# Modal logic for preference based on reasons

Daniel Osherson[1] and Scott Weinstein[2]

[1] Princeton University, Princeton NJ 08544, USA,
osherson@princeton.edu
[2] University of Pennsylvania, Philadelphia PA 19104, USA,
weinstein@cis.upenn.edu

**Abstract.** We discuss the logic of preferences, introducing modal connectives that reflect reasons to prefer that one formula rather than another be true. An axiomatic analysis of two such logics is presented.

## 1 Introduction

The present paper focusses on the modal logic of preference, following up earlier work (Osherson and Weinstein, in press) on the interaction between preference and *reasons*. An example may help to communicate the kind of situation under investigation. You are deciding whether to adopt a certain dog, Fido; alternatively, you might choose the cat Thomasina. To make up your mind, you first imagine how life would be with Fido, taking into account the companionship and safety he would provide but also the expense and bother. Then you do the same for Thomasina. You observe that, compared to Thomasina, life with Fido would have greater value along the first two dimensions but entail less with respect to the second pair. Somehow, you aggregate these four considerations, and plump for Fido.

Our formal reconstruction of this episode is as follows. The world you live in is one of many possibilities including some in which "I adopt Fido" is true and others in which "I adopt Thomasina" is true. In choosing between the two plans, you imagine a world rather similar to yours except that the Fido sentence is true, and another world for the Thomasina sentence. These two worlds are compared for the amount of companionship they provide as well as for safety, expense and bother. A scheme for combining these comparisons is applied, which yields your decision.

The Fido world was delivered by a *selection function* applied to your current world under the thought of adopting Fido, and similarly for the Thomasina world. In other words, selection makes a choice among possible worlds that satisfy whatever proposition is being entertained. In the most basic logic, no conditions regulate how the function operates. But stronger theories impose requirements that fill out the idea that selection seeks a world "close" to its starting point among the worlds that satisfy the target proposition. The most elementary constraint is *reflexivity*, which requires that if the starting world satisfies a proposition $A$, then that world be selected when seeking an $A$-world. A more consequential constraint is that selection be interpreted metrically, in the sense that the chosen world be uniquely nearest to the starting point among $A$-worlds, for some underlying metric that situates all the worlds in play. Several constraints are investigated in Osherson and Weinstein (in press).

The basic logic will be presented shortly, followed by an alternative version that dispenses with selection. It will be seen that the two systems validate the same formulas, a fact not available in Osherson and Weinstein (in press). Before getting started, let us acknowledge some of the prior literature on the logic of preference.

Contemporary work includes several systems that elucidate the interaction between choice and epistemic possibility (see Lang et al., 2003; van Benthem et al., 2009 for an overview). Liu (2008, Ch. 3) is particularly pertinent since it introduces "priorities," which function somewhat like reasons in our theory. Liu's approach is nonetheless different from the one described below inasmuch as selection is absent. A different perspective on the integration of preferences is embodied in the graph-theoretic approach offered in Andréka et al. (2002); different graphs represent different orderings of the alternatives in play, and can be conceived as separate reasons for choice among them. Within another tradition, multi-attribute utility theory (Keeney and Raiffa, 1993) analyses the aggregation of reasons by combining utilities based on separate dimensions. The theory reveals the conditions under which aggregation can proceed additively but stops short of exploring the logical structure of reasons and preference, as we shall do here. Finally (in this abbreviated review), Dietrich and List (2009) provide a representation theorem relating choice to the respective bundles of reasons that apply to the available choices; the simple axioms invoked for their theorem clarify several issues relating to combining reasons.

## 2 The basic theory

Turning to our own proposal, we first introduce the family of languages that are used to express preferences, then provide their formal semantics. A given language is determined by its *signature*, which consists of (a) a non-empty set $\mathbb{P}$ of propositional variables, and (b) a nonempty collection $\mathbb{S}$ of nonempty subsets of $\mathbb{N}$ (the set $\{0, 1, \dots\}$ of natural numbers). The elements of $\mathbb{S}$ serve as indexes for utility functions. The language determined by signature $(\mathbb{P}, \mathbb{S})$ is denoted $\mathcal{L}(\mathbb{P}, \mathbb{S})$, and is built from the following symbols.

