ought to be that Jones tells" is still called "deontic detachment", and likewise for their formal analogues in the standard systems, where KD validates the inference from (8.41') and (8.42') to $\mathbf{O}t$. The crucial thing is the deontic character of the simpler premise in deontic detachment from a deontic conditional.

[94][Smith, 1994] notes that if $\mathbf{O}(q/p)$ is interpreted as $\mathbf{O}(p \to q)$ and we add factual detachment to SDL, we get Mally's collapse: $\vdash \mathbf{O}p \leftrightarrow p$. The part from right to left follows from (FDt) by RND since $\mathbf{O}(p \to p)$ is a thesis. Of greater interest, Smith, crediting Andrew Jones, points out that even a minimal deontic logic that contains merely RED and OD will generate $\mathbf{O}p \to p$. From (FDt), we get $\vdash p \,\&\, \mathbf{O}(p \to \bot) \to \mathbf{O}\bot$, and then from OD, that yields $\vdash \neg(p \,\&\, \mathbf{O}(p \to \bot))$, and then from RED we get $\vdash \neg(p \,\&\, \mathbf{O}\neg p)$, and $\vdash \mathbf{O}p \to p$.

deontic operator, and an ordinary antecedent - the result being an analysis of such conditionals as genuine conditional-deontic *compounds*:

(CDC) $\mathbf{O}(q/p) =_{\text{df}} p \Rightarrow \mathbf{O}q$ for some independent conditional $\Rightarrow$

Non-classical "closest antecedent worlds" conditionals of the sort made famous by Stalnaker and Lewis predominated. The logics of such deontic conditional compounds will then derive from the logic for the non-deontic conditional operator and the logic for the monadic deontic operator. Typically, the conditionals offered, along with the truth of their antecedents, would entail their consequents (a version of modus ponens would hold for the non-material conditional), so (FDt) would hold under this sort of analysis. Its truth-conditions might be formulated in a natural way by means of a selection function $f(I(p), u)$ which selects for each world u and proposition p (a set of possible worlds) presented for consideration by the protasis 'if p'; the apodosis then states that q is true in all situations selected by the protasis:

(8.54) $u \models p \Rightarrow q$ iff $f(I(p), u) \subseteq I(q)$,

where $I(q)$ is the set of possible situations in which q is true.[95] It is important to note here that the set of situations (the proposition) selected by the f-function (i.e., selected by the protasis) depends on the situation u about which the conditional statement is made; thus the sentence in the antecedent may be said to express different propositions in different situations, and the conditionals defined by (8.54) can be said to be "variably strict" rather than strictly necessary conditionals. It is typically assumed that the selection function satisfies the following condition:

(8.55) if $u \models p$ then $u \in f(I(p), u)$

According to (8.54) and (8.55), the following is valid:

(8.56) $(p \Rightarrow q) \rightarrow (p \rightarrow q)$

So a version of modus ponens applies: given p and $p \Rightarrow q$, q follows.[96]

Those favoring this sort of approach usually typically justify their rejection of (DDt) on the following grounds. Conditional obligations like those in (8.42) tell us only what to do in *ideal* circumstances where we keep our primary obligations like those in (8.41); but they thus do not provide guidance or "cues" for action in circumstances where the primary obligation is

[95] Cf. [Moore, 1973; Chellas, 1974; Chellas, 1980].

[96] For other conditions for the f-function and other semantic models for conditionals, see [Lewis, 1973].

not met.[97] In the Chisholm scenarios, combinations like (8.44) and (8.41) entail that the primary obligation, whose execution is hypothesized in the first clause of the conditional obligation (8.42), has been violated. Thus that Jones ought to tell is not entailed by the fact that he ought to go and he ought to tell *if* he goes. If he tells and doesn't go, he makes things worse than if he merely doesn't go. Perhaps the most we can say is that *ideally* he ought to tell. But since on its face, a version of Modus Ponens holds for the conditional in (8.43), if he does not go (8.44) then it follows from that Jones ought to not tell.[98]

In contrast, those favoring (DDt) over (FDt) might object by citing a conditional reminiscent of Forrester's, such as "If Jones will kill his rich aunt now (for the inheritance), then he ought to shoot her to death" (his only immediate means being strangulation or a nearby hunting rifle, say). They then might explore how the picture of those favoring (FDt) holds up for such an example as follows. Suppose Jones will kill his rich aunt as a matter of contingent fact, although he could refrain. Then those favoring (FDt), by parity of reasoning, would have to say that although Jones is obligated to not kill his aunt, nonetheless, because he in fact will do so, he is obligated to shoot her to death, and at most only *ideally* ought to not do so. But the idea that Jones' obligation to not shoot his aunt to death merely expresses an ideal obligation, not an actual obligation, is not easy to accept. Similarly, if it is unqualifiedly obligatory that Jones not kill his aunt, as the friends of (FDt) agree, then it must be impermissible to kill her, and so impermissible to do so by any particular means.[99]

The suggestion then is that *unrestricted* factual detachment seems to allow the mere fact that Jones will do something avoidable and terribly wrong to generate an actual obligation to do something also terribly wrong, though less wrong (even if only infinitesimally less wrong).

It is also to be noted that accounts that allow for factual detachment risks entailing "the pragmatic oddity" [Prakken and Sergot, 1994; Prakken and Sergot, 1996]. Using the Chisholm's quartet, suppose one's analysis

[97]See [Vorobej, 1982] for this idea of "cues" for action.

[98][Dayton, 1981] was influential in arguing that despite the importance of subjunctive conditionals in deontic contexts, the Chisholm puzzle involves a different special deontic conditional not expressible by this means.

[99]In Chisholm's example it is not intuitively clear what is involved—helping the neighbors with a fire, load their moving truck, help them in with the groceries. It is thus easier to accept that letting them know you will help is merely ideal, but not required, since helping might be something you ought to do, but not something you must do or are obligated to do. By default, examples like the one above immediately rule out this sort of "recommended but not required" interpretation. Furthermore, it is not a case where the apparent consequent entails the antecedent as in the case above where what ought to be done if. . . is a way of doing what is hypothetically posited in the antecedent.

countenances the conclusion that "Jones ought to not tell his neighbors he is coming" (from factual detachment applied to (8.43) and (8.44), as well as countenacing the truth of the primary obligation (8.41), "Jones ought to go to his neighbor's assistance"? If the theory allows for aggregation of these two obligations, as all the standard systems do (by KD), we get the conclusion that "Jones ought to go to his neighbor's assistance and not tell them he is coming", which certainly sounds odd, if not false. The original surely does not have this consequence, and yet it looks like any account that embraces factual detachment and RMD will generate this oddity. Prakken and Sergot suggest that this secondary puzzle places pressure on assuming univocality for the "oughts" in Chisholm's paradox; [Carmo and Jones, 2002] make avoiding the pragmatic oddity a desiderata of any adequate account of Chisholm's paradox.

Many who rejected factual detachment represented conditional obligations via a primitive *dyadic* obligation operator (reminiscent of the syntax of a conditional probability operator). They rejected CDC ($\mathbf{O}(q/p) =_{\text{df}} p \Rightarrow \mathbf{O}q$). They believed the logical form of such conditionals was hidden by the surface grammar: the meaning of the compound is not a straightforward function of the meaning of the apparent parts. The underlying intuition is that even if Jones will violate his obligation, that doesn't get him off the hook from obligations that derive from the one he will violate. If he must go help and he must inform his neighbors if he will go, then he must inform them as well, and the fact that he will violate the primary obligation does not block the derivative obligation any more than it does the primary one itself. He is still an agent subject to both constraints.

A "best of the antecedent worlds" semantic picture for the latter approach quickly emerged with Hansson's seminal work:

(BAW) $\mathbf{O}(q/p)$ is true at a world u iff the u-best p-worlds are all q-worlds.[100]

It follows from this by the standard analysis of the monadic operator (UCO) in dyadic contexts that

(8.57) $\mathbf{O}q$ is true iff $\mathbf{O}(q/\top)$, so iff all the unqualifiedly u-best worlds are q-worlds.

This approach, which relies on preference-orderings for the semantics of dyadic conditional obligations, became a widespread trend. (Structurally,

[100] [Hansson, 1971]. See [Spohn, 1975] for a weak completeness theorem for one key system DSDL3 for which Hansson provided a semantics, and see [Parent, 2008] for a strong completeness proof for DSDL3, as well as [Parent, 2010] for such a proof for another system proposed by Hansson, DSDL2.

this ordering semantics approach was also a forerunner of a variety of approaches (to different phenomena) employing what [Makinson, 1993] characterizes as "the notion of minimality under a relation", as in that for defeasible conditionals such as "if p, normally q" that became so central in AI.) Factual detachment does not hold on this picture, since even if our world is one where Jones does not go to the assistance of his neighbors, and the best among those worlds are ones where he doesn't tell them he is coming, it does not follow that the *unqualifiedly* best worlds are ones where he doesn't tell them he is coming; in fact, the best such worlds are ones where he both goes to their aid and lets them know that he will do so.

