SVEN OVE HANSSON

PREFERENCE-BASED DEONTIC LOGIC (PDL)

Abstract. A new possible world semantics for deontic logic is proposed. Its intuitive basis is that prohibitive predicates (such as "wrong" and "prohibited") have the property of negativity, i.e. that what is worse than something wrong is itself wrong. The logic of prohibitive predicates is built on this property and on preference logic. Prescriptive predicates are defined in terms of prohibitive predicates, according to the well-known formula "ought" = "wrong that not". In this preference-based deontic logic (PDL), those theorems that give rise to the paradoxes of standard deontic logic (SDL) are not obtained. (E.g., $O(p \& q) \rightarrow Op \& Oq$ and $Op \rightarrow O(p \vee q)$) are theorems of SDL but not of PDL.) The more plausible theorems of SDL, however, can be derived in PDL.

1. INTRODUCTION

Modern deontic logic began with the introduction by von Wright (1951) of what is now called "standard deontic logic" (SDL). This system provides a logic for a predicate "O" for moral prescription, commonly read "it is obligatory that". The SDL logic for the O predicate is very strong, and much of the efforts devoted to deontic logic have been concerned with the paradoxes that arise in SDL and related systems. (Cf. Føllesdal and Hilpinen 1970, for an overview.) This situation seems to depend, to a large degree, on the lack of a credible semantical basis for a weaker deontic logic that avoids the paradoxes of SDL. The purpose of the present article is to provide such a semantical basis.

2. A SEMANTICAL FRAMEWORK

2.1 *States of Affairs*

The formal language will be taken to contain expressions for (factual) states of affairs. Classical propositional logic and the intersubstitutivity of logically equivalent expressions will be assumed. A consequence

Journal of Philosophical Logic **19**: 75–93, 1990.

relation (Cn) will be introduced to represent logical consequence and to define consistency.

Among the expressions for states of affairs are some that represent human actions. Normative concepts such as "ought" and "wrong" will be assumed to refer exclusively to human actions in the terminology of Oppenheimer (1961, pp. 15–23), i.e. to human will-controlled behaviour. (A normative statement that refers to a state of mind can be construed as referring to will-controlled behaviour that brings about that state of mind.)

DEFINITION D1. The *factual language* L is a set of propositions that is closed under the operations of classical propositional logic.

There is a set I of *individuals*. Some of the elements of L are *action statements*. Each of them represents an action, i.e. will-controlled behaviour, by a specified element or subset of I.

The consequence operator Cn on L is a function from $\mathcal{P}(\mathrm{L})$ to $\mathcal{P}(\mathrm{L})$ such that for all subsets S and T of L and elements p and q of L:

(1) $S \subseteq \mathrm{Cn}(S)$ (inclusion)

(2) $\mathrm{Cn}(S) = \mathrm{Cn}(\mathrm{Cn}(S))$ (iteration)

(3) If $S \subseteq T$, then $\mathrm{Cn}(S) \subseteq \mathrm{Cn}(T)$ (monotony)

(4) If p can be derived from S by classical propositional logic, then $p \in \mathrm{Cn}(S)$.

(5) If $(p \rightarrow q) \in \mathrm{Cn}(S)$, then $q \in \mathrm{Cn}(S \cup \{p\})$ (deduction)

A set $S \subseteq \mathrm{L}$ is *consistent* iff there is no element $p \in \mathrm{L}$ such that both $p \in \mathrm{Cn}(S)$ and $-p \in \mathrm{Cn}(S)$. An element p of L is consistent iff $\{p\}$ is consistent.

The precise logical form of the action statements need not be specified for the present purposes. One of the possible forms is $D_i p$, "i brings it about that p". (Kanger 1957).

2.2 *Perspectives and Alternatives*

In answering the question "What should the person i do?", we see the world from the *perspective* of i. Only such states of affairs that concern actions by i can be part of the answer to this question. The rest of the states of affairs are not considered to be variables, since they are uncontrollable from the perspective of i.

A moral perspective may comprise the actions of more than one individual. In answering the question "What should i and j do?", we take those states of affairs into moral consideration that represent actions by i and/or j.

The normative status of an action may very well vary with the choice of a moral perspective. Actions by i that are morally required from the perspective of i may not necessarily be morally required from the perspective of i-and-j. The present article will deal with normative statements as seen from a *fixed perspective*, that may be either individual or collective. The relationships between moral prescriptions from different perspectives will not be treated here.

DEFINITION D2. Let G be a subset of I. Then a *maximal action set* X for G is a set whose logical closure $\mathrm{Cn}(X)$ is a maximal consistent set of such action statements in L that represent actions by elements or subsets of G.

Further, an *alternative set* A for G is a non-empty set of maximal action sets for G.

Let i be an element of I. Then X is a maximal action set for i iff it is a maximal action set for $\{i\}$. Further, A is an alternative set for i iff it is an alternative set for $\{i\}$.

An alternative set for G may consist of all maximal action sets for G. It may also consist of all feasible or practically possible maximal action sets for G. (In the former case, practicability may be embedded in the consequence relation Cn.)

In what follows, normative discourse will be assumed to presuppose an alternative set A. Intuitively speaking, this set consists of all the completely described alternatives that are under moral consideration.

2.3 *Preferences*

Before introducing normative predicates (such as "ought", "wrong", etc.), a dyadic value predicate for betterness (of actions) will be introduced. The common concept of betterness refers to single actions, rather than to completely specified alternatives. In spite of this, the betterness relation on single actions will be based on a betterness relation for completely described alternatives. The reason for this construction is that it provides a more solid basis for the formal system. Preferences between completely described alternatives depend on less uncertain intuitions, since the problems connected with comparisons of negated and disjunctive states of affairs do not arise. (Hansson 1989).

