Basic literature:
- Risto Hilpinen (ed.): *Deontic Logic: Introductory and systematic readings.* D. Reidel: Dordrecht, 1971.
- Risto Hilpinen (ed.): *New Studies in Deontic Logic.* D. Reidel: Dordrecht 1981,
- Lennart Åqvist: "Deontic Logic", in: Gabbay, D. & Guenthner, F. (eds.): *Handbook of Philosophical Logic*, 2nd ed., vol. 2, 147-264. Kluwer: Dordrecht 2002.
- José Carmo & Andrew Jones: "Deontic Logic and Contrary-to-duties", in: Gabbay, D. & Guenthner, F. (eds.): *Handbook of Philosophical Logic*, 2nd ed., vol. 2, 265-343. Kluwer: Dordrecht 2002.

Biannual Conference "Deontic Logic in Computer Science" (ΔEON), www.deonticlogic.org:

ΔEON '91: J.-J.Ch. Meyer, R.J. Wieringa (eds.), *Deontic Logic in Computer Science*. John Wiley & Sons: Chichester, 1993.
ΔEON '94: Studia Logica, DEON special edition, vol. 57 (1996).
ΔEON '96: Mark A. Brown, José Carmo (eds): *Deontic Logic, Agency and Normative Systems.* Springer: Berlin, 1996.
ΔEON '98: Paul McNamara, Henry Prakken (eds.): *Norms, Logics and Information Systems*. IOS Press: Amsterdam 1999.
ΔEON 2000: Fundamenta Informaticae, DEON special edition vol. 48 (2,3), 2001 and Nordic Journal of Philosophical Logic, DEON special issue, vol. 5 (2), 2000.
ΔEON 2002: Journal of Applied Logic, DEON special issue vol. 2(1), 2004.
ΔEON 2004: Alessio Lomuscio, Donald Nute (eds.): *Deontic Logic in Computer Science*, LNAI 3065, Springer: Berlin, 2004.
ΔEON 2006: Lou Goble, John-Jules Ch. Meyer (eds.): *Deontic Logic and Artificial Normative Systems*, LNAI 4048, Springer: Berlin, 2006.
ΔEON 2008: Ron van der Meyden, Leon van der Torre (eds.): *Deontic Logic in Computer Science*, LNAI 5076, Springer: Berlin, 2008.

Always worth to take a look at:

David Makinson: Bridges from Classical to Nonmonotonic Logic. Kings College Publications: London, 2005.
Alexander Bochman: A Logical Theory of Nonmonotonic Inference and Belief Change. Springer: Berlin, 2001.

Introduction: Norms, Imperatives, Deontic Sentences, and their Logic

I. Norms

- socially constructed objects
- created by agents (normgivers) to direct the behaviour of other agents (norm subjects) over which the former have authority
- in a narrow sense: obligating norms (*O*-norms) that make particular actions, or the realization of a particular state of affairs, obligatory for some agents (where a description of the action or state is sometimes called the 'content' of the *O*-norm); and (perhaps) also permissive norms (*P*-norms) that allow a certain actions, or the realization of some states of affairs for particular agents, even if this action/state of affairs would otherwise be forbidden

[*Problem*: do *P*-norms really exist, or is 'permission' a meta-concept (like contraction) that retracts certain *O*-norms, or adds exceptions to them, within an existing set of *O*-norms, in order to avoid certain deontic consequences (to ensure that the behaviour/state of affairs is no longer recognized as forbidden, but as permitted). Cf. Alchourrón/Bulygin: The Expressive Conception of Norms, in Hilpinen (1981), pp. 95-124.]

- in a wide sense: all concepts that influence the creation, use and expiration/withdrawal of norms, including e.g.: rules of norm creation, rules of norm interpretation (counts-as conditionals), rules that allocate, withdraw and limit the authority (power) of agents (norm-givers), norms that determine hierarchies of authorities, higher-order norms (principles) that do not limit powers/invalidate norms per-se, but may be appealed to in procedures to invalidate norms, etc.