However, a natural objection now emerges: what is the point of such "conditionals" if we are not allowed to detach the apparent consequents from the apparent antecedents - how do we *reason* with them? This suggests that the above line of reasoning for rejecting unqualified (FDt) is not enough, and so it was typically coupled with a *restricted* form of factual detachment, such as:

(RFDt) $(\Box p \,\&\, \mathbf{O}(q/p)) \rightarrow \mathbf{O}q$ (Restricted factual detachment)

$\Box p$ ypically meant that p is now unalterable for the imagined agent.[101] The intuition is that we can conclude $\mathbf{O}q$ from $\mathbf{O}(q/p)$ only if p is not simply true but unalterably so (in the context of evaluation). This is certainly an important complement to the reasoning above for rejecting (FDt), since it does allow for a form of qualified factual detachment, and thus for reasoning from the non-deontic status of the apparent antecedent of deontic conditionals to the apparent deontic consequent. (More nuanced positions emerged, for example in [Loewer and Belzer, 1983], where the authors endorse a special form of factual detachment distinct from those above. This can perhaps be seen as further reflecting the felt need to move to more nuanced positions beyond the dilemma of having to simply choose between (FDt) and (DDt).) However, there is still the question of why this *certainly apparent* composite of a conditional and a deontic operator is actually some sort of primitive idiom and not purely derivative.

We are left with an apparent dilemma: either a) unqualified factual detachment holds and we swallow the consequence that often because someone freely will act horribly, she is obligated to do some slightly less, still horrible, thing; or b) that "if p, then ought q" contrary to appearances, is really an idiom, and the meaning of the whole is not a function of the meaning of apparent conditional and deontic parts, with all the challenges about how we learn the construction if it is not compositional. Neither option seems very satisfying.

[101] [Greenspan, 1975] argues for this position explicitly, and many endorsed it as well.

More nuanced positions emerged, for example in [Loewer and Belzer, 1983], where the authors endorse a special form of factual detachment distinct from those above (cf. [Chisholm, 1964]). This can perhaps be seen as further reflecting the felt need to move to more nuanced positions beyond the dilemma of having to simply choose between (FDt) and (DDt). However, there is still the question of why this certainly apparent composite of a conditional and a deontic operator is actually either some sort of primitive idiom or a composite with a hidden modal antecedent.

We set the issue of (FDt) and (DDt) aside to turn briefly to two key features of the Chisholm scenario that are not represented in the standard systems and that people proposed were central to solving the puzzle.

One popular strategy for solving Chisholm's puzzle has been to carefully distinguish the times of the obligations.[102] This was reinforced by the fact that there are strong independent reasons to be concerned about differentiating the times at which things are obligatory. This was often accompanied by consideration of examples where the candidate "derived" obligations were things to be done *after* the violation (or fulfillment) of the primary obligation (a "forward" version of a CTD case), and indeed this appeared essential to many of the solutions offered. However, Chisholm's own seminal example does not fit so well here. It is naturally interpreted as either a case where at best the obligation to help and the purportedly derivable obligation to tell are *simultaneous* (a "parallel" version)[103], and at worst and more plausibly, where the telling is something to *precede* the going ("backward" versions).[104] After all, "I did help" or "I am now helping" are likely to be obvious to the neighbors, and surely letting them know you are on your way to them to provide aid when you arrive is the natural default reading. As with [Jones, 1990; Prakken and Sergot, 1994; Prakken and Sergot, 1996] stress this shortcoming with temporal solutions with a variety of examples, one being:

(8.58) The children ought not to be cycling on the street

[102] [Thomason, 1981a; Thomason, 1981b] are classics arguing for the general importance of layering deontic logic on top of temporal logic. [Vorobej, 1982; Loewer and Belzer, 1983; Åqvist and Hoepelman, 1981; Feldman, 1986] argued that attention to time is crucial in handling the Chisholm puzzle, or at least some versions thereof (among other puzzles). See also [Chellas, 1980]. For an early dissenting opinion on temporal solutions, see [Castañeda, 1977].

[103] [Jones, 1990] interprets it this way and, more importantly, stresses that such a clearly possible case tells against the suggestion that distinguishing times is at the heart of the puzzle.

[104] [Dayton, 1981; Smith, 1994] contains an illuminating discussion of the three different versions of the Chisholm puzzle (backward, parallel, and forward versions) in evaluating different approaches to solving the Chisholm paradox; in [Smith, 1993], she credits J. J. Meyer for the 'backward'-'forward' terminology.

(8.59) If the children are cycling on the street, then they ought to be cycling on the left hand side of the street

(8.60) The children are cycling on the street

They point out that the intention is surely for the first two to hold at the same time. So there are no times to separate to say the one obligation holds at t_1, but not t_2, and the other holds at t_2 not at t_1. Yet there is surely prima facia reason to think the same phenomena driving the Chisholm paradox is present above.[105]

Alternatively, some suggested that carefully separating the agential components of the example from the (non-agential) circumstancial components would solve the puzzle.[106] Again, there are plainly independently compelling reasons to pursue agency in deontic logic. However, once again, in the case of Chisholm's seminal example, it appears that the two key elements at issue are agential, for each appears to be an action, and one open to the agent as of the time of the puzzle scenario: going to the neighbors' assistance; telling the neighbors' you will help. Furthermore, there are non-agential versions of Chisholm's example such as this variant on others found in [Prakken and Sergot, 1994; Prakken and Sergot, 1996]:

(8.61) There ought to be no hurricane

(8.62) If there is no hurricane, the shutters ought not be closed

(8.63) If there is a hurricane, the shutters ought to be closed

(8.64) There is a hurricane

Here the case seems to parallel that of Chisholm's original example rather well in broad respects, yet there is no reference to actions or agency at all; instead the reference seems to only be to different states of affairs, with the first claim telling us what is *ideally* the case, and the last telling us this ideal circumstance is not realized, with the claims in between telling us what ought to be under the respective *ideal* and *non-ideal* circumstances. (There appears to be no reference to different times here either.)

We note lastly that there have been some attempts to suggest that the problem with Chisholm's paradox might be solved by applying standard concepts of defeasibility from non-monotonic logic, such as that of excep-

[105] The *Forrester paradox* above quintessentially involves a parallel duties case.

[106] [Castañeda, 1981] is a salient and influential instance, arguing that we need to distinguish actions construed as circumstances and actions construed as prescriptions to solve the problem. See also [McNamara, 1988], especially influential for its employment of dynamic logic in deontic contexts; [McNamara, 1988] also offers a solution to Chisholm's puzzle in a broadly similar vein as Castañeda (along with solutions for other puzzles in deontic logic).

tions to normative generalizations, etc.[107] However, this does not seem to jive well with the prima facia difference between violation and defeat.[108] In Chisholm's example, the natural reading is that Jones is obligated to help his neighbors unexceptionally and indefeasibly, but nonetheless, in fact he will not, so it is now true that he will (in the future) violate that obligation. This fact does not defeat that obligation, nor does the corresponding contrary to duty obligation override it or cancel it. Even if defeasibility might figure in part of the story, it seems that no discussion absent of violation concepts will suffice to cover essential features of CTD cases.[109] [Prakken and Sergot, 1996] makes the point nicely with the following example (trivially modified):

(8.65) There must be no fence
(8.66) If there is a fence, it must be a white fence
(8.67) If the cottage property includes a cliff edge, there may be a fence

Suppose (8.65) is meant defeasibly, and a cottage near a cliff edge constitutes the only defeater. Suppose now that Jones has a cottage, but not one near a cliff edge, and he has a red fence. Then he is in violation of (8.65), an undefeated (though defeasible) primary obligation for him, and he is also in violation of (8.66), since he (impermissibly) has a fence, and it is not a white one. Contrast Doe, who has a cottage near a cliff edge and a red fence. Doe is not in violation of (8.65), since it is defeated (undercut) by her exceptional circumstances. Is she in violation of (8.66)? That depends on how (8.66) is evidently meant. Imagine it comes just on the heels of (8.65) in the cottage properties manual, preceded with a "However,"; whereas (8.67) comes in the manual's appendix along with a general discussion of special exceptions to various rules. Then Doe is in full compliance with (8.65)-(8.67), since there are apparently no color restrictions for fences by a cliff

[107]For example, see [Merin, 1992; Ryu and Lee, 1991] McCarty 1992 and Ryu and Lee 1991, and for an earlier work stressing defeasible principles and the Chisholm paradox, see [Loewer and Belzer, 1983] Loewer and Belzer 1983 and Belzer 1986.

[108]Smith 1993 [Smith, 1993] briefly discusses the importance of the difference for the Chisholm puzzle, and at greater length again in [Smith, 1994] Smith 1994, stressing that the central feature of Chisholm puzzles is violation of the primary obligation, not defeat thereof. Prakken and Sergot 1994, Prakken and Sergot 1996, Prakken and Sergot 1997 [Prakken and Sergot, 1994; Prakken and Sergot, 1996; Prakken and Sergot, 1997] also stress the difference and argue for the unresolvability of the puzzle using only defeasiblity. For a dissenting opinion however, see [Bonevac, 1998] Bonevac 1998, which argues that the problem is solvable using defeasiblity, and that this also allows for the analysis of CTDs as composites of a conditional and a monadic obligation operator, pace the dyadic approach.

[109]We assume the point here stands even if the *concept* of violability itself is somehow analyzable via *defeat* concepts, for the key point is that we cannot avoid invoking the difference between a defeated or cancelled primary obligation and a violated one in standard CTD cases, nor the understanding of the CTD as conditional on said violation.

edge. In contrast, if we imagine that (8.66) is intended to cover all cases, not just violations of (8.65), then Doe (along with Jones) is in violation of (8.66), but it is a CTD for Jones only.