In addition to a relation of "better" (strict preference), a relation of "equal in value to" (indifference) will be needed. These are the two basic comparative value concepts. (Halldén 1957, p. 10). However, the most economical axiomatic approach is not to introduce them both directly. Instead, the pre-ordering "better or equal in value to" (weak preference) will be used as a primitive relation in the standard manner. (Sen 1970, pp. 7–11. Cf. Bengt Hansson 1968, pp. 426–427.)

DEFINITION D3. A *weak holistic preference relation* on an alternative set A is a reflexive and transitive relation $\geqslant$ on A, such that if $\mathrm{Cn}(X) = \mathrm{Cn}(Y)$, then $X \geqslant Y$.

$>$ is a strict preference relation, such that for all elements X and Y of A, $X > Y$ iff $X \geqslant Y$ and not $Y \geqslant X$.

$\equiv$ is an indifference relation, such that for all elements X and Y of A, $X \equiv Y$ iff $X \geqslant Y$ and $Y \geqslant X$.

$\geqslant$ is read "better than or equal in value to", $>$ "better than", and $\equiv$ "equal in value to".

It should be noted that $\geqslant$ need not be connected, i.e. the definition does not prescribe that $X \geqslant Y \vee Y \geqslant X$ for all X and Y in A.

A comparison between two single actions may yield different results depending on the rest of the actions performed. Thus, the action p may be better than the action q if, for the rest, the actions of the maximal action set X_1 are performed, whereas q is better than p if for the rest the actions of X_2 are performed. The relation R_x is intended to capture this relative notion of betterness.

DEFINITION D4. (1) An element Y of A is an *X-close representation* of p iff $p \in \mathrm{Cn}(Y)$, and there is no Y' such that $p \in \mathrm{Cn}(Y')$ and $Y \cap X \subset Y' \cap X$.

(2) The ordered pair $\langle Y, Z \rangle$ is a *maximally alike X-close representation* of $\langle p, q \rangle$ iff Y is an X-close representation of p, Z is an X-close representation of q and there is no ordered pair $\langle Y', Z' \rangle$ such that Y' is an X-close representation of p, Z' is an X-close representation of q and $Y \cap Z \cap X \subset Y' \cap Z' \cap X$.

DEFINITION D5. Let X be an element of A. Then the preference relation R_x is defined as follows:

pR_xq iff for all $Y, Z \in A$: if $\langle Y, Z \rangle$ is a maximally alike X-close representation of $\langle p, q \rangle$, then $Y \geqslant Z$.

pR_xq iff pR_xq and not qR_xp

pI_xq iff pR_xq and qR_xp.

pR_xq is read "In (the maximal action set) X, p is better than or equal in value to q". pP_xq is read "In (the maximal action set) X, p is better than q". pI_xq is read "In (the maximal action set) X, p is equal in value to q".

According to definition D5, for p to be at least as good as q in X, it must be the case that a maximal action set Y is always at least as good as a maximal action set Z if (1) Y is the result of a minimal change of X to make p true (Y is an X-close representation of p) (2) Z is the result of a minimal change of X to make q true (Z is an X-close representation of q), and (3) given this, Y and Z are as similar as possible (maximally alike).

This "local" preference relation may be taken as a starting-point for developing global preference relations ("in all maximal action sets, p is better than q"). This, however, will not be needed for our present purposes.

As is stated in the following theorem, the relation R_x is reflexive. For some other logical properties of preference relations of this type, see Hansson (1989).

THEOREM T1. *pR_xp for all p and all $X \in A$.*

The proofs of this and the following theorems are given in the appendix.

2.4 *Normative Predicates*

There are three major classes of normative predicates, namely those that are prescriptive ("ought"), prohibitive ("wrong"), and permissive ("allowed"). In each of these classes, there may be predicates of different strengths (as in the prescriptive case: "advisable" – "ought" – "must").

Let W be a prohibitive predicate, and let p and q be two elements of its domain. If W holds for p, and p is better than or equal in value to q, then W also holds for p. As an example, what is worse than something wrong is also wrong. This property will be called *negativity* (Hansson, 1988). Among the normative predicates, it will be taken as the defining characteristic of prohibitive predicates.

DEFINITION D6. A monadic predicate W over the alternative set A has the property of *negativity* if and only if: Wq holds for all q such that for all $X \in A$ there is a p such that Wp and pR_xq.

A normative predicate W over A that has the property of negativity is also called a *prohibitive* predicate over A.

Prescriptive predicates do not have the corresponding property at the opposite end of the value scale. In other words: What is better than something morally required is not necessarily morally required. This can be seen from many examples of supererogatory actions. Instead, however, prescriptive predicates can be introduced as converses of prohibitive predicates. E.g., "ought" may be defined as "wrong that not".

DEFINITION D7. A monadic predicate O is *prescriptive* iff there is a negative predicate W such that for all p, Op iff $W - p$. Then W is the *converse* predicate of O.

Similarly, permissive predicates may be defined as negations of the converses of prohibitive predicates. Thus, "allowed" may be defined as "not wrong that not".

3. THE LOGIC OF PRESCRIPTIONS

The following theorem holds in general for prescriptive predicates as of definition D7.

THEOREM T2. *Let O be a prescriptive predicate. Then for all* p:

(1) $Op \,\&\, Oq \rightarrow O(p \,\&\, q)$.

(2) $Op \,\&\, O(p \rightarrow q) \rightarrow O(p \,\&\, q)$.

(3) $Op \,\&\, Oq \rightarrow O(p \vee q)$.

Further results can be obtained if it is assumed that, intuitively speaking, it is an option to do nothing wrong. In the formal language, this assumption will be represented by the property of the predicate to be *obeyable*.

DEFINITION D8. A prohibitive predicate W is *obeyable* iff there is at least one $X \in A$ such that $-Wp$ holds for all $p \in \mathrm{Cn}(X)$. A prescriptive predicate is obeyable iff its converse negative predicate is obeyable.