II. Imperatives

- an expression in the grammatical form of a sentence in the imperative mood
- normally used by an agent to create a norm that a certain action is to be carried out / a certain state is to be realized by a certain agent or group of agents (a command)
- norms relate to imperatives in the same way that propositions relate to the (indicative) sentences that express it
- stand aside other expressions that may also be used to create norms, like orders ("I hereby order you to …"), requests ("I ask you (not) to …"), promises ("I hereby promise you that…"), reciprocal promises (contracts), etc.
- have a different 'direction of fit' than indicative sentences: indicatives may be used to indicate what the speaker believes the world to be like – if it is so, than the indicative is qualified as "true", otherwise it is "false"; imperatives may be used to indicate what the speaker wants another agent to do or realize – if the agent acts accordingly, then the action or state of affairs is qualified as "right", or 'satisfactory', otherwise it is "wrong" and constitutes a violation [G.E.M. Anscombe, *Intention*, Blackwell: Oxford, 1957, § 32].
- cannot be meaningfully termed as "true" or "false", but just as "satisfied" or "violated"
- may be successfully or unsuccessfully uttered (succeed in creating norms or not, depending on whether the speaker has authority); a norm is valid (exists) if the utterance was successful, otherwise the intended norm is not valid (does not exist)
- have internal structure (e.g. use of "Boolean" operators 'and', 'or', 'not'; quantifiers, conditionals)
- hypothesis: to each imperative corresponds an indicative that is true (false) if and only if (iff) the imperative is satisfied (violated) (these indicatives have e.g. been called "command termination statement", "actualization" of the imperative,

"Erfüllungsaussage", description of the "envisioned" state of affairs, etc.) (We write !*A* for an imperative with termination statement *A*)
- problem of the shopping list (problem of formalization): imperatives can be separable/inseparable and independent/dependent. Example: Shopping list that says: "Buy walnuts and apples", or "Buy walnuts. Buy apples": is buying walnuts still required if apples are out? Then we would have to speak of a "partial satisfiability"… Aim: formalization in such a way that imperatives are inseparable and independent (do not combine inseparably).
- problem of conditional imperatives (problem of formalization): we can only speak meaningfully of satisfaction/violation of the imperative if the antecedent is true. Example: "if it rains, take an umbrella": unless it rains, it makes no sense to say that the agent has satisfied or violated the imperative (the opportunity for satisfying or violating the imperative does not arise). Contraposition is unwanted, i.e. the above imperative does not demand the same as the imperative "if you do not take an umbrella, see to it that it does not rain".
- the formalization of imperatives may require additional features: the agent that uttered it (authority), the agents that are to realize what it demands (subjects), ranks or priorities, time stamps for features like "lex posteriori derogat lege priori" etc.

III. Deontic sentences ("normative propositions")

- statements that describe some action or the realization of some state of affairs as obligatory, permitted or prohibited
- examples: "it is obligatory that …", "it is permitted that …", "it is forbidden that …", "you ought to realize …", "you may bring it about that …", "you are forbidden to bring it about that …", "there exist norms according to which you must/may/must not do/realize…"
- are true or false depending on whether there exist norms that make the behaviour/state of affairs that is described as obligatory, permitted or forbidden so
- ambivalence: deontic sentences are not to be confused with similar expressions used with the intention to create norms ("it is forbidden to talk during the exam")
- Leibniz' discovery: the 'legal modalities' of what is obligatory, permitted and prohibited appear to be related in a similar manner to each other as the Aristotelian alethic modalities of what is necessary, possible and impossible
- since G. H. von Wright's essay "Deontic Logic" (Mind 60, 1951, 1-15) and monograph "An Essay in Modal Logic" (North Holland: Amsterdam, 1951) formalized using modal operators 'O', 'P', 'F', where $O\alpha$ means that an action α is obligatory or (more standardly) that the state of affairs described by the sentence α is to be realized
- interdefinability of operators: $PA := \neg O \neg A$, $FA := O \neg A$, $OA := \neg P \neg A$
- deviating approach: predicates over actions or states of affairs [Jürgen Rödig, Jaap Hage, Peter Philipp]
- equivalent approach: reduction to alethic modal logic. $OA := \Box (A \rightarrow S)$ or $OA := \Box (Q \rightarrow A)$, where S represents a 'sanction' or simply the fact that something 'bad' is the case or some norm is violated, and Q represents 'what morality prescribes' or simply that everything is 'right' or that no norm is violated [A.R. Anderson, "A Reduction of Deontic Logic to Alethic Modal Logic", Mind 67 (1958), 100-103; S. Kanger, "New Foundations for Ethical Theory: Part 1" in Hilpinen (1971), 36-58]