Thus, however much temporal, agential or action-related aspects of deontic reasoning are important in their own right, it does not appear that any of them hold the key to resolving the general problem that the Chisholm puzzle indicates. Similarly, however important defeasibility is to deontic reasoning, and even if it ultimately has some role to play in a final resolution of Chisholm puzzles, it appears that the difference between defeated and violated obligations will survive that, as will the difference between a deontic conditional intended as telling us what to do if we violate an undefeated obligation, and one telling us what we are obligated to do conditional upon our performing some optional or obligatory action.

The Chisholm puzzle has been highly resistant to simple or even fully satisfying solutions, and using the term "paradox" seems less overstated than in the case of many of the other standard deontic logic puzzles. Many consider Chisholm's paradoxes to be the most important and distinctive puzzle in the development of deontic logic.[110] As noted earlier, SDL is just the normal modal logic, D, and most of the early deontic logics were extensions of SDL. This suggested deontic logic was just an interesting but simple application/interpretation of some simple normal modal logics. But the puzzles with deontic conditionals, especially Chisholm's, helped solidify deontic logic as a distinct specialization in the 1960s and 1970s, one for which normal modal logics like SDL were deemed inadequate.[111] This led to the development of alternative more complex logics for deontic conditionals, and then to a widespread (though as noted, not uncontested) perception that some sort of ordering semantics provided important and promising structures for modeling deontic conditionals, and that in all events, more elaborate expressive and semantic resources were called for.

For the reader interested in seeing what a logic for conditional obligation might look like, we provide one favoring deontic detachment that might be seen as the conditionalized counterpart to SDL, and is indeed called "SDDL" for "Standard Dyadic Deontic Logic" in [Goble, 2003]. We do not provide the semantics here, but refer the interested reader to his article and to the informal remarks above about ordering semantics for SDL.[112] We do not

[110] For example, [Carmo and Jones, 2002].

[111] It is also arguable that the development of conflict-tolerant deontic logics in response to puzzles like Sartre's' dilemma and Plato's dilemma has also been liberating for deontic logic (although none of the aforementioned puzzles has held deontic logicians quite as captivated as Chisholm's puzzle has).

[112] [Goble, 2003; Goble, 2004] goes through the metatheory for ordering semantic approaches to deontic conditionals in the dyadic logic tradition (as well as monadic SDL

present an analog in the Factual Detachment tradition–approaches in the vein of (CDC) above, since [Chellas, 1980] is widely available (deservedly), and contains a nice presentation of such a logic and its semantics.

The system below is deductively equivalent to system CD in [van Fraassen, 1972] and system VN in [Lewis, 1973], but Goble's semantics is more transparently stated via ordering relations, $\geq_u$, like those discussed above regarding the use of ordering semantics for SDL (Section 7.1) and regarding the limit assumption dilemma (Section 8.4).

The system extends a language and logic for PC as follows. A dyadic operator, $\mathbf{O}(/)$, is added and monadic $\mathbf{O}$ is defined as mentioned above: $\mathbf{O}p =_{\text{df}} \mathbf{O}(p/\top)$. Dyadic and monadic permissibility can then be defined in a typical way for dyadic approaches: $\mathbf{P}(p/q) =_{\text{df}} \neg\mathbf{O}(\neg p/q)$ and $\mathbf{P}(p/\top) =_{\text{df}} \neg\mathbf{O}(\neg p/\top)$. However these are not employed in the axiom system below, but for convenience, an ordering relation is defined for the language (not to be confused with the world-relative ordering relation in the ordering semantics, $\geq_u$, and used in the axiomatization:

(Df$\geq$) $\quad p \geq q =_{\text{df}} \neg\mathbf{O}(\neg p/p \vee q)$

(Df$\geq$) says that the proposition that p is as normatively good as the proposition that q just in case it is not obligatory that $\neg p$ on the condition that either p or q. Given the definition of the permissibility operators, this amounts to saying that p is at least as good as q iff p is permissible given $p \vee q$. (At the semantic level, assuming bests for simplicity here, this says roughly p is as good as q just in case there is some best $p \vee q$-world that is a p-world.) SDDL, the dyadic analog to SDL, is then as follows:

(A1) All PC tautologies in the language (TAUT)

(A2) $\mathbf{O}(p \to q/r) \to (\mathbf{O}(p/r) \to \mathbf{O}(q/r))$ (CKD)

(A3) $\mathbf{O}(p/q) \to \neg\mathbf{O}(\neg p/q)$ (CDD)

(A4) $\mathbf{O}(\top/\top)$ (CON)

(A5) $\mathbf{O}(q/p) \to \mathbf{O}(q \& p/p)$ (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 $\vdash \mathbf{O}(r/p) \leftrightarrow \mathbf{O}(r/q)$ (CRED)

(R3) If $\vdash p \to q$ then $\vdash \mathbf{O}(p/r) \to \mathbf{O}(q/r)$ (CRMD)

A1-A4 and R1-R3 are conditional analogues of formulas we used for the standard systems so we just preface those labels (e.g. "KD") with a "C"

itself), but generalizing in interesting ways beyond the standard systems (e.g. to allow for conflicts).

(but note that a conditional analog of VW would not have axiom A4, so that is a "non-standard system" in that respect). A5 and A6 are needed to generate a complete system relative to the intended ordering semantics, and they are more unique to the dyadic conditional. The first says that if q is obligatory at all given p, then p "crosses over" from the condition side to the obligation side and joins with q. (At the semantic level, assuming bests for simplicity, it roughly reflects the idea that if there are best p-worlds and they are all q-worlds, then they all must be $p\&q$-worlds.) The second takes advantage of the definitional abbreviation for "$\geq$" to even more perspicuously reflect in the language a feature of the ordering intended in the semantics and the way "$\mathbf{O}(\,/\,)$" is to be defined via that semantics.

BIBLIOGRAPHY

[Aikhenvald, 2010] A. Y. Aikhenvald. *Imperatives and Commands.* Oxford University Press, Oxford and New York, 2010.

[al Hibri, 1978] A. al Hibri. *Deontic Logic: A Comprehensive Appraisal and a New Proposal.* University Press of America, Washington, DC, 1978.

[Alchourrón and Bulygin, 1971] C. E. Alchourrón and E. Bulygin. *Normative Systems.* Springer-Verlag, Wien/New York, 1971.

[Alchourrón and Bulygin, 1981] C. E. Alchourrón and E. Bulygin. The expressive conception of norms. In R. Hilpinen, editor, *New Studies in Deontic Logic: Norms, Actions, and the Foundations of Ethics*, pages 95–124. Reidel, Dordrecht, 1981.

[Alchourrón and Bulygin, 1993] C. Alchourrón and E. Bulygin. On the logic of normative systems. In H. Stachowiak, editor, *Pragmatik: Handbuch pragmatischen Denkens.* Band iv. *Sprachphilosophie, Sprachpragmatik und formative Pragmatik*, pages 273–294. Felix Meiner Verlag, Hamburg, 1993.

[Alchourrón and Makinson, 1981] C. E. Alchourrón and D. Makinson. Hierarchies of regulations and their logic. In R. Hilpinen, editor, *New Studies in Deontic logic: Norms, Actions, and the Foundations of Ethics*, pages 125–148. Reidel, Dordrecht, 1981.

[Alchourrón, 1969] C. E. Alchourrón. Logic of norms and logic of normative propositions. *Logique et Analyse*, 12:242–268, 1969.

[Alchourrón, 1993] C. E. Alchourrón. Philosophical foundations of deontic logic and the logic of defeasible conditionals. In J.-J. Meyer and R. J. Weiringer, editors, *Deontic Logic in Computer Science: Normative System Specification*, pages 43–84. Jone Wiley and Sons, Ltd., Chichester, 1993.

[Alchourrón, 1996] C. E. Alchourrón. Detachment and defeasibility in deontic logic. *Studia Logica*, 57(1):5–18, 1996.

[Anderson, 1958] A. R. Anderson. A reduction of deontic logic to alethic modal logic. *Mind*, 67:100–103, 1958.

[Anderson, 1967] A. R. Anderson. The formal analysis of normative systems. In N. Rescher, editor, *The Logic of Decision and Action*, pages 147–213. University of Pittsburgh Press, Pittsburgh, 1967. Originally published as Technical Report No. 2, Contract No. SAR/Nonr-609 (16) New Haven: Office of Naval Research, Group Psychology Branch, 1956.

[Apostel, 1982] L. Apostel. Towards a general theory of argumentation. In E. M. Barth and J. L. Martens, editors, *Argumentation. Approaches to Theory Formation*, pages 93–122. John Benjamins B. V., Amsterdam, 1982.

[Åqvist and Hoepelman, 1981] L. Åqvist and J. Hoepelman. Some theorems about a 'tree' system of deontic tense logic. In R. Hilpinen, editor, *New Studies in Deontic Logic: Norms, Actions and the Foundations of Ethics*, pages 187–221. Reidel, Dordrecht, 1981.

[Åqvist and Mullock, 1989] L. Åqvist and P. Mullock. *Causing Harm*. Walter de Gruyter, Berlin-New York, 1989.

[Åqvist, 1963] L. Åqvist. Deonic logic based on a logic of better. In *Proceedings of a Colloquium on Modal and Many-Valued Logics; Acta Philosophica Fennica*, volume XVI, pages 285–290. Societas Philosophica Fennica, 1963.

[Åqvist, 1967] L. Åqvist. Good samaritans, contrary-to-duty imperatives, and epistemic obligations. *Noûs*, 1:361–379, 1967.

[Åqvist, 1974] L. Åqvist. A new approach to the logical theory of actions and causality. In S. Stenlund, editor, *Logical Theory and Semantic Analysis*, pages 73–79. D. Reidel, Dordrecht, 1974.