THEOREM T3. *Let O be a prescriptive predicate. Then*:

(1) $-O(p \,\&\, -p)$ *holds iff O is obeyable.*

(2) *If O is obeyable, then* $Op \rightarrow -O - p$ *holds for all* p.

The logic of obeyable prescriptive predicates will be called *preference-based deontic logic* (PDL).

Stronger logical principles will be obtained if the further assumption is added that, intuitively speaking, whatever only occurs together with something wrong is itself wrong. This will be called the *association principle*.

DEFINITION D9. Let W be any monadic predicate. Then p is *W-associated* iff for all $X \in A$, if $p \in \mathrm{Cn}(X)$, then there is a $q \in \mathrm{Cn}(X)$ such that Wq.

A prohibitive predicate W follows the *association principle* iff: If p is W-associated, then Wp. A prescriptive predicate O follows the association principle iff its converse negative predicate follows the association principle.

THEOREM T4. *Let O be a prescriptive predicate that follows the association principle. Then for all p:*

(1) *If $p \rightarrow q$ is true in all elements of A, then $Op \rightarrow Oq$.*

(2) *If $p \rightarrow q$ is a theorem, then $Op \rightarrow Oq$.*

(3) $O(p \,\&\, q) \rightarrow Op \,\&\, Oq$.

(4) $Op \rightarrow O(p \vee q)$.

(5) $O(p \vee -p)$.

THEOREM T5. *Let O be a prescriptive predicate over a finite alternative set A. Then the association principle holds iff it holds that: If $p \rightarrow q$ is true in all elements of A, then $Op \rightarrow Oq$.*

Standard deontic logic follows from (1) of theorem T2, (2) of theorem T3 as applied to an obeyable predicate, and (3) and (5) of theorem T4. (Føllesdal and Hilpinen 1970, p. 13). Thus, the addition to PDL of the principle of association yields a complete SDL.

Parts (2)–(4) of theorem T4 are all closely connected to the well-known paradoxes of standard deontic logic. None of them is derivable in PDL.

Part (3) of theorem T4 may be called the *principle of division* of duties. From it follows that an obligation to a whole implies an independent obligation to every part of it. The implausibility of this principle was indicated in informal philosophy before the advent of modern deontic logic, for instance by Menger (1934, quoted by Ross 1941, p. 68). Weinberger (1970) and others have argued against its inclusion in deontic logic. As was pointed out by Stranzinger (1978), several of the paradoxes of deontic logic depend on this principle. Von Wright, too, has concluded that "in a deontic logic which rejects the implication from left to right in the equivalence $O(p \,\&\, q) \equiv Op \,\&\, Oq$ while retaining the implication from right to left, the 'paradoxes' would not appear." (von Wright 1981, p. 7).

The strong standing of the principle of division in traditional deontic logic does not seem to be based on considerations of the logic of moral discourse. One source for belief in this principle is the analogy with alethic modal logic (where $\Box(A \,\&\, B) \rightarrow \Box A$ is accepted for good reasons). Another, closely related, source is the fact that the principle of division follows from the semantic principle of "deontically perfect worlds" that has been taken to be the only plausible basis for a possible world semantics for deontic concepts.

Part (2) of theorem T4 was called by Vermazen (1977) the *principle of inheritance* of obligations. It refers to situations "where the attempt to do one thing unavoidably involves one in an attempt to do something else, and where consequently an 'ought' attached to the first action is inherited by the second." (p. 14) It logically implies the principle of division. Even apart from this, it is a highly controversial deontic principle. "[T]he fact that we can't help but bring about the necessary consequences of our action does not mean we have an *obligation* to bring them about." (Sayre-McCord 1986, p. 188).

Part (4) of theorem T4 may be called the *principle of disjunctive extension*. It gives rise to Alf Ross's paradox: If I ought to mail a letter, then I ought to mail or burn it. (Ross 1941) It is, again, a principle whose exclusion from PDL should be welcomed.

In SDL, norms are assumed to refer exclusively to what obtains in the best possible alternatives. As was noted by Dayton (1981, p. 138), "[t]he ideal standard for a given world is an *axiological* standard, not a deontological one: worlds [in deontic logic] are ideals in the sense of being subjunctively *best* worlds." In SDL, only what is compatible with the best is not wrong. This property will be called *perfectionism*.

DEFINITION D10. A prohibitive predicate W fulfils *perfectionism* iff:

> Wp iff for all $X \in A$, if $p \in \mathrm{Cn}(X)$ then there is $Y \in A$ such that $Y > X$.

A prescriptive predicate fulfils perfectionism iff its converse prohibitive predicate fulfils perfectionism.

THEOREM T6. *Perfectionism implies the association principle.*

Perfectionism, as of definition D10, does not imply the existence of at least one best alternative. This property will be introduced as follows:

DEFINITION D11. The *upper limit-assumption* is fulfilled in the alternative set A iff there is at least one element $X \in A$ such that $Y > X$ does not hold for any $Y \in A$.

(Definition D11 yields a weaker condition than that of a "limited value structure" according to Lewis (1974, p. 5) or the "limit-assumption" of von Kutschera (1975, p. 204). These concepts exclude the existence of any infinite sequence of better and better alternatives, whereas definition D11 merely states that there is at least one alternative that is not an element of such a sequence.)

THEOREM T7. *Let O be a prescriptive predicate that fulfils perfectionism and the upper limit-assumption. Then it is obeyable.*

It follows from theorems T6 and T7 that SDL holds for all prescriptive predicates that fulfil perfectionism and the upper limit-assumption.

4. CONCLUSION

Preference-based deontic logic (PDL), is based on plausible semantical principles, mainly on the negativity of prohibitive predicates. The theorems that can be derived from PDL do not give rise to any of the well-known paradoxes of standard deontic logic (SDL). It is proposed, therefore, that PDL reflects the structure of ordinary deontic discourse better than does SDL.