IV. Logic

1. Logic of Norms (?)

- logic := the study of valid and invalid arguments
- arguments = sentences, not objects of (socially constructed) reality, nonsensical: [the norm that demands *A*], ergo [the norm that demands *B*]
- at most: there exist norms that demand *A*, ergo there exist norms that demand *B* (which is an argument with deontic sentences)

2. Logic of Imperatives (?)

- logic := the study of valid and invalid arguments
- arguments consist of premises, a conclusion, and (possibly) intermediate steps designed to show that the conclusion follows from the premises
- valid argument := one for which it is not possible that all premises are true and the conclusion false
- imperatives cannot be true or false, ergo they cannot be premises or conclusions in an argument. So arguments with imperatives cannot exist, though many people seem to think they do. This is Jörgensen's dilemma [Jörgen Jörgensen, Imperatives and Logic, Erkenntnis 7, 1938, 288-296].
- examples of imperative arguments that have been considered valid: "Do this! If one does not do that one cannot do this. So do that!" (Poincaré); "Take all boxes to the station! This is one of the boxes. So take this box to the station!" (Hare)
- various proposals to redefine the definition of a valid argument to accommodate inference where some premises are sentences in the imperative mood and the conclusion may also be in the imperative mood
- logic of satisfaction: '$!A \therefore !B$' means "imperative !*A* is satisfied, ergo imperative !*B* is satisfied" [Hofstadter/McKinsey, On the Logic of Imperatives, *Philosophy of Science* 6 (1939), 446-457]. But on this reading '$!A \therefore A$' is also valid (imperative !*A* is satisfied, ergo *A* is true).
- logic of satisfiability: it is not possible that all imperative premises are satisfied and the imperative in the conclusion is not satisfied
- logic of validity: '$!A \therefore !B$' means "imperative !*A* creates a valid norm, ergo imperative !*B* creates a valid norm". Problems: a) this is not really an imperative, but rather an indicative (deontic) argument, b) norm validity presupposes facts (the uttering of a sentence by a certain agent in a specific situation) that cannot be logically inferred.
- formalistic approaches: '$!A_1, \dots !A_n,\ B_1, \dots, B_n \therefore !C$' is a valid imperative argument iff '$A_1, \dots A_n,\ B_1, \dots, B_n \therefore C$' is valid

- problem: certain conclusions must be ruled out, e.g. the conclusions that follows from the indicative premises alone
- paradoxes: from the imperative ‘post the letter’ derive the imperative ‘post the letter or burn it’ (Ross’s paradox; cf. A. Ross, “Imperatives and Logic”, Theoria 7, 1941, 53-71); from the imperative ‘take the parachute and jump out’ derive the imperative ‘jump out’ (Hare’s paradox; cf. R. M. Hare, “Imperative Sentences”, Mind 58, 1949, 21-39); the problem here is: someone who satisfies only the conclusion satisfies a (derived) imperative, though the action would not be naturally described as ‘right’ (satisfying an imperative). So derived imperatives are somewhat lesser imperatives?
- problem: do imperative arguments really exist in natural language discourses? Are the conjunctives ‘so’ and ‘therefore’ used in Poincaré’s and Hare’s examples really an appeal to logical capabilities of the participants in the discourse, or are they rather used to motivate the giving (creation) of a new imperative by an appeal to existing imperatives and a certain factual situation (this is e.g. claimed by G. A. Wedeking, “Are There Command Arguments?”, Analysis 30, 1970, 161-167; J. Harrison, “Deontic Logic and Imperative Logic”, in P. T. Geach, *Logic and Ethics*, Kluwer: Dordrecht 1991, 79-129)?
- the following should remain possible: a logically assisted *creation* of imperatives or licenses (by e.g. input/output logic). Examples: the automated creation of tax assessments from an input of the general tax laws and a tax declaration, the automated creation of parking permissions from an input of the rules for the parking lot and the insertion of coins into the machine, the automated creation of a letter (email) of contract from an internet login, the contents of a web page and the rules stated there, and a click on an ‘order’ button, etc. Thus interpreted input/output logic should not be confused with pure ‘logical’ reasoning: the obligation to pay taxes, the license to use the parking lot, the obligations due to my internet order etc. all did not ‘exist’ before the output was generated.