[Åqvist, 1987] L. Åqvist. *Introduction to Deontic Logic and the Theory of Normative Systems*. Biblopolis, Napoli, 1987.

[Åqvist, 2002] L. Åqvist. Deontic logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 8, pages 147–264. Kluwer Academic Publishers, Dordrecht, 2nd edition, 2002.

[Armstrong, 2004] D. Armstrong. *Truth and Truthmakers*. Cambridge University Press, Cambridge, 2004.

[Artikis *et al.*, 2012] A. Artikis, R. Craven, N. K. Cicekli, B. Sadighi, and K. Stathis, editors. *Logic Programs, Norms and Action - Essays in Honor of Marek Sergot on the Occasion of His 60th Birthday*, volume 7360 of *Lecture Notes in Computer Science*. Springer, 2012.

[Asher and Bonevac, 1996] N. Asher and D. Bonevac. "Prima facie" obligation. *Studia Logica*, 57(1):19–45, 1996.

[Austin, 1954] J. Austin. *The Province of Jurisprudence Determined and the Uses of the Study of Jurisprudence, with Introduction by L. A. Hart*. Noonday Press, New York, 1954.

[Bartha, 1999] P. Bartha. Moral preference, contrary-to-duty obligation and defeasible oughts. In P. McNamara and H. Prakken, editors, *Norms, Logics and Information Systems*, pages 93–108. IOS Press, Amsterdam, 1999.

[Bell *et al.*, 2001] J. L. Bell, D. DeVidi, and G. Solomon. *Logical Options*. Broadview Press, Peterborough (Ontario) - Orchard Park (N.Y.), 2001.

[Belnap and Perloff, 1988] N. Belnap and M. Perloff. Seeing to it that: A canonical form for agentives. *Theoria*, 54:175–199, 1988.

[Belnap and Perloff, 1992] N. Belnap and M. Perloff. The way of the agent. *Studia Logica*, 51:463–484, 1992.

[Belnap *et al.*, 2001] N. Belnap, M. Perloff, and M. Xu. *Facing the Future: Agents and Choices in our Indeterminist World*. Oxford University Press, New York, 2001.

[Belnap, 1991] N. Belnap. Backwards and forwards in the modal logic of agency. *Philosophy and Phenomenological Research*, 51:777–807, 1991.

[Belzer, 1986] M. Belzer. Reasoning with defeasible principles. *Synthese*, 66(1):135–158, 1986.

[Bentham, 1948] J. Bentham. *An Introduction to the Principles of Morals and Legislation*. Hafner Publishing Company, New York, 1948. First pub. date: 1789.

[Bentham, 1983] J. Bentham. Deontology; together with a table of the springs of action; and the article on utilitarianism. In *Collected Works of Jeremy Bentham*. Clarendon Press, Oxford, 1983. First pub. date: 1834.

[Boh, 1985] I. Boh. Belief, justification, and knowledge – some late-medieval epistemic concerns. *Journal of the Rocky Mountain Medieval and Renaissance Association* (current title *Quidditas*), 6:87–103, 1985.

[Boh, 1993] I. Boh. *Epistemic Logic in the Later Middle Ages*. Routledge, London and New York, 1993.

[Bonevac, 1998] D. Bonevac. Against conditional obligation. *Noûs*, 32(1):37–53, 1998.

[Bostock, 1997] D. Bostock. *Intermediate Logic*. Clarendon Press, Oxford, 1997.

[Brewka, 1989] G. Brewka. Non-monotonic logics - a brief overview. *AI Communications: The European Journal of Artificial Intelligence*, 2(2):88–97, 1989.

[Brink, 1994] D. Brink. Moral conflict and its structure. *The Philosophical Review*, 103:215247, 1994.

[Broome, 1999] J. Broome. Normative requirements. *Ratio*, 12:398419, 1999.

[Broome, 2004] J. Broome. Reasons. In J. Wallace, M. Smith, and et al S. Scheffler, editors, *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, pages 28–55. Oxford University Press, Oxford, 2004.

[Brown, 1992] M. A. Brown. Normal bimodal logics of ability and action. *Studia Logica*, 51:519–532, 1992.

[Brown, 1996a] M. A. Brown. Doing as we ought: Towards a logic of simply dischargeable obligations. In M. A. Brown and J. Carmo, editors, *Deontic Logic, Agency and Normative Systems*, pages 47–65. Springer Verlag, New York, 1996.

[Brown, 1996b] M. A. Brown. A logic of comparative obligation. *Studia Logica*, 57:117–137, 1996.

[Bulygin, 1982] E. Bulygin. Norms, normative propositions, and legal statements. In G. Fløistad, editor, *Contemporary Philosophy. A New survey. Vol. 3: Philosophy of action*, pages 127–152. Martinus Nijhoff, The Hague, 1982.

[Bulygin, 1992] E. Bulygin. On norms of competence. *Law and Philosophy*, 11:201–216, 1992.

[Carmo and Jones, 2002] J. Carmo and A. Jones. Deontic logic and contrary-to-duties. In D. Gabbay and F. Gueunthner, editors, *Handbook of Philosophical Logic*, volume 8, pages 265–343. Kluwer Academic Publishers, Dordrecht, 2nd edition, 2002.

[Carnap, 1956] R. Carnap. *Meaning and Necessity: A Study in Semantics and Modal Logic*. The University of Chicago Press, Chicago, enlarged edition, 1956. First pub. date: 1947.

[Castañeda, 1970] H. N. Castañeda. On the semantics of the ought-to-do. *Synthese*, 21:449–468, 1970. Reprinted in D. Davidson and G. Harman (eds), *Semantics of Natural Language*, pp. 675-694. Dordrecht: D. Reidel, 1972.

[Castañeda, 1977] H. N. Castañeda. Ought, time, and the deontic paradoxes. *Journal of Philosophy*, 74:775–788, 1977.

[Castañeda, 1981] H. N. Castañeda. The paradoxes of deontic logic: The simplest solution to all of them in one fell swoop. In R. Hilpinen, editor, *New Studies in Deontic Logic: Norms, Actions and the Foundations of Ethics*, pages 37–85. Reidel, Dordrecht, 1981.

[Chellas, 1969] B. F. Chellas. *The Logical Form of Imperatives*. Dissertation, Perry Lane Press, Stanford, 1969.

[Chellas, 1974] B. F. Chellas. Conditional obligation. In S. Stenlund, editor, *Logical Theory and Semantic Analysis*, pages 23–33. D. Reidel, Dordrecht, 1974.

[Chellas, 1980] B. F. Chellas. *Modal Logic: An Introduction*. Cambridge University Press, Cambridge, 1980.

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

[Chisholm, 1964] R. M. Chisholm. The ethics of requirement. *American Philosophical Quarterly*, 1:147–153, 1964.

[Chisholm, 1982] R. M. Chisholm. *Brentano and Meinong Studies (Studien Zur Öesterreichischen Philosophie iii)*, chapter Supererogation and offence: A conceptual scheme for ethics, pages 98–113. Rodopi B.V., Amsterdam, 1982. Originally published in *Ratio*, 5, pp. 1-14, 1963.

[Cresswell, 1967] M. J. Cresswell. The interpretation of some Lewis systems of modal logic. *Australasian Journal of Philosophy*, 45(2):198–206, 1967.

[Czelakowski, 1997] J. Czelakowski. Action and deontology. In S. Lindström and E. Ejerhed, editors, *Logic, Action and Cognition*, pages 47–88. Kluwer Academic Publishers, Dordrecht and Boston, 1997.

[Da Costa and Carnielli, 1986] N. C. A. Da Costa and W. A. Carnielli. On paraconsistent deontic logic. *Philosophia*, 16:293–305, 1986.

[Danielsson, 1968] S. Danielsson. *Preference and Obligation: Studies in the Logic of Ethics*. Filosofiska föreningen, Uppsala, 1968.

[Davidson, 1980a] D. Davidson. Agency. In D. Davidson, editor, *Essays on Actions and Events*, pages 43–61. Clarendon Press, Oxford, 1980.

[Davidson, 1980b] D. Davidson. The logical form of action sentences. In D. Davidson, editor, *Essays on Actions and Events*, pages 105–122. Clarendon Press, Oxford, 1980. Originally published in N. Rescher (ed), *The Logic of Decision and Action*, Pittsburgh, University of Pittsburgh Press, 1967.

[DeCew, 1981] J. W. DeCew. Conditional obligation and counterfactuals. *Journal of Philosophical Logic*, 10:55–72, 1981.

[Dignum *et al.*, 1996] F. Dignum, J. J. C. Meyer, and R. Wieringa. Free choice and contextually permitted actions. *Studia Logica*, 57(1):193–220, 1996.

[Dubislav, 1937] W. Dubislav. Zur Unbegründbarkeit der Forderunssätze. *Theoria*, 3:330–342, 1937.

[Elgesem, 1993] D. Elgesem. *Action Theory and Modal Logic*. Ph.D. dissertation, University of Oslo, 1993.

[Elgesem, 1997] D. Elgesem. The modal logic of agency. *Nordic Journal of Philosophical Logic*, 2:1–46, 1997.

[Fehige, 1994] C. Fehige. The limit assumption in deontic (and prohairetic) logic. In G. Meggle and U. Wessels, editors, *Analyomen 1: Proceedings of the 1st Conference "Perspectives in Analytical Philosophy"*. Waiter de Gruyter, Berlin-New York, 1994.

[Feldman, 1986] F. Feldman. *Doing the Best We Can: An Essay in Informal Deontic Logic*. D. Reidel, Dordrecht, 1986.