APPENDIX: PROOFS

Before the proofs of the theorems, two definitions (D12–D13) and four lemmas (L1–L4) will be given. *Definitions D12 and D13* and *Lemma L1* will be used in the rest of the appendix without explicit reference.

DEFINITION D12. (Alchourrón and Makinson 1981). Let X and Y be two sets of propositions. Then $X \perp Y$ is the set of all consistent subsets Z of X such that $Y \cap \text{Cn}(Z) = \varnothing$ and that there is no consistent set Z' such that $Z \subset Z' \subseteq X$ and $Y \cap \text{Cn}(Z') = \varnothing$.

DEFINITION D13. Given an alternative set A:

(1) Cn_A is an operation on sets of propositions such that $p \in \text{Cn}_A(S)$ iff for all $X \in A$, if $S \subseteq \text{Cn}(X)$, then $p \in \text{Cn}(X)$.

(2) A set S is *A-consistent* iff it is a subset of $\text{Cn}(X)$ for some element X of A. It is *A-inconsistent* iff it is not A-consistent.

(3) $S \perp_A T$ is the set of A-consistent subsets V of S such that (1) $T \cap \text{Cn}_A(V) = \varnothing$, and (2) there is no A-consistent set V' such that $V \subset V' \subseteq S$ and $T' \cap \text{Cn}_A(V) = \varnothing$.

LEMMA L1. *Let A be an alternative set. Then Cn_A fulfils the following properties*:

(1) *If $p \in Cn(S)$, then $p \in Cn_A(S)$.*

(2) *$S \subseteq Cn_A(S)$ (inclusion).*

(3) *$Cn_A(S) = Cn_A(Cn_A(S))$ (iteration).*

(4) *If $S \subseteq T$, then $Cn_A(S) \subseteq Cn_A(T)$ (monotony).*

(5) *$(p \rightarrow q) \in Cn_A(S)$ iff $q \in Cn_A(S \cup \{p\})$ (deduction).*

It follows from parts (2)–(4) of the lemma that Cn_A is a consequence operator.

Proof of Lemma L1. (1) Suppose that this part of the lemma does not hold. Then there are p and S such that $p \in \text{Cn}(S)$ and $p \notin \text{Cn}_A(S)$.

The latter means that there must be an $X \in A$ such that $S \subseteq X$ and $p \notin \mathrm{Cn}(X)$, which is impossible.

(2) Suppose this does not hold. Then there must be a p such that $p \in S$ and $p \notin \mathrm{Cn}_A(S)$. From $p \notin \mathrm{Cn}_A(S)$ follows that there is an element X of A such that $S \subseteq X$ and $p \notin \mathrm{Cn}(X)$, which is not possible.

(3) Since $\mathrm{Cn}_A(S) \subseteq \mathrm{Cn}_A(\mathrm{Cn}_A(S))$ follows from part 2 of the present lemma, it remains to show that $\mathrm{Cn}_A(\mathrm{Cn}_A(S)) \subseteq \mathrm{Cn}_A(S)$, i.e. that for all p, if $p \in \mathrm{Cn}_A(\mathrm{Cn}_A(S))$, then $p \in \mathrm{Cn}_A(S)$.

Let p be an element of $\mathrm{Cn}_A(\mathrm{Cn}_A(S))$, and let X be an element of A such that $S \subseteq \mathrm{Cn}(X)$. It follows from $S \subseteq \mathrm{Cn}(X)$ and the definition of Cn_A that $\mathrm{Cn}_A(S) \subseteq \mathrm{Cn}(X)$. Further, since $p \in \mathrm{Cn}_A(\mathrm{Cn}_A(S))$ it follows from $\mathrm{Cn}_A(S) \subseteq \mathrm{Cn}(X)$ that $p \in \mathrm{Cn}(X)$. Thus, for all X such that $S \subseteq \mathrm{Cn}(X)$ it follows that $p \in \mathrm{Cn}(X)$. Then $p \in \mathrm{Cn}_A(S)$.

(4) Let p be an element of $\mathrm{Cn}_A(S)$. Then for all $X \in A$, $S \subseteq \mathrm{Cn}(X)$ implies $p \in \mathrm{Cn}(X)$. Since $S \subseteq T$ if follows that for all $X \in A$, $T \subseteq \mathrm{Cn}(X)$ implies $p \in \mathrm{Cn}(X)$. Thus $p \in \mathrm{Cn}_A(T)$.

(5) First suppose $(p \rightarrow q) \in \mathrm{Cn}_A(S)$. Then for all $X \in A$, $S \subseteq \mathrm{Cn}(X)$ implies $(p \rightarrow q) \in \mathrm{Cn}(X)$. Now suppose q is not an element of $\mathrm{Cn}_A(S \cup \{p\})$. Then there is an $X \in A$ such that $(S \cup \{p\}) \subseteq \mathrm{Cn}(X)$ and $q \notin \mathrm{Cn}(X)$. Since $\mathrm{Cn}(X)$ is inclusion-maximal, from $q \notin \mathrm{Cn}(X)$ follows $-q \in \mathrm{Cn}(X)$. Thus $\{p, -q, p \rightarrow q\} \subseteq \mathrm{Cn}(X)$, making X inconsistent, contrary to the conditions.

Next, suppose $q \in \mathrm{Cn}_A(S \cup \{p\})$. Let X be any element of A such that $S \subseteq \mathrm{Cn}(X)$. From $q \in \mathrm{Cn}_A(S \cup \{p\})$ and the consistency of X follows that $(p \,\&\, -q) \notin \mathrm{Cn}(X)$. Then, since $\mathrm{Cn}(X)$ is inclusion-maximal, follows $-(p \,\&\, -q) \in \mathrm{Cn}(X)$, i.e. $p \rightarrow q \in \mathrm{Cn}(X)$. Thus $p \rightarrow q \in \mathrm{Cn}_A(S)$.

