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 second author is grateful for the opportunity to contribute to this volume in honor of Peter Buneman, whose colleagueship he enjoyed for many years at Penn. Peter's passionate and powerful intellect, his generosity, and his kindness have enriched the lives of all who worked with him. We dedicate this chapter to him. Our hope is that he, or one among his legion of distinguished students, find something here which they can elaborate in ways beyond our capacity to imagine.

The present paper focusses on the modal logic of preference, following up earlier work (Osherson and Weinstein, 2012) 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 (companionship and safety) but entail less with respect to the second pair (expense and bother). 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 (2012).

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 (2012). We then proceed to extend the basic logic in another direction, by introducing quantifiers. 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.

We shall not attempt to further summarize the extensive literature on the logic of preference. An excellent review up to 1989 is offered by Hansson (1989). Surveys of later work are available in Liu (2008) and Dietrich and List (2009). It is, however, worth emphasizing that the formalisms presented below share many features with earlier work. For example, Hansson (1989) introduces a *selection function* for choosing worlds relevant to an affirmation of preference; a somewhat different kind of selection function (based on the analysis of counterfactuals in Stalnaker, 1968) is central to our own theory. Similarly, the idea of attaching values to possible worlds in order to analyze preference among statements appears in several works (e.g., Rescher, 1967), and is pivotal here as well. Our approach thus builds on many earlier discussions; but (so far as we can see) it puts familiar pieces together in a novel way.

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 semantics.

2.1 Syntax

Signatures. 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 numbers appearing in $X \in \mathbb{S}$ represent specific reasons for preference such as the desire for companionship in our introductory example. A set X of reasons influences preference through aggregation of its members. If $\bigcup \mathbb{S} \in \mathbb{S}$ then preference according to $\bigcup \mathbb{S}$ amounts to preference *tout court*; for, such preference takes into account all reasons in play.

Formulas. 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 $(\psi \succ_X \varphi)$.

To illustrate a formula with modal embedding, suppose that utility indexes refer to commercial agents like businesses. Then in a domain that represents economic conditions (availability of raw materials, tax laws, etc.), $\varphi \succ_i \psi$ might mean that φ is more conducive to the profitability of business i than is ψ. Due to competition (e.g., for scarce resources or market share), i might be better off if j does not benefit from the same economic situations as i, yielding, for example:

$$(\varphi \succ_i \psi) \ \rightarrow \ ((\varphi \succ_j \psi) \prec_i (\psi \succ_j \varphi)).$$

A similar interpretation concerns the fitness of species i, j in a given ecological environment.

2.2 Semantics

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 φ, and a world v satisfying ψ 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. Let φ, ψ express the adoption of Fido and Thomasina, respectively. 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.

Models. 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 $\mathbb{P}$.

Propositions. We may now specify the proposition (set of worlds) expressed by a given formula φ 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 φ is the negation $\neg\theta$ then $\varphi[\mathcal{M}] = \mathbb{W} \setminus \theta[\mathcal{M}]$.
(c) If φ is the conjunction $(\theta \wedge \psi)$ then $\varphi[\mathcal{M}] = \theta[\mathcal{M}] \cap \psi[\mathcal{M}]$.
(d) If φ 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 θ or none that satisfies ψ. Thus, we read $(\theta \succeq_X \psi)$ with existential import ("the θ-world is weakly X-better than the ψ-world," where the definite description is Russellian). In the nontrivial case, let $A \neq \emptyset$ be the proposition expressed by θ in $\mathcal{M}$, and $B \neq \emptyset$ the one expressed by ψ. 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* φ just in case $\varphi[\mathcal{M}] \neq \emptyset$, φ is *valid* in $\mathcal{M}$ just in case $\varphi[\mathcal{M}] = \mathbb{W}$, and φ is *valid* just in case φ is valid in every model. The *basic theory* is the set of φ that are valid in every basic model.

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

Then applying (5)e 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 in (8), 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 [using the modality defined in (6)], together with all $\mathcal{L}(\mathbb{P}, \mathbb{S})$-instances of the following additional axiom schemata. $X \in \mathbb{S}$, and $\varphi, \psi, \theta \in \mathcal{L}(\mathbb{P}, \mathbb{S})$:

(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 φ as to the proposition expressed by ψ. 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)e. 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)e′ If φ 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 φ 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 (2012) 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) φ is a theorem of O.
- (b) φ is a generalized validity.
- (c) φ is a basic validity.