(a) the set $\mathbb{P}$ of propositional variables
(b) the unary connective $\neg$
(c) the binary connective $\wedge$
(d) for every set $X \in \mathbb{S}$, the binary connective $\succeq_X$
(e) the two parentheses

Formulas are defined inductively via:

$$p \in \mathbb{P} \quad | \quad \neg\varphi \quad | \quad (\varphi \wedge \psi) \quad | \quad (\varphi \succeq_X \psi) \text{ for } X \in \mathbb{S}.$$

We rely on obvious abbreviations for the boolean connectives including the constants $\top$, $\bot$. We also write: $(\varphi \succ_X \psi)$ for $(\varphi \succeq_X \psi) \wedge \neg(\psi \succeq_X \varphi)$, $(\varphi \approx_X \psi)$ for $(\varphi \succeq_X \psi) \wedge (\psi \succeq_X \varphi)$, $(\varphi \preceq_X \psi)$ for $(\psi \succeq_X \varphi)$, and $(\varphi \prec_X \psi)$ for $(\psi \succ_X \varphi)$.

According to the semantics provided below, $\varphi \succ_1 \psi$ can be understood as follows. As a function of the world you actually inhabit, a world $w$ satisfying $\varphi$, and a world $v$ satisfying $\psi$ are selected. The formula is true just in case $u_1(w) > u_1(v)$, where $u_1$ is a utility function from worlds to numbers, with index 1. If the indexes $1 \ldots 4$ measure companionship, safety, expense, and bother then $X = \{1 \ldots 4\}$ is the aggregate index for all four together. So, if $w$ is the world in which Fido is adopted, and $v$ is the world for Thomasina then Fido is your choice if $u_X(w) > u_X(v)$, in which case $\varphi \succ_X \psi$ is true at the world you inhabit.

A *model* for signature $(\mathbb{P}, \mathbb{S})$ is based on a nonempty set of points called "worlds." Subsets of worlds are termed *propositions*. As discussed above, given a nonempty proposition $A$ and world $w$, we pick an alternative to $w$ among the worlds in $A$. (If $w \in A$ then the "alternative" might be $w$ itself.) Such choices are formalized as follows.

(1) Definition: A *selection function* $s$ over a set $\mathbb{W}$ of worlds is a mapping from $\mathbb{W} \times \{A \subseteq \mathbb{W} \mid A \neq \emptyset\}$ to $\mathbb{W}$ such that for all $w \in \mathbb{W}$ and $\emptyset \neq A \subseteq \mathbb{W}$, $s(w, A) \in A$.

Intuitively, $s$ chooses a member of $A$ that is similar to $w$.

Next, recall that each world can be evaluated according to different utility scales, indexed by members of $\mathbb{S}$.

(2) Definition: A *utility function $u$ over $\mathbb{W}$ and $\mathbb{S}$* is a mapping from $\mathbb{W} \times \mathbb{S}$ to $\Re$ (the reals).

For $w \in \mathbb{W}$ and $\{i\}, X \in \mathbb{S}$, we write $u(w, \{i\})$ as $u_i(w)$, and $u(w, X)$ as $u_X(w)$.

In a given signature $(\mathbb{P}, \mathbb{S})$, $\mathbb{P}$ is a nonempty set of propositional variables. The last component of a model is the assignment of a proposition to each variable in $\mathbb{P}$.

(3) Definition: A *truth-assignment (over $\mathbb{W}$ and $\mathbb{P}$)* is a mapping from $\mathbb{P}$ to the power set of $\mathbb{W}$.

For a truth-assignment $t$, the idea is that $p \in \mathbb{P}$ is true in $w \in \mathbb{W}$ just in case $w \in t(p)$.

(4) Definition: A *(basic) model* for a signature $(\mathbb{P}, \mathbb{S})$ is a quadruple $(\mathbb{W}, s, u, t)$ where
- (a) $\mathbb{W}$ is a nonempty set of worlds;
- (b) $s$ is a selection function over $\mathbb{W}$;
- (c) $u$ is a utility function over $\mathbb{W}$ and $\mathbb{S}$;
- (d) $t$ is a truth-assignment over $\mathbb{W}$ and $\mathbb{P}$.

We may now specify the proposition (set of worlds) expressed by a given formula $\varphi$ in a model $\mathcal{M}$. This proposition is denoted $\varphi[\mathcal{M}]$, and defined as follows.

(5) Definition: Let signature $(\mathbb{P}, \mathbb{S})$, $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$, and model $\mathcal{M} = (\mathbb{W}, s, u, t)$ for $(\mathbb{P}, \mathbb{S})$ be given.