[Føllesdal and Hilpinen, 1971] D. Føllesdal and R. Hilpinen. Deontic logic: An introduction. In R. Hilpinen, editor, *Deontic Logic: Introductory and Systematic Readings*, pages 1–35. D. Reidel, Dordrecht, 2nd edition, 1971.

[Forrester, 1984] J. W. Forrester. Gentle murder, or the adverbial samaritan. *Journal of Philosophy*, 81:193–196, 1984.

[Garson, 2001] J. Garson. Quantification in modal logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, pages 267–323. D. Reidel, Dordrecht, 2nd edition, 2001.

[Girle, 2009] R. Girle. *Modal Logics and Philosophy*. McGill-Queens University Press, Montreal and Kingston and Ithaca, 2nd edition, 2009.

[Goble, 1990a] L. Goble. A logic of good, should, and would: Part I. *Journal of Philosophical Logic*, 19(2):169–199, 1990.

[Goble, 1990b] L. Goble. A logic of good, should, and would: Part II. *Journal of Philosophical Logic*, 19:253–276, 1990.

[Goble, 1991] L. Goble. Murder most gentle: The paradox deepens. *Philosophical Studies*, 64(2):217–227, 1991.

[Goble, 1994] L. Goble. Quantified deontic logic with definite descriptions. *Logique et Analyse*, 37:239–253, 1994.

[Goble, 1996] L. Goble. 'Ought' and extensionality. *Noûs*, 30:330–355, 1996.

[Goble, 2003] L. Goble. Preference semantics for deontic logic part I- simple models. *Logique et Analyse*, 46(183-184):383–418, 2003.

[Goble, 2004] L. Goble. Preference semantics for deontic logic part II - multiplex models. *Logique et Analyse*, 47(335-363), 2004.

[Goble, 2005] L. Goble. A logic for deontic dilemmas. *Journal of Applied Logic*, 3:461–483, 2005.

[Goble, 2009] L. Goble. Normative conflicts and the logic of 'ought'. *Noûs*, 43(3):450–489, 2009.

[Goldman, 1976] H. S. Goldman. Dated rigthness and moral perfection. *The Philosophical Review*, 85(4):449–487, 1976.

[Greenspan, 1975] P. S. Greenspan. Conditional oughts and hypothetical imperatives. *Journal of Philosophy*, 72:259–276, 1975.

[Grelling, 1939] K. Grelling. Zur Logik der Sollsätze. *Unity of Science Forum*, January:44–47, 1939.

[Grue-Sörensen, 1939] K. Grue-Sörensen. Imperativsätze und Logik. Begegnung einer Kritik. *Theoria*, 5:195–202, 1939.

[Hamblin, 1987] C. L. Hamblin. *Imperatives*. Basil Blackwell, Oxford, 1987.

[Hansen, 2004] J. Hansen. Problems and results for logics about imperatives. *Journal of Applied Logic*, 2(1):39–61, 2004.

[Hansen, 2008] J. Hansen. *Imperatives and Deontic Logic: On the Semantic Foundations of Deontic Logic*. Dissertation (available from the author as a pdf), University of Leipzig, 2008.

[Hansson, 1971] B. Hansson. An analysis of some deontic logics. In R. Hilpinen, editor, *Deontic logic: Introductory and Systematic Readings*, pages 121–147. D. Reidel, Dordrecht, 1971. Reprinted from *Noûs* 3, pp. 373-98, 1969.

[Hansson, 1990] S. O. Hansson. Preference-based deontic logic (PDL). *Journal of Philosophical Logic*, 19(1):75–93, 1990.

[Hansson, 2001] S. O. Hansson. *Structures of Values and Norms*. Cambridge University Press, Cambridge, 2001.

[Hardin, 1968] G. Hardin. The tragedy of the commons. *Science*, 162(3859):1243 and 1248, 1968.

[Hart and Honoré, 1959] H. L. A. Hart and A. M. Honoré. *Causation in the Law*. Clarendon Press, Oxford, 1959.

[Hartmann, 1987] R. Hartmann. *Die Religion des Islam*. Wissenschaftliche Buchgesellschaft, Darmstadt, 1987. Originally published by Mittler and Sohn, Berlin, 1944.

[Hedenius, 1941] I. Hedenius. *Om Rätt och Moral ('On Law and Morals')*. Tidens Förlag, Stockholm, 1941.

[Henry, 1967] P. D. Henry. *The Logic of St. Anselm*. Clarendon Press, Oxford, 1967.

[Herrestad and Krogh, 1995] H. Herrestad and C. Krogh. Obligations directed from bearers to counterparts. In *Proceedings of the Fifth International Conference on Artificial Intelligence and Law (International Conference on Artificial Intelligence and Law)*, pages 210–218. ACM Press, College Park, Maryland, USA, 1995.

[Hilpinen, 1969] R. Hilpinen. An analysis of relativized modalities. In J. W. David, D.J. Hockney, and W.K. Wilson, editors, *Philosophical Logic*, Synthese library, pages 181–193. D. Reidel, Dordrecht, 1969.

[Hilpinen, 1974] R. Hilpinen. On the semantics of personal directives. In C. H. Heidrich, editor, *Semantics and Communication*, pages 162–179. North-Holland, Amsterdam, 1974.

[Hilpinen, 1993] R. Hilpinen. On deontic logic, pragmatics, and modality. In H. Stachowiak, editor, *Pragmatik: Handbuch Pragmatischen Denkens*. Band iv: *Sprachphilosophie, Sprachpragmatik und formative Pragmatik*, pages 295–319. Felix Meiner Verlag, Hamburg, 1993.

[Hilpinen, 1997] R. Hilpinen. On action and agency. In E. Ejerhed and S. Lindström, editors, *Logic, Action and Cognition - Essays in Philosophical Logic*, pages 3–27. Kluwer Academic Publishers, Dordrecht, 1997.

[Hilpinen, 2006] R. Hilpinen. Norms, normative utterances, and normative propositions. *Análisis Filosófico (Special Issue: Homenaje a Carlos E. Alchourrón)*, 26(2):229–241, 2006. Guest editors: E.Bulygin and G. Palau.

[Hintikka, 1957] J. Hintikka. Quantifiers in deontic logic. *Societas Scientiarum Fennica, Commentationes Humanarum Litterarum*, 23(4):1–23, 1957.

[Hintikka, 1971] J. Hintikka. Some main problems of deontic logic. In R. Hilpinen, editor, *Deontic Logic: Introductory and Systematic Readings*, pages 59–104. D. Reidel, Dordrecht, 1971.

[Hofstadter and McKinsey, 1939] A. Hofstadter and J. C. C. McKinsey. On the logic of imperatives. *Philosophy of Science*, 6:446–457, 1939.

[Hohfeld, 1919] W. N. Hohfeld. *Fundamental Legal Conceptions as Applied in Judicial Reasoning*. Yale University Press, New Haven, 1919. Edited by Walter Wheeler Cook.

[Horty, 1994] J. F. Horty. Moral dilemmas and nonmonotonic logic. *Journal of Philosophical Logic*, 23(1):35–65, 1994.

[Horty, 1996] J. F. Horty. Combining agency and obligation (preliminary version). In M. A. Brown and J. Carmo, editors, *Deontic Logic, Agency and Normative Systems*, pages 98–122. Springer Verlag, New York, 1996.

[Horty, 1997] J. F. Horty. Non-monotonic foundations for deontic logic. In D. Nute, editor, *Defeasible Deontic Logic*, pages 17–44. Kluwer Academic, Dordrecht, 1997.

[Horty, 2001] J. F. Horty. *Agency and Deontic Logic*. Oxford University Press, Oxford, 2001.

[Horty, 2003] J. F. Horty. Reasoning with moral conflicts. *Noûs*, 37(4):557–605, 2003.

[Hourani, 1975] G. F. Hourani. Ethics in medieval Islam: A conspectus. In G. Hourani, editor, *Essays on Islamic Philosophy and Science*, pages 128–135. State University of New York Press, Albany, 1975.

[Hourani, 1985] G. F. Hourani. *Reason and Tradition in Islamic Ethics*. Cambridge University Press, Cambridge, 1985.

[Hruschka, 1986] J. Hruschka. Das deontologische Sechseck bei Gottfried Achenwall im Jahre 1767. In *Joachim Jungius Gesellschaft der Wissenschaften e.V., Hamburg, 4 (1986) Heft 2*. Vandenhoeck and Ruprecht, Göttingen, 1986.

[Hursthouse, 1999] R. Hursthouse. *On Virtue Ethics*. Oxford University Press, Oxford, 1999.

[Jackson and Pargetter, 1986] F. Jackson and R. Pargetter. Oughts, options, and actualism. *Philosophical Review*, 95:233–255, 1986.

[Jackson, 1985] F. Jackson. On the semantics and logic of obligation. *Mind*, 94:177–195, 1985.

[Jackson, 1988] F. Jackson. Understanding the logic of obligation. *Aristotelian Society*, SUPP 62:255–270, 1988.

[Jennings, 1974] R. E. Jennings. A utilitarian semantics for deontic logic. *Journal of Philosophical Logic*, 3:445–456, 1974.

[Jones and Pörn, 1985] A. Jones and I. Pörn. Ideality, sub-ideality and deontic logic. *Synthese*, 65:275–290, 1985.

[Jones and Pörn, 1986] A. Jones and I. Pörn. 'Ought' and 'must'. *Synthese*, 66:89–93, 1986.