The proof is provided in Appendix 1. The small model property for basic and generalized satisfiability is a corollary to the proof, from which decidability follows immediately.

The axioms 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 O.

5 Quantified preference logic

The basic system described above can be seen as a propositional calculus extended with modal binary connectives. Our present purpose is to show how the propositional part can be replaced with predicate calculus. We start with syntax.

5.1 Syntax for quantified preference logic

Signatures. A quantified language is built from its "signature."

(11) Definition: By a *signature (for quantified preference logic* is meant a pair $(\mathbb{L}, \mathbb{S})$ where
(a) $\mathbb{L}$ is a collection of predicates and function symbols of various arities.
(b) $\mathbb{S}$ is a nonempty collection of nonempty subsets of natural numbers $(0, 1 \ldots)$.

As before, members of $\mathbb{S}$ stand for sets of reasons thought of as dimensions for evaluating possible worlds.

Formulas. We may now specify the language $\mathcal{L}(\mathbb{L}, \mathbb{S})$ parameterized by the signature $(\mathbb{L}, \mathbb{S})$. Formulas are built from the following symbols.

(a) the members of $\mathbb{L}$ along with the identity sign $=$
(b) for each $X \in \mathbb{S}$, the binary connective $\succeq_X$
(c) the binary connective $\wedge$ and the unary connective $\neg$
(d) the quantifier $\exists$
(e) the two parentheses, (,)
(f) a denumerable collection $v_0, v_1 \ldots$ of individual variables (denoted below by x, y, z).

The set of *terms* is constructed from functions and variables as usual. The set $\mathcal{L}(\mathbb{L}, \mathbb{S})$ of *formulas* is likewise built in the usual way except that we add the clause:

Given $\varphi, \psi \in \mathcal{L}(\mathbb{L}, \mathbb{S})$ and $X \in \mathbb{S}$, $\varphi \succeq_X \psi$ also belongs to $\mathcal{L}(\mathbb{L}, \mathbb{S})$.

In addition to our earlier abbreviations, we write $\forall x \varphi$ for $\neg \exists x \neg \varphi$. Also, the global modalities $\Box \varphi$ and $\Diamond \varphi$ are defined as before [via $\neg(\neg \varphi \succeq_X \neg \varphi)$ and $\varphi \succeq_X \varphi$, respectively].

Examples of formulas. The following formulas serve as illustration.

(12) (a) $\exists x(Px \succ_X \forall y Py)$
(b) $\exists x Px \succ_X \forall y Py$

In the domain of people, (12)a affirms that there is someone for whom satisfying P is preferable to everyone satisfying it. This might well be true. For example, from my perspective, it's better that I discover a metric ton of gold than that everyone does (where the reasons encoded in X are basely materialistic). In contrast, (12)b entails that someone getting the gold is better than everyone getting it, which might be false if it doesn't strike me as plausible that I'm the lucky person. We return to (12) later on.

The next example is more complicated inasmuch as it exhibits modal embedding. Let the domain of discourse consist of citizens in a modern state. Suppose that the predicate P picks out the charismatic, socialist politicians (if any) in a given possible world. Suppose Q picks out the fabulously wealthy citizens in a given world. We'll also rely on two utility scales. Let u_c measure the level of consumer confidence in a given world (greater consumer confidence yielding greater u_c utility); let u_j measure the level of social and economic justice in a given world (more justice means greater u_j utility). Now consider:

$$\forall x(\ (\ \exists y Qy\ \succ_c\ Px\ )\ \succ_j\ Px)$$

According to the semantics provided below, this formula is true in a given world w_0 just in case the following circumstances obtain. For all citizens (say, Tom), there is greater social justice in the nearest world w_1 to w_0 in which

(13) the existence of charismatic socialist leaders provokes more consumer confidence than does Tom's being fabulously wealthy

compared to the nearest world w_2 to w_0 in which Tom is fabulously wealthy. That is, $u_j(w_1) > u_j(w_2)$. Of course, we must also interpret (13) according to our semantics. It means that the nearest world w_3 to w_1 with charismatic socialist leaders has greater consumer confidence than the nearest world w_4 to w_1 in which Tom is fabulously wealthy. That is, $u_c(w_3) > u_c(w_4)$. Of course, the nearest world that satisfies a certain formula might be your own.