(a) If $\varphi \in \mathbb{P}$ then $\varphi[\mathcal{M}] = t(\varphi)$.
(b) If $\varphi$ is the negation $\neg\theta$ then $\varphi[\mathcal{M}] = \mathbb{W} \setminus \theta[\mathcal{M}]$.
(c) If $\varphi$ is the conjunction $(\theta \wedge \psi)$ then $\varphi[\mathcal{M}] = \theta[\mathcal{M}] \cap \psi[\mathcal{M}]$.
(d) If $\varphi$ has the form $(\theta \succeq_X \psi)$ for $X \in \mathbb{S}$, then $\varphi[\mathcal{M}] = \emptyset$ if either $\theta[\mathcal{M}] = \emptyset$ or $\psi[\mathcal{M}] = \emptyset$. Otherwise:
$\varphi[\mathcal{M}] = \{w \in \mathbb{W} \mid u_X(s(w, \theta[\mathcal{M}])) \geq u_X(s(w, \psi[\mathcal{M}]))\}$.

Note that $(\theta \succeq_X \psi)[\mathcal{M}]$ is defined to be empty if there is no world that satisfies $\theta$ or none that satisfies $\psi$. Thus, we read $(\theta \succeq_X \psi)$ with existential import ("the $\theta$-world is weakly $X$-better than the $\psi$-world," where the definite description is Russellian). In the nontrivial case, let $A \neq \emptyset$ be the proposition expressed by $\theta$ in $\mathcal{M}$, and $B \neq \emptyset$ the one expressed by $\psi$. Then world $w$ satisfies $(\theta \succeq_X \psi)$ in $\mathcal{M}$ iff the world selected from $A$ has $X$-utility no less than that of the world selected from $B$.

In the sequel, we rely on standard model theoretic locutions, notably: model $\mathcal{M}$ *satisfies* $\varphi$ just in case $\varphi[\mathcal{M}] \neq \emptyset$, $\varphi$ is *valid* in $\mathcal{M}$ just in case $\varphi[\mathcal{M}] = \mathbb{W}$, and $\varphi$ is *valid* just in case $\varphi$ is valid in every model. The *basic theory* is the set of $\varphi$ that are valid in every basic model.

Finally, observe that the "global modality" (Blackburn et al., 2001, §2.1) can be expressed in the following manner. Choose any $X \in \mathbb{S}$, and for $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ let:

(6) $$\Box\varphi \stackrel{\text{def}}{=} \neg(\neg\varphi \succeq_X \neg\varphi) \quad \text{and} \quad \Diamond\varphi \stackrel{\text{def}}{=} (\varphi \succeq_X \varphi).$$

Then applying (5)d yields:

(7) Proposition: For all $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ and models $\mathcal{M} = (\mathbb{W}, s, u, t)$:
(a) $\Box\varphi[\mathcal{M}] \neq \emptyset$ iff $\Box\varphi[\mathcal{M}] = \mathbb{W}$ iff $\varphi[\mathcal{M}] = \mathbb{W}$.
(b) $\Diamond\varphi[\mathcal{M}] \neq \emptyset$ iff $\Diamond\varphi[\mathcal{M}] = \mathbb{W}$ iff $\varphi[\mathcal{M}] \neq \emptyset$.

Proposition (7) implies that the axioms of S5 are valid for $\Box$ and $\Diamond$. Other validities are shown below.

## 3 Axioms for the basic theory

The axioms for the basic theory, which we call O, include all $\mathcal{L}(\mathbb{P}, \mathbb{S})$-instances of any standard schematic axiomatization of S5, together with all $\mathcal{L}(\mathbb{P}, \mathbb{S})$-instances of the following additional axiom schemata.

(8) (a) $((\varphi \succeq_X \psi) \wedge (\psi \succeq_X \theta)) \rightarrow (\varphi \succeq_X \theta)$
(b) $(\Diamond\varphi \wedge \Diamond\psi) \leftrightarrow ((\varphi \succeq_X \psi) \vee (\psi \succeq_X \varphi))$
(c) $\Box(\varphi \leftrightarrow \psi) \rightarrow (((\varphi \succeq_X \theta) \leftrightarrow (\psi \succeq_X \theta)) \wedge ((\theta \succeq_X \varphi) \leftrightarrow (\theta \succeq_X \psi)))$

The theorems of O consist of the closure of these axioms under the rules of *modus ponens* and necessitation. The adequacy of O follows from Theorem (10) below.

## 4 Generalized models

In the basic theory, $\varphi \succeq_X \psi$ asserts that $u_X$ attributes at least as much value to the proposition expressed by $\varphi$ as to the proposition expressed by $\psi$. The latter two propositions are represented by elements of each, selected on the basis of the world at which the formula is evaluated. In the present section, we generalize this idea by comparing the value of propositions directly, without recourse to selected worlds as representatives. To begin, let $(\mathbb{P}, \mathbb{S})$ be our background signature, and recall that a *total preorder* is transitive and connected over its domain.