3. Deontic Logic

a) Preliminaries: Propositional Logic

<u>*Alphabet:*</u>
Set of proposition letters Prop={ $p_1, p_2, p_3, \ldots$}
Operators: $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$
Auxiliary signs: (,)

<u>*Language:*</u>
(a) All prosition letters belong to $\mathscr{L}_{\mathrm{PL}}$.
(b) If A and B belong to $\mathscr{L}_{\mathrm{PL}}$, then so do $\neg A$, $(A\wedge B)$, $(A\vee B)$, $(A\rightarrow B)$ and $(A\leftrightarrow B)$.
(c) (closure)

<u>*Semantics:*</u>

As usual: A *valuation* v is a function v: Prop $\rightarrow \{0,1\}$ that assigns any proposition letter a truth value represented by 0 or 1. We then define the *truth* of a PL-sentence recursively by the familiar clauses, so for all $p \in$ Prop and $A, B \in \mathscr{L}_{\mathrm{PL}}$:

(a)	p	is true	iff	$v(p_n) = 1$
(b)	$\neg A$	is true	iff	A is not true

etc. If a PL-sentence A is true given a valuation v we write $v \vDash_{\mathrm{PL}} A$, and if a PL-sentence A is true given *any* valuation v we write $\vDash_{\mathrm{PL}} A$ and call A a *tautology*. If there is no such valuation, we call A a *contradiction*. The sentence $\neg A$ is called the *opposite* of A. As usual, we use ‘$\top$’ as abbreviation for an arbitrary tautology, and ‘$\bot$’ as abbreviation for an arbitrary contradiction. A set $\Gamma = \{A_1, A_2, \ldots\} \subseteq \mathscr{L}_{\mathrm{PL}}$ of PL-sentences is called *satisfiable* iff there is a valuation v such that for any $A_i \in \Gamma$, $v \vDash_{\mathrm{PL}} A_i$, and Γ is then said to be *satisfied* by v. Likewise the set Γ is said to *entail* another sentence $A \in \mathscr{L}_{\mathrm{PL}}$ iff for every valuation v that satisfies Γ, $v \vDash_{\mathrm{PL}} A$ (we then write $\Gamma \vDash_{\mathrm{PL}} A$). A sentence A is called *contingent* iff both $\{A\}$ and $\{\neg A\}$ are satisfiable.

We suppose a sound and complete *calculus* of PL (e.g. Frege’s axioms and rules for *modus ponens* and substitution). If A is provable within this calculus, we write $\vdash_{\mathrm{PL}} A$.

b) Monadic Deontic Logic

<u>*Alphabet:*</u>
like that of $\mathscr{L}_{\mathrm{PL}}$, plus the additional *operator symbol: O*

<u>*Language:*</u>

(a) If $A \in \mathscr{L}_{\mathrm{PL}}$ then $A \in \mathscr{L}_{\mathrm{DL}}$ and $OA \in \mathscr{L}_{\mathrm{DL}}$.
(b) If $A, B \in \mathscr{L}_{\mathrm{DL}}$ then so are $\neg A$, $(A\&B)$, $(A\vee B)$, $(A\rightarrow B)$, and $(A\leftrightarrow B)$.
(c) (closure)

<u>*Definitions:*</u> $PA \overset{\mathrm{def}}{=} \neg O\neg A$ $\quad FA \overset{\mathrm{def}}{=} O\neg A$ $\quad IA \overset{\mathrm{def}}{=} PA \wedge P\neg A$

Note: here I use a language that contains ‘mixed’ expressions like $A \rightarrow OB$, but not ‘nested’ deontic operators as in $OOA\rightarrow OA$.

Informal semantics:

OA	it is obligatory that A	FA	it is forbidden that A
PA	it is permitted that A	IA	it is indifferent that A

Axiomatic system: Standard Deontic Logic (SDL)

SDL is the smalles set of DL-sentences such that

(a) SDL contains all $\mathscr{L}_{\mathrm{DL}}$-instances of PL-theorems as well as all PL-instances of the axiom schemes

(M) $O(A\wedge B) \rightarrow (OA \wedge OB)$
(C) $(OA \wedge OB) \rightarrow O(A\wedge B)$
(N) $O\top$
(D) $OA \rightarrow PA$

(b) SDL is closed under *modus ponens* and the 'rule of extensionality':

(Ext) If $\vdash_{\mathrm{PL}} A \leftrightarrow B$ then $OA \leftrightarrow OB \in$ SDL.