5.2 Semantics in quantified preference logic

Models in quantified preference logic. Recall that a signature $(\mathbb{L}, \mathbb{S})$ consists of vocabulary ($\mathbb{L}$) and sets of utility indices ($\mathbb{S}$).

(14) DEFINITION: Let a signature $(\mathbb{L}, \mathbb{S})$ be given. By a *model* for the signature is meant a quintuple $\mathcal{M} = \langle D, \mathbb{W}, t, u, s\rangle$ where:
(a) D is a nonempty set, the *domain* of $\mathcal{M}$.

(b) $\mathbb{W}$ is a nonempty set of points, the *worlds* of $\mathcal{M}$.
(c) t maps $\mathbb{W} \times \mathbb{L}$ to the appropriate set-theoretic objects over D. (For example, if $Q \in \mathbb{L}$ is a binary relation symbol then $t(w, Q)$ is a subset of $D \times D$.) Identity is assigned to $=$.
(d) u is a function from $\mathbb{S} \times \mathbb{W}$ to the real numbers. For $X, \{i\} \in \mathbb{S}$ we write $u_X(w)$ in place of $u(X, w)$ and $u_i(w)$ in place of $u(\{i\}, w)$.
(e) s is a function from $\mathbb{W} \times \{A \subseteq \mathbb{W} \mid \emptyset \neq A\}$ such that for all $w \in \mathbb{W}$ and $\emptyset \neq A \subseteq W$, $s(w, A) \in A$.

Thus, $\mathbb{W}$ corresponds to a set of potential situations; via t, each gives extensions in D to the vocabulary in $\mathbb{L}$. The function u_X measures the utility of worlds according to the considerations encoded in $X \in \mathbb{S}$. Finally, given a world w_0 and a set A of worlds, s selects a "cognitively salient" member of A, where salience may depend on the vantage point w_0.

Propositions in quantified preference logic. Subsets of worlds are called *propositions*. In the context of a given model, our semantic definition assigns a proposition (subset of $\mathbb{W}$) to each closed formula. To explain, fix a signature $(\mathbb{L}, \mathbb{S})$, and let a model $\mathcal{M} = \langle D, \mathbb{W}, t, u, s\rangle$ be given. By an *assignment (for $\mathcal{M}$)* is meant a map of the individual variables of $\mathcal{L}(\mathbb{L}, \mathbb{S})$ into D. Given a variable x and assignment d, an x *variant* of d is any assignment that differs from d at most in the member of D assigned to x. Assignments are extended to terms of $\mathcal{L}(\mathbb{L}, \mathbb{S})$ in the usual way.

(15) Definition: Let a model $\mathcal{M} = \langle D, \mathbb{W}, t, u, s\rangle$ and assignment d be given. For $\varphi \in \mathcal{L}(\mathbb{L}, \mathbb{S})$, the proposition $\varphi[\mathcal{M}, d]$ is defined as follows.
(a) If φ is $Pt_1 \ldots t_n$ for $P \in \mathbb{L}$ and terms $t_1 \ldots t_n$ then:
$$\varphi[\mathcal{M}, d] = \{w \in \mathbb{W} \mid \langle d(t_1) \ldots d(t_n)\rangle \in t(w, P)\}.$$
(b) If φ is the negation $\neg\theta$ then $\varphi[\mathcal{M}, d] = \mathbb{W} \setminus \theta[\mathcal{M}, d]$.
(c) If φ is the conjunction $(\theta \wedge \psi)$ then $\varphi[\mathcal{M}, d] = \theta[\mathcal{M}, d] \cap \psi[\mathcal{M}, d]$.
(d) If φ is the existential $\exists x\psi$ then $\varphi[\mathcal{M}, d]$ is the set of $w \in \mathbb{W}$ such that $w \in \psi[\mathcal{M}, d']$ for some x variant d' of d.
(e) If φ has the form $(\theta \succeq_X \psi)$ for $X \in \mathbb{S}$, then $\varphi[\mathcal{M}, d] = \emptyset$ if either $\theta[\mathcal{M}, d] = \emptyset$ or $\psi[\mathcal{M}, d] = \emptyset$. Otherwise:
$$\varphi[\mathcal{M}, d] = \{w \in \mathbb{W} \mid u_X(s(w, \theta[\mathcal{M}, d])) \geq u_X(s(w, \psi[\mathcal{M}, d]))\}.$$