(9) Definition: Let a set $\mathbb{W}$ of worlds be given.

(a) By a *value-ordering for* $\mathbb{W}$ *and* $\mathbb{S}$ is meant a function $v$ from $\mathbb{W} \times \mathbb{S}$ to the set of total preorders over the class of nonempty subsets of $\mathbb{W}$.

(b) Let a truth-assignment $t$ and a value-ordering $v$ for $\mathbb{W}$ and $\mathbb{S}$ be given. Then $(\mathbb{W}, t, v)$ is a *generalized model.*

Thus, a value-ordering arranges propositions by utility, relative to index $X \in \mathbb{S}$ and vantage point $w \in \mathbb{W}$. The semantics of generalized models is given by Definition (5) with the following substitution for clause (5)d. Let $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ and generalized model $\mathcal{M} = (\mathbb{W}, t, v)$ for $(\mathbb{P}, \mathbb{S})$ be given.

(5)d′ If $\varphi$ has the form $(\theta \succeq_X \psi)$ for $X \in \mathbb{S}$, then $\varphi[\mathcal{M}] = \emptyset$ if either $\theta[\mathcal{M}] = \emptyset$ or $\psi[\mathcal{M}] = \emptyset$. Otherwise:

$$\varphi[\mathcal{M}] = \{w \in \mathbb{W} \mid \theta[\mathcal{M}] \text{ comes no earlier than } \psi[\mathcal{M}] \text{ in } v(w, X)\}.$$

We call $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ a *generalized validity* just in case $\varphi$ is valid in all generalized models (that is, just in case for all generalized models $\mathcal{M} = (\mathbb{W}, t, v)$, $\varphi[\mathcal{M}] = \mathbb{W}$).

Here is the sense in which Definition (9) generalizes the basic theory presented in Section 2. Let (basic) model $\mathcal{M} = (\mathbb{W}, s, u, t)$ be given. Then a value-ordering $v$ is induced by the following condition. For $w \in \mathbb{W}$, $X \in \mathbb{S}$, and nonempty $A, B \subseteq \mathbb{W}$, $A$ is (weakly) ordered after $B$ iff $u_X(w_A) \geq u_X(w_B)$ where $w_A = s(w, A)$ and $w_B = s(w, B)$. (The truth-assignment $t$ plays no role.) In Osherson and Weinstein (in press) we exhibit classes of generalized models whose value orderings cannot be induced in this way. The excess of generalized models, however, does not affect the class of generalized validities. For, the latter class is axiomatized by the same system presented in Section 3 for the basic theory.

(10) Theorem: For all $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ the following are equivalent.

(a) $\varphi$ is a theorem of $\mathsf{O}$.

(b) $\varphi$ is a generalized validity.

(c) $\varphi$ is a basic validity.

The proof is provided in the appendix. The small model property for basic and generalized satisfiability is a corollary to the proof, from which decidability follows immediately.

## 5 Discussion

The axioms $\mathsf{O}$ are striking for their simplicity, expressing little more than the preordering of $\succeq_X$, an obvious substitution property, and the apparatus of S5 (along with familiar rules of inference). Apparently, both basic and generalized models represent a wide range of reason-based preferences. As noted in Section 4, there are natural classes of generalized models that are not induced by any basic model. So the fact that the two kinds of models define the same set of validities is perhaps the most noteworthy aspect of Theorem (10).

The generality of the basic theory provides reason to study subclasses of models, such as the metrical models (mentioned in the Introduction). Each such subclass can be evaluated as a theory of rational preference, as well as inviting additions to $\mathsf{O}$.

## Appendix

It is easy to see that (10)a implies (10)b and that (10)b implies (10)c. We proceed to establish that (10)c implies (10)a. For this purpose, we first establish that (10)b implies (10)a. For the latter, we prove the dual, namely,

(11) For all $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$, if $\varphi$ is consistent, then $\varphi$ is satisfiable in a generalized model.

The proof of (11) will be based on a canonical model construction. In order to explain the construction we require the notion of *modal depth*.