[Jones and Sergot, 1993] A. Jones and M. Sergot. On the characterization of law and computer systems: The normative systems perspective. In J.-J. C. Meyer and R. J. Wieringa, editors, *Deontic Logic in Computer Science*, pages 275–307. John Wiley and Sons, Chichester, 1993.

[Jones and Sergot, 1996] A. Jones and M. Sergot. A formal characterization of institutionalized power. *Logic Journal of the IGPL*, 4(3):429–445, 1996.

[Jones, 1990] A. Jones. Deontic logic and legal knowledge representation. *Ratio Juris*, 3(2):237–244, 1990.

[Jørgensen, 1937 and 1938] J. Jørgensen. Imperatives and logic. *Erkenntnis*, 7:288–296, 1937 and 1938.

[Jørgensen, 1938] J. Jørgensen. Imperativer og logik. *Theoria*, 4:183–90, 1938.

[Kamp, 1974] H. Kamp. Free choice permission. *Proceedings of the Aristotelian Society*, 74:57–74, 1974.

[Kamp, 1979] H. Kamp. Semantics versus pragmatics. In F. Guenthner and S. J. Schmidt, editors, *Formal Semantics and Pragmatics for Natural Languages*, pages 255–287. D. Reidel, Dordrecht, 1979.

[Kanger and Kanger, 2001] S. Kanger and H. Kanger. Rights and parliamentarism. In G. Holmström-Hintikka, S. Lindström, and R. Sliwinski, editors, *Collected Papers of Stig Kanger with Essays on his Life and Work*, volume I, pages 120–145. Kluwer Academic Publishers, Dordrecht, 2001. Originally published in 1966 in *Theoria* 32, pp. 85-115.

[Kanger, 1971] S. Kanger. New foundations for ethical theory. In R. Hilpinen, editor, *Deontic Logic: Introductory and Systematic Readings*, pages 36–58. D. Reidel, Dordrecht, 1971. Originally published in mimeographic form in 1957.

[Kanger, 1984] H. Kanger. *Human Rights in the U.N. Declaration (Acta Universitatis Upsaliensis)*. University of Uppsala, Uppsala, 1984.

[Kanger, 2001] S. Kanger. Law and logic. In G. Holmström-Hintikka, S. Lindström, and R. Sliwinski, editors, *Collected Papers of Stig Kanger with Essays on his Life and Work*, volume 1, pages 146–169. Kluwer Academic Publishers, Dordrecht, 2001. Originally published in 1972 in *Theoria* 38, pp. 105-132.

[Kelsen, 1967] H. Kelsen. *Pure Theory of Law*. University of California Press, Berkeley and Los Angeles, 1967. Translation by M. Knight of Reine Rechtslehre, 1960, 2nd ed., Verlag Franz Deuticke, Vienna.

[Kempson, 1977] R. M. Kempson. *Semantic Theory*. Cambridge texts in linguistics. Cambridge University Press, Cambridge, 1977.

[Knebel, 1991] S. K. Knebel. Necessitas moralis ad optimum. Zum historischen Hintergrund der Wahl der besten aller möglichen Welten. *Studia Leibnitiana*, 23:3–24, 1991.

[Knuuttila and Hallamaa, 1995] S. Knuuttila and O. Hallamaa. Roger Roseth and medieval deontic logic. *Logique et Analyse*, 38(149):75–87, 1995.

[Knuuttila, 1981] S. Knuuttila. The emergence of deontic logic in the fourteenth century. In R. Hilpinen, editor, *New Studies in Deontic Logic: Norms, Actions, and the Foundations of Ethics*, pages 225–248. Reidel, Dordrecht, 1981.

[Knuuttila, 1993] S. Knuuttila. *Modalities in Medieval Philosophy*. Routledge, London and New York, 1993.

[Kripke, 1963] S. Kripke. Semantical analysis of modal logic I: Normal modal propositional calculi. *Mathematical Logic Quarterly*, 9(5-6):67–96, 1963.

[Kripke, 1965] S. Kripke. Semantical analysis of modal logic II: Non-normal modal propositional calculi. In L. Hemkin, J.W. Addison, and A. Tarski, editors, *The Theory of Models (Proceedings of the 1963 International Symposium at Berkeley)*, pages 206–220. North Holland Publishing Co, Amsterdam, 1965.

[Krogh and Herrestad, 1996] C. Krogh and H. Herrestad. Getting personal: Some notes on the relationship between personal and impersonal obligation. In M. A. Brown and J. Carmo, editors, *Deontic Logic, Agency and Normative Systems*, pages 134–153. Springer Verlag, New York, 1996.

[Leibniz, 1930] G. W. Leibniz. Elementa juris naturalis. In *Sämtliche Schriften und Briefe. Sechste Reihe: Philosophische Schriften. Bd. 1*, pages 431–485. Otto Reichl Verlag, Darmstadt, 1930. First pub. date: 1671.

[Lemmon, 1962a] E. J. Lemmon. Moral dilemmas. *Philosophical Review*, 71:139–158, 1962.

[Lemmon, 1962b] E. J. Lemmon. On sentences verifiable by their use. *Analysis*, 22:86–89, 1962.

[Lewis, 1973] D. K. Lewis. *Counterfactuals*. Blackwell, Oxford, 1973.

[Lewis, 1974] D. K. Lewis. Semantic analyses for dyadic deontic logic. In S. Stenlund, editor, *Logical Theory and Semantic Analysis: Essays Dedicated to Stig Kanger on his Fiftieth Birthday*, pages 1–14. D.Reidel, Dordrecht, 1974.

[Lewis, 1978] D. K. Lewis. Reply to McMichael's "Too much of a good thing: A problem in deontic logic". *Analysis*, 38:85–86, 1978.

[Lewis, 1979] D.K. Lewis. A problem about permission. In E. Saarinen, R. Hilpinen, I. Niiniluoto, and M. Provence, editors, *Essays in Honour of Jaakko Hintikka on the Occasion of his Fiftieth Birthday*, pages 163–175. Reidel, Dordrecht, 1979.

[Lindahl and Odelstad, 2000] L. Lindahl and J. Odelstad. Normative systems represented by boolean quasi-orderings. *Nordic Journal of Philosophical Logic*, 5:161–174, 2000.

[Lindahl, 1977] L. Lindahl. *Position and Change*. D. Reidel, Dordrecht, 1977.

[Lindahl, 2001] L. Lindahl. Stig Kanger's theory of rights. In G. Holmström-Hintikka, S. Lindström, and R. Sliwinski, editors, *Collected Papers of Stig Kanger with Essays on his Life and Work*, pages 151–171. Kluwer Academic Publishers, Dordrecht, 2001. Originally published in D. Prawitz et al. (eds), *Logic, Methodology and Philosophy of Science IX*, 1994, Amsterdam: Elsevier Science B. V., pp. 889-911.

[Loewer and Belzer, 1983] B. Loewer and M. Belzer. Dyadic deontic detachment. *Synthese*, 54:295–318, 1983.

[Loewer and Belzer, 1986] B. Loewer and M. Belzer. Help for the good samaritan paradox. *Philosophical Studies*, 50:117–128, 1986.

[Loewer and Belzer, 1991] B. Loewer and M. Belzer. "Prima facie" obligation. In E. Lepore, editor, *John Searle and his Critics*, pages 359–370. Blackwell, Cambridge, 1991.

[Lokhorst and Goble, 2004] G.-J. Lokhorst and L. Goble. Mally's deontic logic. *Grazer Philosophische Studien*, 67:37–57, 2004.

[Lokhorst, 2006] G.-J. C. Lokhorst. Propositional quantifiers in deontic logic. In L. Goble and J. J. C. Meyer, editors, *Deontic Logic and Artificial Normative Systems (Deon 2006 Proceedings)*, volume 47 of *Lecture Notes in Artificial Intelligence*, pages 201–209. Springer-Verlag, Berlin and Heidelberg, 2006.

[Lokhorst, 2008] G.-J. Lokhorst. Mally's deontic logic. In E. Zalta, editor, *Stanford Encyclopedia of Philosophy*. 2008.

[Lyons, 1977] J. Lyons. *Introduction to Theoretical Linguistics*. Cambridge University Press, Cambridge, 1977.

[Makinson and van der Torre, 2000] D. Makinson and L. van der Torre. Input/output logics. *Journal of Philosophical Logic*, 29(4):383–408, 2000.

[Makinson and van der Torre, 2003] D. Makinson and L. van der Torre. What is input/output logic? In B. Löwe, W. Malzkorn, and T. Räsch, editors, *Foundations of the Formal Sciences II: Applications of Mathematical Logic in Philosophy and Linguistics*, volume 17 of *Trends in logic*, pages 163–174. Springer-Verlag, Berlin, 2003.

[Makinson, 1981] D. Makinson. Quantificational reefs in deontic waters. In R. Hilpinen, editor, *New Studies in Deontic Logic: Norms, Actions, and the Foundations of Ethics*, pages 87–91. Reidel, Dordrecht, 1981.

[Makinson, 1986] D. Makinson. On the formal representation of rights relations: Remarks on the work of Stig Kanger and Lars Lindahl. *Journal of Philosophical Logic*, 15:403–425, 1986.

[Makinson, 1993] D. Makinson. Five faces of minimality. *Studia Logica*, 52(3):339–379, 1993.

[Makinson, 1999] D. Makinson. On the fundamental problem of deontic logic. In P. McNamara and H. Prakken, editors, *Norms, Logic and Information Systems*, pages 29–53. IOS Press, Amsterdam, 1999.