Thus, relative to $\mathcal{M}$ and d, the formula $(\theta \succeq_X \psi)$ expresses the null proposition if evaluating it requires that s choose a world from $\emptyset$. (Preference makes a covert existential claim in the present theory, namely, that there is something to choose between.) Otherwise $w \in \mathbb{W}$ belongs to the proposition expressed by $(\theta \succeq_X \psi)$ just in case the world chosen by s to represent $\theta[\mathcal{M}, d]$ has greater X-utility than the world chosen by s to represent $\psi[\mathcal{M}, d]$ — where s's choices depend on the current situation w. Informally, we think of s as choosing the most similar world to w among those available in the proposition at issue.

We extract the assignment-invariant core of a formula's proposition in the standard way.

(16) DEFINITION: Let $\varphi \in \mathcal{L}(\mathbb{L}, \mathbb{S})$ and model $\mathcal{M} = \langle D, \mathbb{W}, t, u, s \rangle$ be given. We write $\varphi[\,\mathcal{M}\,]$ for the intersection of $\varphi[\,\mathcal{M}, d\,]$ over all assignments d.

It follows that for closed $\varphi \in \mathcal{L}(\mathbb{L}, \mathbb{S})$ (no free variables), $\varphi[\,\mathcal{M}\,] = \varphi[\,\mathcal{M}, d\,]$ for any assignment d. As usual, we call closed $\varphi \in \mathcal{L}(\mathbb{L}, \mathbb{S})$ *satisfiable* just in case $\varphi[\,\mathcal{M}\,] \neq \emptyset$ for some model $\mathcal{M}$; and φ is valid iff $\neg\varphi$ is not satisfiable.

Analysis of the formulas in Example (12). The formula (12)a is true at w_0 in model $\langle D, \mathbb{W}, t, u, s \rangle$ just in case there is $a \in D$ such that

the nearest world w_1 (according to s) in which $a \in t(w_1, P)$

has higher u_X value than

the nearest world w_2 (according to s) in which $t(w_2, P) = D$.

On the other hand, (12)b is true at w_0 in $\langle D, \mathbb{W}, t, u, s \rangle$ just in case

the nearest world w_1 (according to s) in which $t(w_1, P) \neq \emptyset$

has higher u_X value than

the nearest world w_2 (according to s) in which $t(w_2, P) = D$.

6 Basic properties of quantified preference logic

6.1 Expressive power of modal formulas

It is worth verifying that our modal vocabulary allows additional propositions to be expressed.

(17) DEFINITION:
- (a) The *modal depth* of formulas is defined inductively. First-order (non-modal) formulas have modal depth zero. If $\varphi, \psi \in \mathbb{L}$ have respective modal depths m, n then $\varphi \succeq_X \psi$ has modal depth $1 + \max\{m, n\}$.
- (b) We say that a model $\mathcal{M}$ *has a modal hierarchy* just in case there are closed formulas $\varphi_0, \varphi_1 \ldots$ such that for all $n \geq 0$:
  - i. φ_n has modal depth n;
  - ii. for all closed $\psi \in \mathcal{L}$ of modal depth n or less, $\varphi_{n+1}[\,\mathcal{M}\,] \neq \psi[\,\mathcal{M}\,]$.

(18) DEFINITION: Let $\mathcal{N} = \langle D, \mathbb{W}, t \rangle$ be the first three components of a model, missing just the utility and selection functions, u, s. Notice that $\langle D, \mathbb{W}, t \rangle$ assigns a proposition $\psi[\,\mathcal{N}\,] \subseteq \mathbb{W}$ to each non-modal $\psi \in \mathcal{L}$. We call $\mathcal{N}$ a *normal core* just in case D is countable, $\mathbb{W}$ is countably infinite, and there is non-modal, closed $\psi \in \mathcal{L}$ with $\emptyset \neq \psi[\,\mathcal{N}\,] \neq \mathbb{W}$.

Now fix a countable signature $(\mathbb{L}, \mathbb{S})$. The following proposition reveals the near ubiquity of modal hierarchies.