(12) Definition: We define $\mu(\varphi)$, the modal depth of $\varphi$, by recursion on $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ as follows.

$$\mu(\varphi) = \begin{cases} 0 & \text{if } \varphi \in \mathbb{P} \\ \mu(\psi) & \text{if } \varphi = \neg\psi \\ \mathsf{max}\{\mu(\psi), \mu(\theta)\} & \text{if } \varphi = (\psi \wedge \theta) \\ \mathsf{max}\{\mu(\psi), \mu(\theta)\} + 1 & \text{if } \varphi = (\psi \preceq_X \theta) \end{cases}$$

Since the satisfiability of single formulas is at issue, we may assume that our signature $(\mathbb{P}, \mathbb{S})$ is finite. For any such signature, it is easy to verify that if $(\mathbb{P}, \mathbb{S})$ is finite, then for any $n \in \mathbb{N}$, there are only finitely many $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ with $\mu(\varphi) \leq n$ up to equivalence in sentential logic. In light of this, we may enforce the convention that any set of formulas of bounded modal depth that we mention is finite. To reduce notational clutter, we fix throughout a finite signature $(\mathbb{P}, \mathbb{S})$ and omit further reference to it. Moreover, we suppose that $\mathbb{S}$ is a singleton and suppress the subscripts on occurrences of $\preceq$. Likewise, they are suppressed on utility functions $u$. It will be seen that these simplifications affect nothing of substance in our construction.

If $\Sigma$ is a set of formulas, we let $\nu(\Sigma) = \{\Box\varphi \mid \Box\varphi \in \Sigma\}$. If $\Sigma$ and $\Sigma'$ are sets of formulas, we say $\Sigma$ is *compatible* with $\Sigma'$ just in case $\nu(\Sigma) = \nu(\Sigma')$. A set of formulas $\Sigma$ is *consistent* just in case $\perp$ is not $\mathsf{O}$-derivable from $\Sigma$; a set of formulas $\Sigma$ is *maximally consistent* just in case it is consistent and no proper extension of it is consistent. We say a set of formulas

$\Gamma$ is *n-maximally consistent* if and only if there is a maximally consistent set $\Sigma$ such that $\Gamma = \{\varphi \in \Sigma \mid \mu(\varphi) \leq n\}$. We abbreviate "$n$-maximally consistent set of formulas" to "$n$-mcs." Note that by our aforementioned convention, every $n$-mcs is finite. We repeatedly use the following fundamental property of maximally consistent sets of formulas.

(13) For every maximally consistent set of formulas $\Gamma$ and formula $\varphi$, if $\varphi$ is $\mathsf{O}$-derivable from $\Gamma$, then $\varphi \in \Gamma$. Moreover, for every $n \in \mathbb{N}$, $n$-mcs $\Sigma$, and $\varphi$ of modal depth $\leq n$, if $\varphi$ is $\mathsf{O}$-derivable from $\Sigma$, then $\varphi \in \Sigma$.

For each $n, m \geq 0$ and $n$-mcs $\Sigma$, we define the *canonical generalized model*, $\mathcal{M}^{n,m}(\Sigma) = (\mathbb{W}^{m,n}, v^{n,m}, t^{n,m})$ of depth $n$ and width $m$ generated by $\Sigma$. Given $n$-mcs $\Sigma$, let $\Xi^n(\Sigma)$ be the family of $n$-mcs's which are compatible with $\Sigma$. The collection of worlds $\mathbb{W}^{n,m}$ of $\mathcal{M}^{n,m}(\Sigma)$ is $\Xi^n(\Sigma) \times \{0, \dots, m\}$. In order to specify the remaining components of $\mathcal{M}^{n,m}(\Sigma)$, we fix an $n$-mcs $\Sigma_0$. We also fix $m \in \mathbb{N}$ to be "large enough" (a lower bound for $m$ appears at the end of the proof). For brevity, we write $\mathcal{M}^n$ for our canonical generalized model $\mathcal{M}^{n,m}(\Sigma_0)$ and we write $\mathbb{W}^n$, $v^n$, and $t^n$ for $\mathbb{W}^{n,m}$, $v^{n,m}$, and $t^{n,m}$, respectively. Moreover, if $w \in \mathbb{W}^n$, we call $w$ an $n$-mcs (ignoring its second coordinate) and likewise we write $\varphi \in w$ just in case $\varphi$ is a member of the first coordinate of $w$. For each $p \in \mathbb{P}$, $t^n(p) = \{w \in \mathbb{W}^n \mid p \in w\}$. Toward defining the value ordering $v^n$, we begin by defining a sequence of partial value orderings $v^n_j$ and partial models $\mathcal{M}^n_j$ simultaneously by induction on $j$, for $0 \leq j \leq n$. Let $v^n_0 = \emptyset$ (the empty partial function) and $\mathcal{M}^n_0 = (\mathbb{W}^n, v^n_0, t^n)$. Note that for every $\varphi$ of modal depth 0, $\varphi[\mathcal{M}^n_0]$ is well-defined since the evaluation of such formulas does not make use of the value ordering. Moreover, for all $w \in \mathbb{W}^n$ and for all $\varphi$ of modal depth 0, $w \in \varphi[\mathcal{M}^n_0]$ if and only if $\varphi \in w$. This follows immediately from (13), the definition of $t^n$, and the fact that each $w \in \mathbb{W}^n$ is an $n$-mcs, since every formula of modal depth 0 is a boolean combination of sentence letters.