[Mally, 1971] E. Mally. Elemente des Sollens. Grundgesetze der Logik des Willens. In K. Wolf and P. Weingartner, editors, *E. Mally, Logische Scriften. Grosses Logikfragment Grundgesetze des Sollens*, pages 229–324. D. Reidel, Dordrecht, 1971. Originally published in 1926, *Gurndgesetz Des Sollens. Elemente Der Logik Des Willens*. Graz, Leuschner and Lubensky.

[Marcus, 1980] R. B. Marcus. Moral dilemmas and consistency. *The Journal of Philosophy*, 77:121–136, 1980.

[Marcus, 1996] R. B. Marcus. More about moral dilemmas. In H. E. Mason, editor, *Moral Dilemmas and Moral Theory*, pages 23–35. Oxford University Press, New York, 1996.

[Mares and McNamara, 1997] E. D. Mares and P. McNamara. Supererogation in deontic logic: Metatheory for dwe and some close neighbours. *Studia Logica*, 59(3):397–415, 1997.

[Mates, 1965] B. Mates. *Elementary Logic*. Oxford University Press, Oxford, 1965.

[McCarty, 1992] L. T. McCarty. Defeasible deontic reasoning. In *Proceedings of the Fourth international Workshop on Nonmonotonic Reasoning*, pages 139–147, Plymouth, Vermont, USA, 1992.

[McConnell and Brue, 2002] C. R. McConnell and S. R. Brue. *Economics. Principles, Problems, and Policies*. McGraw-Hill, Boston, 2002.

[McMichael, 1978] A. McMichael. Too much of a good thing: A problem in deontic logic. *Analysis*, 38:83–84, 1978.

[McNamara, 1988] P. McNamara. A determination theorem for a von Wright inspired weakening of SDL. 1988.

[McNamara, 1990] P. McNamara. *The Deontic Quaddecagon*. Dissertation, University of Massachusetts, 1990.

[McNamara, 1995] P. McNamara. The confinement problem: How to terminate your mom and her trust. *Analysis*, 55(4):310–313, 1995.

[McNamara, 1996a] P. McNamara. Doing well enough: Toward a logic for commonsense morality. *Studia Logica*, 57(1):167–192, 1996.

[McNamara, 1996b] P. McNamara. Making room for going beyond the call. *Mind*, 105(419):415–450, 1996.

[McNamara, 1996c] P. McNamara. Must I do what I ought? (or will the least I can do do?). In M. A. Brown and J. Carmo, editors, *Deontic Logic, Agency and Normative Systems*, pages 154–173. Springer Verlag, New York, 1996.

[McNamara, 1999] P. McNamara. Doing well enough in an Andersonian-Kangerian framework. In P. McNamara and H. Prakken, editors, *Norms, Logics and Information Systems: New Studies in Deontic Logic and Computer Science*, pages 181–198. IOS Press, Washington, DC, 1999.

[McNamara, 2000] P. McNamara. Towards a framework for agency, inevitability, praise and blame. *Nordic Journal of Philosophical Logic*, 5:135–160, 2000.

[McNamara, 2004] P. McNamara. Agential obligation as non-agential personal obligation plus agency. *Journal of Applied Logic*, 2(1):117–152, 2004.

[McNamara, 2006] P. McNamara. Deontic logic. In D. Gabbay and J. Woods, editors, *The Handbook of the History of Logic*, volume 7, pages 197–288. Elsevier Press, Amsterdam, 2006.

[McNamara, 2010] P. McNamara. Deontic logic. In E. Zalta, editor, Stanford Encyclopedia of Philosophy. 2010. `http://plato.stanford.edu.archives/fall2010/entries/logic-deontic/`.

[McNamara, 2011a] P. McNamara. Praise, blame, obligation, and DWE: Toward a framework for the classical conception of supererogation and kin. *Journal of Applied Logic*, 9:153170, 2011.

[McNamara, 2011b] P. McNamara. Supererogation, inside and out: Toward an adequate scheme for common sense morality. In M. Timmons, editor, *Oxford Studies in Normative Ethics*, volume I, pages 202–235. Oxford University Press, Oxford, 2011.

[Meinong, 1968a] A. Meinong. Ethische Bausteine (Nacgelassenes Fragment). In R. Haller and R. Kindinger, editors, *Alexius Meinong Gesamtausgabe,* vol. III*, Abhandlungen zur Werttheorie*, pages 659–724. Akademische Druck- u. Verlagsanstalt, Graz, 1968. The page numbers in the text are those of the article not the page numbers of the *Gesamtausgabe*.

[Meinong, 1968b] A. Meinong. Psychologisch-ethische Untersuchungen zur Werth-Theorie. In R. Haller and R. Kindinger, editors, *Alexius Meinong Gesamtausgabe,* vol. III*, Abhandlungen zur Werttheorie*, pages 3–94. Akademische Druck- u. Velagsanstalt, Graz, 1968. Originally published by Leuschner and Lubensky, Graz, 1894. The page numbers in the text are those of the original publication, reproduced in the *Gesamtausgabe*.

[Menger, 1939] K. Menger. A logic of the doubtful: On optative and imperative logic. In K. Menger, editor, *Reports of a Mathematical Colloquium,* 2nd series, 2nd issue, pages 53–64. Indiana University Press, 1939.

[Meyer and Weiringer, 1993] J.-J. Meyer and R. J. Weiringer. *Deontic Logic in Computer Science: Normative System Specification*. Jone Wiley and Sons, Ltd., Chichester, 1993.

[Meyer, 1988] J.-J. Meyer. A different approach to deontic logic: Deontic logic viewed as a variant of dynamic logic. *Notre Dame Journal of Formal Logic*, 29:109–136, 1988.

[Montague, 1970] R. Montague. Universal grammar. *Theoria*, 36:373–98, 1970.

[Moreau, 1996] M. Moreau. Prima facie and seeming duties. *Studia Logica*, 57:47–71, 1996.

[Mott, 1973] P. L. Mott. On Chisholm's paradox. *Journal of Philosophical Logic*, 2:197–210, 1973.

[Mulligan *et al.*, 1984] K. Mulligan, P. Simons, and B. Smith. Truth-makers. *Philosophy and Phenomenological Research*, 44:287–321, 1984.

[Nowell Smith and Lemmon, 1960] P. H. Nowell Smith and E. J. Lemmon. Escapism: The logical basis of ethics. *Mind*, 69:289–300, 1960.

[Nute, 1997] D. Nute. *Defeasible Deontic Logic*. Synthese library, 263. Kluwer Academic, Dordrecht ; Boston, 1997.

[Parent, 2008] X. Parent. On the strong completeness of Åqvist's dyadic deontic logic G. In R. van der Meyden and L. van der Torre, editors, *Deontic Logic in Computer Science. 9th International Donference, Deon 2008*, volume 5076 of *Lecture Note in Computer Science*, pages 189–202, Berlin Heidelberg, 2008. Springer-Verlag.

[Parent, 2010] X. Parent. A complete axiom set for Hansson's deontic logic DSDL2. *Logic Journal of the IGPL*, 18(3):422–429, 2010.

[Parent, 2012] X. Parent. Why be afraid of identity? - Comments on Sergot and Prakken's views. In Artikis et al. [2012], pages 295–307.

[Peirce, 1998] C. S. Peirce. New elements. In N. Houser, J. R. Eller, and et al A. C. Lewis, editors, *The Essential Peirce. Selected Philosophical Writings, 1893-1913: Vol 2*, pages 300–324. Indiana University Press, Bloomington and Indianapolis, 1998. (Written in 1904).

[Pörn, 1977] I. Pörn. Action theory and social science: Some formal models. *Synthese Library*, 120, 1977.

[Pörn, 1989] I. Pörn. On the nature of a social order. In J. E. Fenstad, T. Frolov, and R. Hilpinen, editors, *Logic, Methodology and Philosophy of Science, VIII*, pages 553–67. Elsevier Science, New York, 1989.

[Powers, 1967] L. Powers. Some deontic logicians. *Noûs*, 1:381–400, 1967.

[Prakken and Sergot, 1994] H. Prakken and M. Sergot. Contrary-to-duty imperatives, defeasibility and violability. In A. Jones and M. Sergot, editors, *Proceedings of the Second International Workshop on Deontic Logic in Computer Science*, pages 296–318. Tano Publishers, Oslo, 1994.

[Prakken and Sergot, 1996] H. Prakken and M. Sergot. Contrary-to-duty obligations. *Studia Logica*, 57(1):91–115, 1996.

[Prakken and Sergot, 1997] H. Prakken and M. Sergot. Dyadic deontic logic and contrary-to-duty obligations. In D. Nute, editor, *Defeasible deontic logic*, Synthese library, pages 223–262. Kluwer Academic, Dordrecht-Boston, 1997.

[Prakken, 1996] H. Prakken. Two approaches to the formalisation of defeasible deontic reasoning. *Studia Logica*, 57(1):73–90, 1996.

[Priest, 2008] G. Priest. *An Introduction to Non-Classical Logics. From If to Is.* Cambridge University Press, Cambridge, second edition, 2008.

[Prior, 1954] A. N. Prior. The paradoxes of derived obligation. *Mind*, 63:64–65, 1954.

[Prior, 1955] A. N. Prior. *Formal logic*. Oxford University Press, Oxford, 1955. Second edition, 1962.

[Prior, 1958] A. N. Prior. Escapism: The logical basis of ethics. In A. I. Melden, editor, *Essays in Moral Philosophy*, pages 135–146. University of Washington Press, Seattle, 1958.