LEMMA L2. *Suppose* $-p \in \mathrm{Cn}(X)$, *where* $X \in A$, *and* $S \in X \perp_A \{-p\}$. *Then there is an element* Z *of* A *such that* $\mathrm{Cn}(Z) = \mathrm{Cn}_A(S \cup \{p\})$.

Proof of Lemma L2. Since $-p \notin \mathrm{Cn}_A(S)$, $S \cup \{p\}$ is A-consistent. To prove maximality, let z be an expression such that $S \cup \{p, z\}$ is A-consistent. We have to prove that $z \in \mathrm{Cn}_A(S \cup \{p\})$.

Since $-p \in \mathrm{Cn}(X)$ and $\mathrm{Cn}(X)$ is maximally consistent, $(p \rightarrow z) \in \mathrm{Cn}(X)$. Now suppose that $(p \rightarrow z) \notin \mathrm{Cn}_A(S)$. Then, since $S \in X \perp_A \{-p\}$, $-p \in \mathrm{Cn}_A(S \cup \{p \rightarrow z\})$, i.e. $((p \rightarrow z) \rightarrow -p) \in \mathrm{Cn}_A(S)$, i.e. $-(p \mathbin{\&} z) \in \mathrm{Cn}_A(S)$, so that $S \cup \{p, z\}$ is A-inconsistent, contrary to the conditions. Thus it is not the case that $(p \rightarrow z) \notin \mathrm{Cn}_A(S)$. Thus $(p \rightarrow z) \in \mathrm{Cn}_A(S)$, i.e. $z \in \mathrm{Cn}_A(S \cup \{p\})$.

LEMMA L3. *Y is an X-close representation of p iff $p \in \mathrm{Cn}(Y)$, $Y \in A$, and $Y \cap X \in X \perp_A \{-p\}$.*

Proof of Lemma L3. Part 1: Suppose that $p \in \mathrm{Cn}(Y)$, $Y \in A$, and $Y \cap X \in X \perp_A \{-p\}$. Further suppose Y is not an X-close representation of p. Then, by *Definition D4*, there is an element Y' of A such that $p \in \mathrm{Cn}(Y')$ and that $Y \cap X \subset Y' \cap X$. Then there is a z such that $z \in Y' \cap X$ and $z \notin Y \cap X$. Since $z \in X$ it follows by $Y \cap X \in X \perp_A \{-p\}$ that $-p \in \mathrm{Cn}_A((Y \cap X) \cup \{z\})$. Thus $(z \rightarrow -p) \in \mathrm{Cn}_A(Y \cap X)$ and, since $Y \cap X \subset Y' \cap X$, also $(z \rightarrow -p) \in \mathrm{Cn}_A(Y' \cap X)$. Thus, since $\mathrm{Cn}_A(Y' \cap X) \subseteq \mathrm{Cn}_A(Y')$, it follows that $(z \rightarrow -p) \in \mathrm{Cn}_A(Y')$. Since $p \in \mathrm{Cn}(Y')$ and $z \in Y'$ it follows that Y' is inconsistent, contrary to the conditions.

Part 2: Suppose that Y is an X-close representation of p, and that $Y \cap X$ is not an element of $X \perp_A \{-p\}$. Since $p \in \mathrm{Cn}(Y)$, $-p \notin \mathrm{Cn}_A(Y \cap X)$. Thus there is a set S such that $Y \cap X \subset S$ and $S \in X \perp_A \{-p\}$. Let Y' be an element of A' such that $-p \in \mathrm{Cn}(Y')$ and $S \subseteq Y' \cap X$. (Cf. Lemma L2). Then $Y' \in A$, $p \in \mathrm{Cn}(Y')$, and $Y \cap X \subset S \subseteq Y' \cap X$, so that $Y \cap X \subset Y' \cap X$. Thus Y is not an X-close representation of p.

LEMMA L4. *If Y is an X-close representation of p, and Z is an X-close representation of q, then $\langle Y, Z\rangle$ is a maximally alike X-close representation of $\langle p, q\rangle$ iff $Y \cap Z \cap X \in X \perp_A \{-p, -q\}$.*

Proof of Lemma L4. In the proof we assume that Y is an X-close representation of p and that Z is an X-close representation of q.

Part 1: Suppose $Y \cap Z \cap X \in X \perp_A \{-p, -q\}$. Further suppose that $\langle Y, Z\rangle$ is not a maximally alike X-close representation of $\langle p, q\rangle$. Then, by definition D4, there are Y' and Z' such that Y' is an X-close represention of p and Z' an X-close representation of q, and that $Y \cap Z \cap X \subset Y' \cap Z' \cap X$. Then there is a z such that

$z \in Y' \cap Z' \cap X$ and $z \notin Y \cap Z \cap X$. Since $z \in X$ and $Y \cap Z \cap X \in X \perp_A \{-p, -q\}$, it follows that either $-p$ or $-q$ is an element of $\mathrm{Cn}_A((Y \cap Z \cap X) \cup \{z\})$. Without loss of generality, we may assume that $-p \in \mathrm{Cn}_A((Y \cap Z \cap X) \cup \{z\})$. Then $(z \to -p) \in \mathrm{Cn}_A(Y \cap Z \cap X)$. Further, since $Y \cap Z \cap X \subset Y' \cap Z' \cap X$, it follows that $(z \to -p) \in \mathrm{Cn}_A(Y' \cap Z' \cap X)$. Then, since $\mathrm{Cn}_A(Y' \cap Z' \cap X) \subseteq \mathrm{Cn}_A(Y')$, it follows that $(z \to -p) \in \mathrm{Cn}_A(Y')$. Since, however, $p \in \mathrm{Cn}(Y')$ and $z \in Y'$, this makes Y' inconsistent, contrary to the conditions.