Suppose that our construction has proceeded to some stage $j$, with $0 \leq j < n$ resulting in a partial model $\mathcal{M}^n_j = (\mathbb{W}^n, v^n_j, t^n)$. Moreover, suppose, as induction hypothesis, that for every formula of modal depth $\leq j$,

(14) $w \in \varphi[\mathcal{M}^n_j]$ if and only if $\varphi \in w$.

Let $\Omega^n_j = \{\varphi[\mathcal{M}^n_j] \mid \mu(\varphi) \leq j\} - \{\emptyset\}$. We proceed to specify $v^n_{j+1}$. For each $w \in \mathbb{W}^n$, $v^n_{j+1}(w)$ is the relation on $\Omega^n_j$ defined as follows.

(15) For all $\varphi$ and $\psi$ with $\mu(\varphi), \mu(\psi) \leq j$ and $\varphi[\mathcal{M}^n_j]$, $\psi[\mathcal{M}^n_j]$ non-empty,

$$\langle \varphi[\mathcal{M}^n_j], \psi[\mathcal{M}^n_j] \rangle \in v^n_{j+1}(w) \text{ if and only if } (\varphi \preceq \psi) \in w.$$

To complete the construction, we must verify that

(16) for all $w \in \mathbb{W}^n$ and all formulas $\varphi$ of modal depth $\leq j+1$,
  (a) $v^n_{j+1}(w)$ is a pre-order of $\Omega^n_j$, and
  (b) $w \in \varphi[\mathcal{M}^n_{j+1}]$ if and only if $\varphi \in w$.

In order to establish (16)a, we argue as follows. Fix $w \in \mathbb{W}^n$. We first show that $v^n_{j+1}(w)$ is well-defined, that is, if $\varphi$, $\psi$, and $\theta$ are formulas of modal depth $\leq j$ and $\varphi[\mathcal{M}^n_j] = \psi[\mathcal{M}^n_j]$, then

(17) $\langle \varphi[\mathcal{M}^n_j], \theta[\mathcal{M}^n_j] \rangle \in v^n_{j+1}(w)$ if and only if $\langle \psi[\mathcal{M}^n_j], \theta[\mathcal{M}^n_j] \rangle \in v^n_{j+1}(w)$,

and similarly with $\varphi$ and $\theta$ and $\psi$ and $\theta$ reversed. So suppose that

(18) $\varphi$ and $\psi$ are formulas of modal depth $\leq j$ and $\varphi[\mathcal{M}^n_j] = \psi[\mathcal{M}^n_j]$.

It follows at once from (18), (14), and (13), recalling the fact that every $w' \in \mathbb{W}^n$ is an $n$-mcs, that

(19) for all $w' \in \mathbb{W}^n$, $(\varphi \leftrightarrow \psi) \in w'$.

Let $\chi$ be the conjunction of the formulas in $\nu(w)$. It follows from (19) and the definition of $\mathbb{W}^n$ that

(20) $\chi \rightarrow (\varphi \leftrightarrow \psi)$ is a theorem of $\mathsf{O}$,

for otherwise, there would be an $n$-mcs $w' \in \mathbb{W}^n$ with $\neg(\varphi \leftrightarrow \psi) \in w'$ contradicting (19). Since the theorems of $\mathsf{O}$ are closed under necessitation, (20) implies that

(21) $\Box(\chi \rightarrow (\varphi \leftrightarrow \psi))$ is a theorem of $\mathsf{O}$.

Moreover, since each $w' \in \mathbb{W}^n$ is an $n$-mcs, $\Box\theta \rightarrow \Box\Box\theta$ is a theorem of S5, and each of the conjuncts of $\chi$ is a "boxed" formula, it follows from (13) that

(22) for all $w' \in \mathbb{W}^n$, and all maximally consistent sets of formulas $\Gamma \supset w'$, $\Box\chi \in \Gamma$.

It follows from (21), (22), and (13), and the S5 modal principle

$$(\Box\chi \wedge \Box(\chi \rightarrow (\varphi \leftrightarrow \psi))) \rightarrow \Box(\varphi \leftrightarrow \psi),$$

that

(23) $\Box(\varphi \leftrightarrow \psi) \in w$.