[Reach, 1939] K. Reach. Some comments on Grelling's paper 'zur logik der sollsatze'. *Unity of Science Forum*, April:72, 1939.

[Ross, 1939] W. D. Ross. *Foundations of Ethics*. Oxford University Press, Oxford, 1939.

[Ross, 1941] A. Ross. Imperatives and logic. *Theoria*, 7:53–71, 1941.

[Ross, 1944] A. Ross. Imperatives and logic. *Philosophy of Science*, 11:30–46, 1944.

[Ross, 1968] A. Ross. *Directive and Norms*. Routledge and Kegan Paul, London, 1968. B. Loar, editor.

[Ryu and Lee, 1991] Y. U. Ryu and R. M. Lee. Defeasible deontic reasoning: A logic programming model. In J.-J. Meyer and R. J. Weiringer, editors, *Deontic Logic in Computer Science: Normative System Specification*, pages 225–241. Jone Wiley and Sons, Ltd., Chichester, 1991.

[Sajama, 1988] S. Sajama. Meinong on the foundations of deontic logic. *Grazer Philosophische Studien*, 32:69–81, 1988.

[Sandu and Tuomela, 1996] G. Sandu and R. Tuomela. Joint action and group action made precise. *Synthese*, 105:319–345, 1996.

[Santos and Carmo, 1996] F. Santos and J. Carmo. Indirect action, influence and responsibility. In M. A. Brown and J. Carmos, editors, *Deontic Logic, Agency and Normative Systems*, pages 194–215. Springer Verlag, New York, 1996.

[Schotch and Jennings, 1981] P. K. Schotch and R. E. Jennings. Non-kripkean deontic logic. In R. Hilpinen, editor, *New Studies in Deontic Logic: Norms, Actions, and the Foundations of Ethics*, pages 149–162. Reidel, Dordrecht, 1981.

[Scott, 1970] D. Scott. Advice on modal logic. In K. Lambert, editor, *Philosophical Problems in Logic: Some Recent Developments*, pages 143–174. D. Reidel, Dordrecht, 1970.

[Segerberg, 1982] K. Segerberg. A deontic logic of action. *Studia Logica*, 41:269–282, 1982.

[Segerberg, 1990] K. Segerberg. Validity and satisfaction in imperative logic. *Notre Dame Journal of Formal Logic*, 31:203–221, 1990.

[Segerberg, 1992] K. Segerberg. Getting started: Beginnings in the logic of action. *Studia Logica*, 51:347–378, 1992.

[Sergot, 1999] M. Sergot. Normative positions. In P. McNamara and H. Prakken, editors, *Norms, Logic and Information Systems*, pages 289–308. IOS Press, Amsterdam, 1999.

[Sinnott-Armstrong, 1985] W. Sinnott-Armstrong. A solution to Forrester's paradox of gentle murder. *Journal of Philosophy*, 82:162–168, 1985.

[Sinnott-Armstrong, 1988] W. Sinnott-Armstrong. *Moral Dilemmas*. Basil Blackwell, Oxford, 1988.

[Slote, 1997] M. Slote. Agent-based virtue ethics. In R. Crisp and M. Slote, editors, *Virtue Ethics*. Oxford University Press, Oxford, 1997.

[Smith, 1993] T. Smith. Violation of norms. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*, pages 60 – 65. ACM, Amsterdam, The Netherlands, 1993.

[Smith, 1994] T. Smith. *Legal Expert Systems: Discussion of Theoretical Assumptions*. Dissertation, University of Utrecht, 1994.

[Spohn, 1975] W. Spohn. An analysis of Hansson's dyadic deontic logic. *Journal of Philosophical Logic*, 4:237–252, 1975.

[Stenius, 1963] E. Stenius. The principles of the logic of normative systems. In *Proceedings of a Colloquium on Modal and Many-Valued Logics, Helsinki, 23-26 August, 1962, Acta Philosophica Fennica*, volume 16, pages 247–260, 1963.

[Stout, 1980] R. Stout. *The League of Frightened Men*. Bantam Books, New York, 1980. First pub. date: 1935.

[Tannsjö, 2010] T. Tannsjö. *From Reasons to Norms. On the Basic Question of Ethics*. Springer, Dordrecht and New York, 2010.

[Thomason, 1981a] R. H. Thomason. Deontic logic and the role of freedom in moral deliberation. In R. Hilpinen, editor, *New Studies in Deontic Logic: Norms, Actions, and the Foundations of Ethics*, pages 177–186. Reidel, Dordrecht, 1981.

[Thomason, 1981b] R. H. Thomason. Deontic logic as founded on tense logic. In R. Hilpinen, editor, *New Studies in Deontic Logic: Norms, Actions, and the Foundations of Ethics*, pages 165–176. Reidel, Dordrecht, 1981.

[Thomson, 1971] J. Thomson. The time of a killing. *The Journal of Philosophy*, 68:115–132, 1971.

[Tomberlin, 1981] J. E. Tomberlin. Contrary-to-duty imperatives and conditional obligations. *Noûs*, 15:357–375, 1981.

[Urmson, 1958] J. O. Urmson. Saints and heroes. In A. I. Melden, editor, *Essays in Moral Philosophy*, pages 198–216. University of Washington Press, Seattle, 1958.

[van der Torre and Tan, 1997] L. van der Torre and Y. H. Tan. The many faces of defeasibility in defeasible deontic logic. In D. Nute, editor, *Defeasible Deontic Logic*, pages 79–122. Kluwer Academic, Dordrecht-Boston, 1997.

[van der Torre and Tan, 2000] L. van der Torre and Y. H. Tan. Two-phase deontic logic. *Logique et Analyse*, 171-172:411–456, 2000.

[van der Torre, 1997] L. van der Torre. *Reasoning about Obligations: Defeasibility in Preference-based Deontic Logic*. Dissertation, Tinbergen institute research series, Erasmus University Rotterdam, 1997.

[van Eck, 1982] J. van Eck. A system of temporally relative modal and deontic predicate logic and its philosophical applications. *Logique et Analyse*, 25:249–290, 339–381, 1982. Original publication, as dissertation, Groningen, University of Groningen, 1981.

[van Fraassen, 1972] B. C. van Fraassen. The logic of conditional obligation. *Journal of Philosophical Logic*, 1:417–438, 1972.

[van Fraassen, 1973] B. C. van Fraassen. Values and the heart's command. *Journal of Philosophy*, 70:5–19, 1973.

[von Wright, 1951] G. H. von Wright. Deontic logic. *Mind*, 60:1–15, 1951. Reprinted in [von Wright, 1957, pp. 58-74].

[von Wright, 1956] G. H. von Wright. A note on deontic logic and derived obligation. *Mind*, 65:507–509, 1956.

[von Wright, 1957] G. H. von Wright. *Logical Studies*. Routledge and Kegan Paul, London, 1957.

[von Wright, 1963] G. H. von Wright. *Norm and Action: A Logical Enquiry*. Humanities Press, New York, 1963.

[von Wright, 1964] G. H. von Wright. A new system of deontic logic. *Danish Yearbook of Philosophy*, 1:173–182, 1964. Reprinted in R. Hilpinen (ed), *Deontic Logic: Introductory and Systematic Readings*, D. Reidel, Dordrecht, 1971, pp. 105-115.

[von Wright, 1965] G. H. von Wright. A correction to a new system of deontic logic. *Danish Yearbook of Philosophy*, 2:103–107, 1965. Reprinted in R. Hilpinen (ed), *Deontic Logic: Introductory and Systematic Readings*, D. Reidel, Dordrecht, 1971, pp. 115-120.

[von Wright, 1968] G. H. von Wright. *An Essay in Deontic Logic and the General Theory of Action*, volume 21 of *Acta Philosophica Fennica*. North Holland Publishing Co., Amsterdam, 1968.

[von Wright, 1983] G. H. von Wright. Norms, truth, and logic. In *G. H. Von wright, Practical Reason. Philosophical Papers, vol. I*, pages 130–209. Cornell University Press, Ithaca, 1983.

[von Wright, 1996] G. H. von Wright. Is there a logic of norms? In *Six Essays in Philosophical Logic. Acta Philosophica Fennica 60*, pages 35–54. Societas Philosophica Fennica, Helsinki, 1996. Originally published in *Ratio Juris*, 4, 265-283, 1991.

[Vranas, 2010] P. Vranas. In defense of imperative inference. *Journal of Philosophical Logic*, 39:5971, 2010.

[Vranas, 2011] P. Vranas. New foundations for imperative logic: Pure imperative inference. *Mind*, 120:369–446, 2011.

[Weinberger, 1985] O. Weinberger. Freedom, range for action and the ontology of norms. *Synthese*, 65:307–324, 1985.

[Williams, 1965] B. Williams. Ethical consistency. *Proceedings of the Aristotelian Society*, Supp, vol. 39:103–124, 1965.

[Wolenski, 1990] J. Wolenski. Deontic logic and possible world semantics: A historical sketch. *Studia Logica*, pages 273–282, 1990.

[Xu, 1995] M. Xu. On the basic logic of Stit with a single agent. *Journal of Symbolic Logic*, 60(2):459–483, 1995.

[Zagzebski, 1996] L. Zagzebski. *Virtues of the Mind*. Cambridge University Press, New York, 1996.

Risto Hilpinen
University of Miami (Coral Gables)
Email: hilpinen@miami.edu

Paul McNamara
University of New Hampshire (Durham)
Email: paulm@unh.edu