Part 2: Suppose $Y \cap Z \cap X$ is not an element of $X \perp_A \{-p, -q\}$. Then, since neither $-p$ nor $-q$ is an element of $\mathrm{Cn}_A(Y \cap Z \cap X)$, there is a set S such that $Y \cap Z \cap X \subset S$ and $S \in X \perp_A \{-p, -q\}$. Since S is a subset of X and $-p \notin \mathrm{Cn}_A(S)$, there is a set T_1 such that $S \subseteq T_1$ and $T_1 \in X \perp_A \{-p\}$. Similarly, there is a set T_2 such that $S \subseteq T_2$ and $T_2 \in X \perp_A \{-q\}$.

Let Y' and Z' be elements of A such that $\mathrm{Cn}(Y') = \mathrm{Cn}_A(T_1 \cup \{p\})$ and $\mathrm{Cn}(Z') = \mathrm{Cn}_A(T_2 \cup \{q\})$. (Cf. Lemma L2.) Then Y' is an X-close representation of p and Z' an X-close representation of q, and $Y \cap Z \cap X \subset S$ and $S \subseteq Y' \cap Z' \cap X$, i.e. $Y \cap Z \cap X \subset Y' \cap Z' \cap X$. It follows that $\langle Y, Z \rangle$ is not a maximally alike X-close representation of $\langle p, q \rangle$.

Proof of Theorem T1. Let $\langle Y, Z \rangle$ be a maximally alike X-close representation of $\langle p, p \rangle$. Then Y is an X-close representation of p so that, by *Lemma L3*, $Y \cap X \in X \perp_A \{-p\}$. Similarly, $Z \cap X \in X \perp_A \{-p\}$. Further, by *Lemma L4*, $Y \cap Z \cap X \in X \perp_A \{-p\}$. It follows, since no element of $X \perp_A \{-p\}$ is a proper subset of another element of $X \perp_A \{-p\}$, that $Y \cap X = Z \cap X$. Thus there is an element $S \in X \perp_A \{-p\}$ such that $S \subseteq Y$ and $S \subseteq Z$. Further, $p \in \mathrm{Cn}(Y)$ and $p \in \mathrm{Cn}(Z)$. Thus $\mathrm{Cn}_A(S \cup \{p\})$ is both a subset of $\mathrm{Cn}(Y)$ and a subset of $\mathrm{Cn}(Z)$.

There are now two cases:

(i) If $-p \notin \mathrm{Cn}(X)$, then $S = X = Y = Z$.

(ii) If $-p \in \mathrm{Cn}(X)$, then by *Lemma L2*, $\mathrm{Cn}(Y) = \mathrm{Cn}(Z)$.

In both cases it follows by Definition D3 that $Y \geqslant Z$. Since this holds for all $\langle Y, Z \rangle$ that are maximally alike X-close representations of $\langle p, p \rangle$, it can be concluded that pR_xp.

Proof of Theorem T2. Part 1: This part of the theorem will be proved in its equivalent form that for all negative predicates W, Wp & $Wq \rightarrow W(p \vee q)$. Suppose Wp and Wq, and let X be any element of A.

Case (i), $p \in \text{Cn}(X)$: Let $\langle Y, Z \rangle$ be a maximally alike X-close representation of $\langle p, p \vee q \rangle$. Then $\text{Cn}(Y) = \text{Cn}(Z) = \text{Cn}(X)$, so that $Y \geqslant Z$ follows by Definition D3. Thus $pR_x(p \vee q)$.

Case (ii), $q \in \text{Cn}(X)$: Let $\langle Y, Z \rangle$ be a maximally alike X-close representation of $\langle q, p \vee q \rangle$. Then $\text{Cn}(Y) = \text{Cn}(Z) = \text{Cn}(X)$, so that $Y \geqslant Z$ follows by Definition D3. Thus $qR_x(p \vee q)$.

Case (iii), $-p \;\&\; -q \in \text{Cn}(X)$: Let $\langle Y, Z \rangle$ be a maximally alike X-close representation of $\langle p, p \vee q \rangle$.

By *Lemma L3*, $Y \cap X \in X \perp_A \{-p\}$. By *Lemma L4*, $Y \cap Z \cap X \in X \perp_A \{-p, -p \;\&\; -q\}$. Then, since $X \perp_A \{-p, -p \;\&\; -q\} = X \perp_A \{-p\}$, $Y \cap Z \cap X \in X \perp_A \{-p\}$.

From $Y \cap X \in X \perp_A \{-p\}$, $Y \cap Z \cap X \in X \perp_A \{-p\}$, $Y \cap Z \cap X \subseteq Y \cap X$ and the fact that no element of $X \perp_A \{-p\}$ is a proper subset of another element of $X \perp_A \{-p\}$ follows that $Y \cap Z \cap X = Y \cap X$.

Since $p \in \text{Cn}(Y)$ it follows that $(q \rightarrow p) \in \text{Cn}(Y)$. Since $(-p \;\&\; -q) \in \text{Cn}(X)$ it follows that $(q \rightarrow p) \in \text{Cn}(X)$. Thus $(q \rightarrow p) \in \text{Cn}(Y \cap X)$. From this and $Y \cap Z \cap X = Y \cap X$ follows $(q \rightarrow p) \in \text{Cn}(Z)$. From $(p \vee q) \in \text{Cn}(Z)$ and $(q \rightarrow p) \in \text{Cn}(Z)$ follows $p \in \text{Cn}(Z)$. From $p \in \text{Cn}(Z)$, $p \in \text{Cn}(Y)$, and $Y \cap Z \cap X \in X \perp_A \{-p\}$ follows, by *Lemma L2*, that $\text{Cn}(Y) = \text{Cn}(Z)$. Then $Y \geqslant Z$ follows by Definition D3, so that $pR_x(p \vee q)$.