(19) PROPOSITION: Let $\mathcal{N} = \langle D, \mathbb{W}, t\rangle$ be a normal core. Then there is a utility function $u : \mathbb{S} \times \mathbb{W} \to \Re$ and a selection function $s : \mathbb{W} \times \{A \subseteq \mathbb{W} \mid A \neq \emptyset\} \to \mathbb{W}$ such that the model $\mathcal{M} = \langle D, \mathbb{W}, t, u, s\rangle$ has a modal hierarchy.

PROOF: Choose utility index $X \in \mathbb{S}$, let $\mathcal{N} = \langle D, \mathbb{W}, t\rangle$ be a normal core, and fix closed, non-modal $\psi \in \mathcal{L}$ with $\emptyset \neq \psi[\mathcal{N}] \neq \mathbb{W}$. By replacing ψ with its negation if necessary, we can ensure that $\psi[\mathcal{N}]$ has at least two elements. Let φ_0 be ψ and let φ_{n+1} be $(\top \prec_X \varphi_n)$. Observe that for all $n \in \mathbb{N}$, φ_n has modal depth n. We will define s and u in such a way that $\varphi_0, \varphi_1 \ldots$ is a modal hierarchy for $\mathcal{M} = \langle D, \mathbb{W}, t, u, s\rangle$.

Let $\{w_0, w_1, \ldots\}$ enumerate $\mathbb{W}$. Since $\psi[\mathcal{N}] = \varphi_0[\mathcal{N}]$ has at least two elements, we may assume without loss of generality that $\{w_0, w_1\} \subseteq \varphi_0[\mathcal{N}]$. Let u be any utility function that meets the conditions:

(20) $u_X(w_0) = 0$ and for all $i > 0$, $u_X(w_i) = 1$.

It remains to specify the selection function s, and to show that it generates a modal hierarchy. This is achieved by inductively defining a sequence of "partial selection" functions s_n, $n \in \mathbb{N}$. At stage n, the partial selector s_n defines a partial model $\mathcal{M}_n = \langle D, \mathbb{W}, t, u, s_n\rangle$ which yields a proposition $\chi[\mathcal{M}_n, d]$ for each assignment d, and each $\chi \in \mathcal{L}$ of modal depth n or below. It will be easy to see that for each such χ and d, $\chi[\mathcal{M}_n, d] = \chi[\mathcal{M}, d]$ where $\mathcal{M} = \langle D, \mathbb{W}, t, u, s\rangle$ with $\bigcup_n s_n \subset s$. Let $\mathfrak{P}_n$ denote the family of nonempty propositions expressed by formulas of modal depth n or below with arbitrary assignments of members of D to their free variables. It is easy to verify that $\mathfrak{P}_n$ is countable. At stage $n = 0$, we let $s_0 = \emptyset$.

For stage $n + 1$, we will define s_{n+1} so that:

(a) s_{n+1} is defined for every pair (w, X) where $w \in \mathbb{W}$ and $X \in \mathfrak{P}_n$; hence, for every assignment d and $\chi \in \mathcal{L}$ of modal depth n or below, $\chi[\mathcal{M}_n, d]$ is well defined.
(b) $\varphi_{n+1}[\mathcal{M}_{n+1}] \notin \mathfrak{P}_n$ hence $\varphi_{n+1}[\mathcal{M}] \notin \mathfrak{P}_n$;

Moreover, at every stage n, it will be the case that $\{w_0, w_1\} \subseteq \varphi_n[\mathcal{M}_n]$. In particular, $\{w_0, w_1\} \subseteq \varphi_0[\mathcal{M}_0] = \psi[\mathcal{N}]$ follows from our choice of ψ.

Now we complete stage $n + 1$. For all $w \in \mathbb{W}$, set $s_{n+1}(w, \mathbb{W}) = w_0$ (hence we always draw w_0 from the proposition expressed by $\top$). For all $w \in \mathbb{W}$ and all $C \in \mathfrak{P}_n - \{\varphi_n[\mathcal{M}_n], \mathbb{W}\}$, choose $s_{n+1}(w, C)$ to be an arbitrary member of C. For the remainder of s_{n+1}, choose $A \subseteq \mathbb{W} - \{w_0, w_1\}$ such that $A \notin \{B - \{w_0, w_1\} \mid B \in \mathfrak{P}_n\}$. Such an A exists because $\mathfrak{P}_n$ is countable. For all $w \in \mathbb{W}$, we define:

$$s_{n+1}(w, \varphi_n[\mathcal{M}_n]) = \begin{cases} w_1 & \text{if } w \in A \cup \{w_0, w_1\} \\ w_0 & \text{otherwise.} \end{cases}$$