For $A \in$ SDL we write $\vdash_{\mathrm{SDL}} A$ and call A *provable* in SDL. (Observation: SDL is a normal modal logic of type KD.)

Standard possible worlds semantics:

A model M is a tripel $\langle W,R,V\rangle$, where

- $W\neq\varnothing$ (set of "possible worlds")
- $R\subseteq W\times W$ (binary relation; we write wRv for "v is a 'deontic alternative' for w")
- V: Prop$\times W\rightarrow\{1,0\}$ (valuation function that associates one truth value to any proposition letter in any world w)

Restriction for SDL-models: R is serial ($\forall w\in W{:}\exists v\in W{:}\ wRv$)

Truth definitions: (we write $M, w \vDash A$ for "A is true in the world w in the model M")
$M, w \vDash p$ iff $V(p,w)=1$,
$M, w \vDash \neg A$ iff not $M, w \vDash A$,
etc. for all Boolean operators, and
$M, w \vDash OA$ iff $\forall v\in W$: $(wRv \Rightarrow M, v \vDash A)$

A sentence A is SDL-satisfiable iff there exists a SDL-model $M=\langle W,R,V\rangle$ and a world $w\in W$ such that $M, w \vDash A$.
A sentence A is SDL-valid iff for all SDL-models $M=\langle W,R,V\rangle$ and all worlds $w\in W$, $M, w \vDash A$ (we write $\vDash_{\mathrm{SDL}} A$)
SDL is consistent, sound, strongly complete and decidable.

Some paradoxes of monadic deontic logic:

$OA\rightarrow O(A\vee B)$	(deontic version of Ross's paradox)
$PA\rightarrow P(A\vee B)$	(paradox of free choice permission)
$FA\rightarrow F(A\wedge B)$	(Good Samaritan or Penitent's paradox)
$\neg(OA \wedge O\neg A)$	("Sartre's Paradox")

c) Towards dyadic deontic logic: Prior's and Chisholm's paradoxes

The paradox of derived obligation (Prior's paradox)
(Arthur N. Prior: "The Paradoxes of Derived Obligation", Mind 63, 1954, 64-65)

Often it is the case that something (A) is forbidden, but that if it – contrary to duty – nevertheless obtains, there is some other obligation, e.g. that some act of reparation (B) be done. If $O\neg A$ expresses that there is a (primary) obligation not to realize A, how is the secondary obligation that in case A nevertheless obtains, it is true that B ought to be realized, to be formalized? It cannot be $O(A\rightarrow B)$ since we have

/PP) $O\neg A\rightarrow O(A\rightarrow B)$

for arbitrary B as an SDL-theorem. One may think the situation can be remedied by using the formalization $A\rightarrow OB$. However, the next paradox shows that this solution does not work:

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

Chisholm's paradox consists of the following sentences:

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

Observation: The four natural-language sentences appear to be (a) consistent and (b) independent from each other. From the sentences it appears we can conclude that (c) the man ought not to tell he is coming.

Formalization:

(1') Op
(2') $O(p\to q)$
(3') $\neg p\to O\neg q$ [since (3) is a 'contrary-to-duty obligation']
(4') $\neg p$

Problem: From (1') und (2') we can derive Oq in SDL. From (3') and (4') we obtain $O\neg q$ with modus ponens. With (C) we obtain $O(q\wedge\neg q)$. But this collides with (D). So (1') - (4') are SDL-inconsistent, which violates observation (a).

First correction: we replace (2') by

(2") $p\to Oq$

But (2") is follows deductively from (4'). This violates observation (b).

Second Correction: We replace (3') by

(3") $O(\neg p\to\neg q)$

But (3") derives from (1') in SDL (cf. Prior's paradox). This violates observation (b). Also, with such a formalization, we could no longer derive $O\neg q$ from the set of sentences, which violates obseration (c).

Further observation: Every set of norms that ties *specific* sanctions to the violation of each of its norms, i.e. a sanction that should be applied if this norm is violated, and should not be applied otherwise, makes true deontic sentences of the form (1) - (3).