Thus in all three cases (and for all $X \in A$) there is an r such that Wr and $rR_x(p \vee q)$. Thus $W(p \vee q)$.

Part 2 follows directly from part 1.

Part 3: This part of the theorem will be proved in its equivalent form that for all negative predicates W, Wp & $Wq \rightarrow W(p \;\&\; q)$.

Suppose that Wp and Wq, and let X be any element of A. The proof will be divided into four cases.

Case (i): $p \in \mathrm{Cn}(X)$ and $q \in \mathrm{Cn}(X)$. Let $\langle Y, Z \rangle$ be any maximally alike X-close representation of $\langle p, p \,\&\, q \rangle$. Then $\mathrm{Cn}(Y) = \mathrm{Cn}(Z) = \mathrm{Cn}(X)$, and $Y \geqslant Z$ by Definition D3, so that Wp and $pR_x(p \,\&\, q)$.

Case (ii): $p \in \mathrm{Cn}(X)$ and $-q \in \mathrm{Cn}(X)$. Let $\langle Y, Z \rangle$ be any maximally alike X-close representation of $\langle q, p \,\&\, q \rangle$.

By *Lemma L3*, $Z \cap X \in X \perp_A \{-p \vee -q\}$. Further, by *Lemma L4*, $Y \cap Z \cap X \in X \perp_A \{-q, -p \vee -q\}$. Since $X \perp_A \{-p \vee -q\} = X \perp_A \{-q, -p \vee -q\}$ it follows that both $Z \cap X$ and $Y \cap Z \cap X$ are elements of $X \perp_A \{-p \vee -q\}$. From this, from $Y \cap Z \cap X \subseteq Z \cap X$ and from the fact that no element of $X \perp_A \{-p \vee -q\}$ is a proper subset of another element of $X \perp_A \{-p \vee -q\}$ follows that $Y \cap Z \cap X = Z \cap X$.

From $p \in \mathrm{Cn}(Z)$ and $p \in \mathrm{Cn}(X)$ follows $p \in \mathrm{Cn}(Z \cap X)$. Thus $p \in \mathrm{Cn}(Y \cap Z \cap X)$, thus $p \in \mathrm{Cn}(Y)$.

It follows that $p \,\&\, q$ is an element of both $\mathrm{Cn}(Y)$ and $\mathrm{Cn}(Z)$ and that Y and Z also have a subset in common, namely $Y \cap Z \cap X$, that is an element of $X \perp_A \{-p \vee -q\}$. It follows then, by *Lemma L2*, that $\mathrm{Cn}(Y) = \mathrm{Cn}(Z)$. By Definition D3, $Y \geqslant Z$, so that Wq and $qR_x(p \,\&\, q)$.

Case (iii): $-p \in \mathrm{Cn}(X)$ and $q \in \mathrm{Cn}(X)$. Then, in similar manner as in case (ii) it follows that Wp and $pR_x(p \,\&\, q)$.

Case (iv): $-p \in \mathrm{Cn}(X)$ and $-q \in \mathrm{Cn}(X)$. Let $\langle Y, Z \rangle$ be any maximally alike X-close representation of $\langle p, p \,\&\, q \rangle$. It follows in the same manner as in case (ii) that $Z \cap X = Y \cap Z \cap X$. Since $(p \rightarrow q) \in \mathrm{Cn}(X)$ and $(p \rightarrow q) \in \mathrm{Cn}(Z)$ it follows that $(p \rightarrow q) \in \mathrm{Cn}(Z \cap X)$, and thus $(p \rightarrow q) \in \mathrm{Cn}(Y \cap Z \cap X)$, thus $(p \rightarrow q) \in \mathrm{Cn}(Y)$. From this and $p \in \mathrm{Cn}(Y)$ follows $p \,\&\, q \in \mathrm{Cn}(Y)$. It can now be concluded, exactly as in case (ii), that $Y \geqslant Z$, and that Wp and $pR_x(p \,\&\, q)$.

Thus in all four cases there is an r such that Wr and $rR_x(p \,\&\, q)$. It follows that $W(p \,\&\, q)$.

Proof of Theorem T3. Part 1: We prove the theorem in its equivalent form that for all negative predicates W, $-W(p \vee -p)$ holds for all p iff W is obeyable.

(i): Suppose W is not obeyable. Let X be any element of A. Then there is an element q of $\mathrm{Cn}(X)$ such that Wq. Further let $\langle Y, Z \rangle$ be a maximally alike X-close representation of $\langle q, p \vee -p \rangle$. Then

$\text{Cn}(Y) = \text{Cn}(Z) = \text{Cn}(X)$, so that $Y \geqslant Z$ follows by Definition D3. Thus Wq and $qR_x(p \vee -p)$. Since there is, for all X, such a q, $W(p \vee -p)$.

(ii): Suppose W is obeyable. Then there is an $X \in A$ such that $-Wp$ holds for all $p \in \text{Cn}(X)$.

Part 2: We prove the theorem in its equivalent form that if the negative predicate W is obeyable, then $Wp \rightarrow -W-p$ holds for all p.

Suppose W is obeyable. Then there is an $X \in A$ such that for all p, if $p \in \text{Cn}(X)$, then $-Wp$.

For any p, either $p \in \text{Cn}(X)$ or $-p \in \text{Cn}(X)$. In the first case, $-Wp$. In the latter case, $-W-p$. Thus, in both cases $-Wp \vee -W-p$, i.e. $Wp \rightarrow -W-p$.

Proof of Theorem T4. Part 1: We prove this part of the theorem in its equivalent form that the following schema holds for all negative predicates W that follow the association principle: If $p \rightarrow q$ is an element of the logical closures of all elements of A, then $Wq \rightarrow Wp$.