But then, by (13), (23), Axiom (8)c and the fact that $w$ is an $n$-mcs,

(24) $(\varphi \preceq \theta) \in w$ if and only if $(\psi \preceq \theta) \in w$.

Therefore $v^n_{j+1}$ is well-defined, since (17) follows directly from (24) and (15).

In order to see that $v^n_{j+1}(w)$ is a pre-order of $\Omega^n_j$, it suffices to show that

(25) (a) $\emptyset$ is not in the field of $v^n_{j+1}(w)$,
(b) $v^n_{j+1}(w)$ is transitive on $\Omega^n_j$, and
(c) $v^n_{j+1}(w)$ is connected on $\Omega^n_j$.

Toward establishing condition (25)a, we show that if $A$ is in the field of $v^n_{j+1}(w)$, then $A \neq \emptyset$. So suppose that

(26) $\langle \varphi[\mathcal{M}^n_j], \psi[\mathcal{M}^n_j]\rangle \in v^n_{j+1}(w)$,

with $\mu(\varphi), \mu(\psi) \leq j$. We show that $\varphi[\mathcal{M}^n_j] \neq \emptyset$. (The argument for $\psi[\mathcal{M}^n_j] \neq \emptyset$ is virtually identical.) To show this, it suffices, by (14), to show that for some $w' \in \mathbb{W}^n$, $\varphi \in w'$. Suppose, for *reductio*, that for all $w' \in \mathbb{W}^n$, $\varphi \notin w'$. Since all $w' \in \mathbb{W}^n$ are $n$-mcs's, it follows at once that for all $w' \in \mathbb{W}^n$, $\neg\varphi \in w'$. As before, let $\chi$ be the conjunction of the formulas in $\nu(w)$. Arguing as we did for (23), we may conclude that $(\chi \rightarrow \neg\varphi)$ is a theorem of $\mathsf{O}$, and thence that $\Box\neg\varphi \in w'$ for all $w' \in \mathbb{W}^n$. It follows immediately by (13) that

(27) $\neg\Diamond\varphi \in w'$, for all $w' \in \mathbb{W}^n$.

On the other hand, it is a direct consequence of (15) and (26) that

(28) $\varphi \preceq \psi \in w$.

It follows from (13), (28), and the right-to-left direction of Axiom (8)b that

(29) $\Diamond\varphi \in w$.

But (29) contradicts (27), thereby establishing that $\varphi[\mathcal{M}^n_{j+1}] \neq \emptyset$.

In order to establish (25)b, suppose that $\varphi, \psi$, and $\theta$ are formulas of modal depth $\leq j$, $w \in \mathbb{W}^n$ and that

(30) $\langle \varphi[\mathcal{M}^n_j], \psi[\mathcal{M}^n_j]\rangle \in v^n_{j+1}(w)$ and $\langle \psi[\mathcal{M}^n_j], \theta[\mathcal{M}^n_j]\rangle \in v^n_{j+1}(w)$.

It follows immediately from (30) and (15) that

(31) $\varphi \preceq \psi \in w$ and $\psi \preceq \theta \in w$.

Therefore, by Axiom (8)a and (13),

(32) $\varphi \preceq \theta \in w$.

Hence, by (32) and (15)

(33) $\langle \varphi[\mathcal{M}^n_j], \theta[\mathcal{M}^n_j]\rangle \in v^n_{j+1}(w)$.

We leave the argument for (25)c to the reader – it is virtually the same as the argument for (25)b, using the left-to-right direction of Axiom (8)b in place of Axiom (8)a.

We now verify (16)b. Note that by (16)a, for every $\varphi$ with $\mu(\varphi) \leq j+1$, $\varphi[\mathcal{M}^n_{j+1}]$ is a well-defined. It is clear from (15) and the choice of $v^n_0$ as the empty partial function that for all $0 \leq i \leq j$ and all $w \in \mathbb{W}^n$, $v^n_i(w) \subseteq v^n_{i+1}(w)$. It follows at once that

(34) for all $\varphi$ of modal depth $\leq j$, $\varphi[\mathcal{M}^n_j] = \varphi[\mathcal{M}^n_{j+1}]$.

Hence, by (14) and (34), it follows at once that in order to prove (16)b, we need only show that for every $w \in \mathbb{W}^n$ and every formula $\varphi$, if $\mu(\varphi) = j+1$, then

(35) $w \in \varphi[\mathcal{M}^n_{j+1}]$ if and only if $\varphi \in w$.