Consequence: Neither $O(A\to B)$ nor $A\to OB$ are adequate formalizations for statements about commitments. Contrary-to-duty norms cannot adequately be formalized in SDL.

d) Dyadic deontic logic

Alphabet:
like that of $\mathcal{L}_{DL}$, plus the additional *auxiliary sign:* /

Language:
(a) If $A,B\in\mathcal{L}_{PL}$ then $O(A/B)\in\mathcal{L}_{DDL}$.
(b) If $A, B\in\mathcal{L}_{DDL}$ then so are $\neg A$, $(A\wedge B)$, $(A\vee B)$, $(A\to B)$, and $(A\leftrightarrow B)$.
(c) (closure)

Definitions: $P(A/B) \overset{def}{=} \neg O(\neg A/B)$ $F(A/B) \overset{def}{=} O(\neg A/B)$ $I(A/B) \overset{def}{=} P(A/B)\wedge P(\neg A/B)$

Informal semantics:
$O(A/B)$ it is obligatory that A given B $F(A/B)$ it is forbidden that A given B
$P(A/B)$ it is permitted that A given B $I(A/B)$ it is indifferent that A given B

Note: the language is restricted to fully modalized formulas, i.e. mixed expressions like $(C\wedge O(A/C))\to O(A/\top)$ are not allowed, and neither are nested deontic operators. A conservative extension of Hansson's dyadic deontic logic that allows for both is Åqvist's (2002) system G, but it requires an additional operator of universal necessity

Axiomatic system: Hansson's Dyadic Standard Deontic Logic (DSDL3)
DSDL3 is the smalles set of DDL-sentences such that

(a) DSDL3 contains all $\mathcal{L}_{DL}$-instances of PL-theorems as well as all PL-instances of the following axiom schemes:

(DM) $O(A\wedge B/C)\to(O(A/C)\wedge O(B/C))$
(DC) $(O(A/C)\wedge O(B/C))\to O(A\wedge B/C)$
(DN) $O(\top/C)$
(Cond) $O(A/C\wedge D)\to O(D\to A/C)$
(CCMon) $O(A\wedge D/C)\to O(A/C\wedge D)$
(RMon) $P(D/C)\to(O(A/C)\to O(A/C\wedge D))$

(b) SDL is closed under *modus ponens* and these rules:

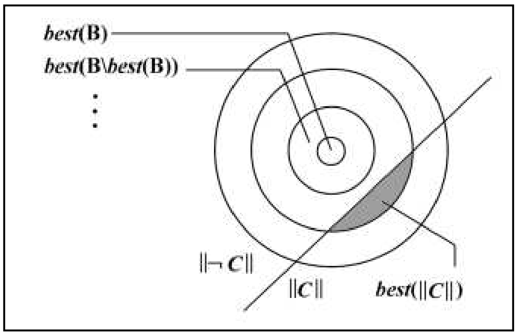
(CExt) If $\vdash_{PL} C\to(A\leftrightarrow B)$ then $O(A/C)\leftrightarrow O(B/C)\in$ DSDL3
(ExtC) If $\vdash_{PL}(C\leftrightarrow D)$ then $O(A/C)\leftrightarrow O(A/D)\in$ DSDL3
(DD-R) if not $\vdash_{PL}\neg C$ then $O(A/C)\to P(A/C)\in$ DSDL3

For $A\in$ DSDL3 we write $\vdash_{DSDL3} A$ and call A *provable* in DSDL3.

<u>*Hansson's preference semantics for dyadic deontic logic:*</u>

Motivation: "The problem is what happens if somebody commits a forbidden act. Ideal worlds are excluded. But it may be the case that among the still achievable worlds some are better than others. There should then be an obligation to make the best out of the sad circumstances." (Bengt Hansson: "An Analysis of Some Deontic Logics", *Noûs* 3, 1969, 373-398, reprinted in Hilpinen (1971) 121 –147.)