Suppose the association principle holds. Further suppose that $p \rightarrow q$ is an element of the logical closures of all elements of A and that Wq. Then for all $X \in A$, if $p \in \text{Cn}(X)$ then $q \in \text{Cn}(X)$. Thus it follows by the association principle that Wp.

Parts 2, 3 and 4 follow directly from part (1).

Part 5: Directly from the definition of the association principle.

Proof of Theorem T5. One direction of the theorem follows from part (1) of *Theorem T4*. For the other direction, we first note that $Op \rightarrow Oq$ can be replaced by $Wq \rightarrow Wp$ in the theorem.

Suppose it holds that if $p \rightarrow q$ is an element of the logical closures of all elements of A, then $Wq \rightarrow Wp$. Further let p be such that for all X, if $p \in \text{Cn}(X)$, then there is a $q \in \text{Cn}(X)$ such that Wq. Let $X_1, \ldots X_n$ be the set of elements of A that imply p. Then for each X_k there is an element $q_k \in \text{Cn}(X_k)$ such that Wq_k.

Now let r be $q_1 \vee \ldots q_n$. Then for all $X \in A$, if $p \in \text{Cn}(X)$ then $r \in \text{Cn}(X)$. Thus $p \rightarrow r$ is true in all elements of A. By the condition assumed, $Wr \rightarrow Wp$.

By Wq_1 & Wq_2 & . . . Wq_n follows, by part (1) of *Theorem T2*, Wr. By Wr and $Wr \rightarrow Wp$ follows Wp. This is sufficient to show that the association principle holds.

Proof of Theorem T6. Let $B(A)$ be the set of logical closures of the best elements of A, i.e. $\mathrm{Cn}(X) \in B(A)$ iff there is no Y such that $Y > X$. Then perfectionism can be expressed by the formula $Wp \leftrightarrow p \notin \bigcup(B(A))$. (I.e., p is wrong iff it is not true in any best alternative).

Suppose p is such that for all X, if $p \in \mathrm{Cn}(X)$ then there is a $q \in \mathrm{Cn}(X)$ such that Wq. From $q \in \mathrm{Cn}(X)$ and Wq follows by perfectionism that $\mathrm{Cn}(X) \notin B(A)$. Thus if $p \in \mathrm{Cn}(X)$ then $\mathrm{Cn}(X) \notin B(A)$. Thus $p \notin \bigcup(B(A))$, from which follows Wp. Thus the association principle holds.

Proof of Theorem T7. $B(A)$ is defined as in the proof of *Theorem T6*. Let W be the converse prohibitive predicate of O. Suppose perfectionism and the upper limit-assumption hold. By the latter, there is at least one element X of A such that $\mathrm{Cn}(X) \in B(A)$. Then by perfectionism (in the form given in the proof of *Theorem T6*), if $p \in \mathrm{Cn}(X)$, then $-Wp$. By *Definition D8*, this is sufficient to prove that W, and thus O, is obeyable.

REFERENCES

Alchourrón, Carlos E. and David Makinson (1981) Hierarchies of Regulations and Their Logic, pp. 125–148 in Risto Hilpinen (ed.), *New Studies in Deontic Logic*.

Dayton, Eric (1981) Two Approaches to Deontic Logic, *Journal of Value Inquiry* **15**: 137–147.

Føllesdal, Dagfinn and Hilpinen, Risto (1970) Deontic Logic: An Introduction. In: R. Hilpinen (ed.) *Deontic Logic: Introductory and Systematic Readings*, pp. 1–35.

Halldén, Sören (1957) *On the Logic of Better*.

Hansson, Bengt (1968) Fundamental Axioms for Preference Relations, *Synthese* **18**: 423–442.

Hansson, Sven Ove (1988) Defining 'Good' and 'Bad' in Terms of 'Better', to be published in *Notre Dame Journal of Formal Logic*.

Hansson, Sven Ove (1989) A New Semantical Approach to the Logic of Preference, *Erkenntnis* **31**: 1–42.

Kanger, Stig (1957) New Foundations for Ethical Theory, reprinted in Risto Hilpinen (ed), *Deontic Logic: Introductory and Systematic Readings*, 1971, pp. 36–58.

von Kutschera, Franz (1975) Semantic Analysis of Normative Concepts, *Erkenntnis* **9**: 195–218.

Lewis, David (1974) Semantical Analyses for Dyadic Deontic Logic, pp. 1–14 in Sören Stenlund (ed.) *Logical Theory and Semantic Analysis.*

Menger, Karl (1934) *Moral, Wille, Weltgestaltung.*

Oppenheim, Felix E. (1961), *Dimensions of Freedom.*

Ross, Alf (1941) Imperatives and Logic, *Theoria* **7**: 53–71.

Sayre-McCord, Geoffrey (1986) Deontic Logic and the Priority of Moral Theory, *Noûs* **20**: 179–197.

Sen, Amartya K. (1970) *Collective Choice and Social Welfare.*

Stranzinger, Rudolf (1978) Ein paradoxienfreies deontisches System, pp. 183–192 in *Strukturierungen und Entscheidungen im Rechtsdenken*, Wien.

Vermazen, Bruce (1977) The Logic of Practical 'Ought'-Sentences, *Philosophical Studies* **32**: 1–77.

Weinberger, Ota (1970) Normenlogik anwendbar im Recht, *Logique et Analyse* **13**: 94–106.

von Wright, Georg Henrik (1951) Deontic Logic, *Mind* **60**: 1–15.

von Wright, Georg Henrik (1981) On the Logic of Norms and Actions, in R. Hilpinen (ed.) *New Studies in Deontic Logic*, pp. 3-35.

Uppsala University,
Department of Philosophy,
Villavägen 5,
S-752 36 Uppsala,
Sweden.