Every formula of modal depth $j+1$ is a boolean combination of formulas of the form $\psi \preceq \theta$, with $\mu(\psi), \mu(\theta) \leq j$. Thus, by (13) and the fact that all $w \in \mathbb{W}^n$ are $n$-mcs's, in order to establish (35), it suffices to show that for all $\psi$ and $\theta$ with $\mu(\psi), \mu(\theta) \leq j$,

(36) $w \in (\psi \preceq \theta)[\mathcal{M}^n_{j+1}]$ if and only if $(\psi \preceq \theta) \in w$.

But (36) is an immediate consequence of (15). This concludes the construction of the partial generalized model $\mathcal{M}^n_n$. By (16)b, it has the "canonical model property"

(37) for all $\varphi$ of modal depth $\leq n$, $w \in \varphi[\mathcal{M}^n_n]$ if and only if $\varphi \in w$.

Let $v^n$ be a value ordering such that for every $w \in \mathbb{W}^n$, $v^n(w)$ extends $v^n_n(w)$ and let $\mathcal{M}^n = (\mathbb{W}^n, v^n, t^n)$. It follows immediately from (37) that $\mathcal{M}^n$ satisfies $\Sigma_0$. Since every formula $\varphi$ is contained in an $n$-mcs for some $n$, this concludes the proof of (11).

We proceed to establish that (10)c implies (10)a. In order to do so, we will make use of the neglected parameter $m$ in our definition of the model $\mathcal{M}^n (= \mathcal{M}^{n,m})$. In particular, recall that the collection of worlds $\mathbb{W}^{n,m}$ of $\mathcal{M}^{n,m}$ is $\Xi^n(\Sigma_0) \times \{0, \ldots, m\}$. By our proof above that (10)b implies (10)a, it will suffice to show that for a sufficiently large choice of $m$, there is a basic partial model $\mathcal{M} = \langle \mathbb{W}^n, s, u, t^n \rangle$ such that $v^n_n$ is the value ordering of $\Omega^n_{n-1}$ induced by $\mathcal{M}$, for this will establish that every consistent $\varphi$ is satisfied by some basic model. It is easy to see that no matter how $m$ is chosen,

(38) for every proposition $A \in \Omega^n_{n-1}$, $\mathsf{card}(A) \geq m$.

Let $\Pi$ be the set of pre-orderings of $\Omega^n_{n-1}$, and choose $m \geq \mathsf{card}(\Pi) \cdot \mathsf{card}(\Omega^n_{n-1})$. It then follows from (38) that there is a function $f : \Pi \times \Omega^n_{n-1} \mapsto \mathbb{W}^n$ such that

(39) (a) for all $\pi \in \Pi$ and $A \in \Omega^n_{n-1}$, $f(\pi, A) \in A$, and
(b) for all distinct $\pi, \pi' \in \Pi$ and all distinct $A, B \in \Omega^n_{n-1}$, $f(\pi, A) \neq f(\pi', B)$.

It follows at once from (39) that we may define $u$ in such a way that

(40) for all $\pi \in \Pi$ and all $A, B \in \Omega^n_{n-1}$,

$$u(f(\pi, A)) \leq u(f(\pi, B)) \text{ if and only if } \langle A, B\rangle \in \pi.$$

Finally, define the selector $s$ as follows.

(41) For all $w \in \mathbb{W}^n$ and $A \in \Omega^n_{n-1}$, $s(w, A) = f(v^n_n(w), A)$.

It follows at once from (40) and (41) that if we let $\mathcal{M}$ be the partial basic model $\langle \mathbb{W}^n, s, u, t^n\rangle$, then $v^n_n$ is the value ordering of $\Omega^n_{n-1}$ induced by $\mathcal{M}$. □

# Bibliography


H. Andréka, M. Ryan, and P.-Y. Schobbens. Operators and laws for combining preferential relations. *Journal of Logic and Computation*, 12:12–53, 2002.

P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, 2001.

Franz Dietrich and Christian List. A reason-based theory of rational choice. Technical Report, *London School of Economics*, 2009.

R. L. Keeney and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. Cambridge University Press, Cambridge UK, 1993.

J. Lang, L. van der Torre, and E. Weydert. Hidden uncertainty in the logical representation of desires. In *Proceedings of eighteenth international joint conference on artificial intelligence (IJCAI03)*, 2003.

F. Liu. *Changing for the Better: Preference Dynamics and Agent Diversity*. PhD thesis, ILLC, University of Amsterdam, 2008.

D. Osherson and S. Weinstein. Preference based on reasons. *Review of Symbolic Logic*, in press.

J. van Benthem, P.k Girard, and O. Roy. Everything else being equal: A modal logic for *Ceteris Paribus* preferences. *Journal of Philosophical Logic*, 38:83–125, 2009.