Let **B** be the set of all valuation functions v: Prop $\rightarrow$ {0,1}. We write $\|A\|$ for $\{v \in \mathbf{B} \mid v \vDash_{PL} A\}$.
Let $\leq \subseteq \mathbf{B} \times \mathbf{B}$ be a relation on the set **B**. Intuitively, $v' \leq v$ means that v is at least as good as v'. We assume that $\leq$ is



(a) reflexive, i.e. $v \leq v$
(b) transitive, i.e. if $v_1 \leq v_2$ and $v_2 \leq v_3$, then $v_1 \leq v_3$
(c) connected, i.e. for all v_1, $v_2 \in \mathbf{B}$: either $v_1 \leq v_2$ oder $v_2 \leq v_1$

and additionally satisfies the *Limit Assumption*:

(LA) If $\|A\| \neq \emptyset$ then $best(\|A\|) \neq \emptyset$

where $best(X)$ is the set of 'best' valuations of $X \subseteq \mathbf{B}$:

$$best(X) =_{def} \{v \in X \mid \text{for all } v', \text{ if } v' \in X, \text{ then } v' \leq v\}.$$

If $\leq$ is a relation as defined, we call $\leq$ a connected preference relation (CPR).

The crucial truth definition is then (the truth definitions for Boolean connectives being as usual):

$$\leq \vDash O(A/C) \quad \text{iff} \quad best(\|C\|) \subseteq \|A\|$$

i.e. A is obligatory given C iff A is true in the best worlds in which C is true.

A sentence A is DSDL3-satisfiable iff there exists a CPR $\leq$ such that $\leq \vDash A$.
A sentence A is DSDL3-valid iff for all CPRs $\leq$, $\leq \vDash A$ (we write $\vDash_{DSDL3} A$).
DSDL3 is consistent, sound, strongly complete and decidable (cf. Wolfgang Spohn: "An Analysis of Hansson's Dyadic Deontic Logic", *Journal of Philosophical Logic* 4, 1975, 237 - 252; Xavier Parent: "On the Strong Completeness of Åqvist's Dyadic Deontic Logic G", in: ΔEON'08, 289-202).

<u>*Paradoxes of dyadic deontic logic:*</u>

While dyadic deontic logic solves the problem of contrary-to-duty obligations, it is important to see that '$O(A/C)$' does *not* describe the existence of a conditional norm, created e.g. by use of a conditional imperative 'if C is the case then see to it that A'. If $O(A/C)$ is understood in this sense, then e.g. the theorem $O(A/C) \wedge P(\neg A/\top) \rightarrow O(\neg C/\neg A)$, derived by use of (Cond), (RMon) and (CExt), looks counterintuitive (example: if it rains you ought to take an umbrella, in the default circumstances you may take an umbrella (or not), ergo if you do not take an umbrella then you must see to it that it does not rain).

<u>***Homework:***</u>

Here is an alternative, norm- or imperative-based semantics for dyadic deontic logic:

Let $I \subseteq \mathscr{L}_{PL}$ be a subset of PL-sentences; they are meant to correspond to the termination statements of a set of (unconditional) imperatives, or the 'contents' of a set of unconditonal O-norms. Note that I is not assumed to be consistent.

Let $I \bot A$ (the A-remainders of I) be the set of maximal subsets Γ of I such that Γ does not derive A, i.e. $I \bot A$ contains all $\Gamma \subseteq I$ such that i) not $\Gamma \vdash_{PL} A$, and ii) there is no $\Delta \subseteq I$ such that $\Gamma \subseteq \Delta$ and not $\Delta \vdash_{PL} A$.

Consider the following truth definition for $O(A/C)$ (truth definitions for Boolean connectives being as usual):

$$I \vDash O(A/C) \quad \text{iff} \quad \text{for all } \Gamma \in I \bot \neg C\text{: } \Gamma \cup \{C\} \vdash_{PL} A$$

i.e. A is obligatory given C iff A derives – potentially with C – from a maximal set of contents of norms that still can be collectively satisfied in the circumstances described by C. Thus again we 'make the best out of the sad circumstances'.

Then the axiomatic system that is defined just like DSDL3, except without axiom _________, is sound and (weakly) complete with respect to this truth definition.

Exercise: provide the (only) DSDL3 axiom from the list above that is *not* valid under the new truth definition.

Hint: the resulting axiomatic system is equivalent to Kraus, Lehmann and Magidor's system P with the dyadic-deontic axiom DD-R added; cf. Kraus, Lehmann, Magidor: "Nonmonotic Reasoning, Preferential Models and Cumulative Logics", Artificial Intelligence 44, 1990, 167-207.