It follows immediately from (20) that $\varphi_{n+1}[\mathcal{M}_{n+1}] = A \cup \{w_0, w_1\} \notin \mathfrak{P}_n$. □

A natural question about Proposition (19) is whether modal hierarchies still appear when models satisfy various *frame properties*. To illustrate, model $\mathcal{M} = \langle D, \mathbb{W}, t, u, s\rangle$ is called "reflexive" just in case for all $w \in \mathbb{W}$ and $A \subseteq \mathbb{W}$, if $w \in A$ then $s(w, A) = w$. Reflexivity embodies the idea that the actual world is closer to home than any other world. Several frame properties are examined in Osherson and Weinstein (2012), and also below. In the case of reflexivity, the foregoing proof can be adjusted to show that any normal core can be extended to a reflexive model with modal hierarchy. We leave unexplored the larger project of characterizing the frame properties that allow modal hierarchies, or identifying natural properties that do not.

6.2 Undecidability of satisfaction

Suppose that the signature $(\mathbb{L}, \mathbb{S})$ contains two unary predicates $P, Q \in \mathbb{L}$. Then it follows from the argument in Kripke (1962) that:

(21) Proposition: The satisfiable subset of $\mathcal{L}(\mathbb{L}, \mathbb{S})$ is not decidable.

Kripke's argument hinges on a mapping from first-order sentences with just the binary relation symbol R to modal sentences that replace Rxy with $\Diamond(Px \wedge Qy)$. On the other hand, the validities are axiomatizable:

(22) Proposition: If the signature is effectively enumerable then so is the set of valid formulas in quantified preference logic.

This fact follows from Proposition (27), below.

6.3 Size of models

Suppose that the signature contains a binary predicate G. Then the upward Löwenheim-Skolem property fails to apply to quantified preference logic. Indeed:

(23) Proposition: There is $\varphi \in \mathcal{L}(\mathbb{L}, \mathbb{S})$ such that:
   (a) Some model $\langle D, \mathbb{W}, t, u, s\rangle$ with D countable satisfies φ.
   (b) No model $\langle D, \mathbb{W}, t, u, s\rangle$ with D uncountable satisfies φ.

Proof: Basically, φ says that $\prec$ is a lexicographical order on $D \times D$; such an order cannot be embedded in $\langle \Re, <\rangle$ if D is uncountable. For typographical simplicity, we choose $X \in \mathbb{S}$, and write $\prec$ in place of $\prec_X$.

Specifically, we take φ to be the conjunction of the following formulas.

(24) (a) $\forall x \forall y(x \neq y \rightarrow ((Gxx \prec Gyy) \vee (Gyy \prec Gxx))$
   (b) $\forall x_1 y_1 x_2 y_2((Gx_1y_1 \prec Gx_2y_2) \leftrightarrow ((Gx_1x_1 \prec Gx_2x_2) \vee ((x_1 = x_2) \wedge (Gy_1y_1 \prec Gy_2y_2))))$

Let a model $\mathcal{M} = \langle D, \mathbb{W}, t, u, s\rangle$ and $w_0 \in \mathbb{W}$ be given with $w_0 \in \varphi[\mathcal{M}]$. We define:

$$X = \{u(s(w_0, Gxx[\mathcal{M}, d(a/x)])) \mid a \in D\}.$$

Then (24)a implies that X (a set of reals) has the same cardinality as D. Define:

$$Y = \{u(s(w_0, Gxy[\mathcal{M}, d(a/x, b/y)])) \mid a, b \in D\}.$$

Then (24)b implies that $\langle Y, <\rangle$ is isomorphic to the lexicographic ordering of $X \times X$.

We leave to the reader the verification that φ is satisfiable in a model with countable domain. On the other hand, suppose that the domain is uncountable, whence X is uncountable. Then the existence of an isomorphism between $\langle Y, <\rangle$ and the lexicographic ordering of $X \times X$ contradicts the separability of the real line. □

6.4 Preorder models

We can recover the upward Löwenheim-Skolem property by introducing a more general way to compare the value of worlds. Recall that a (*total*) *preorder* is transitive, connected, and reflexive over its domain. Given a signature $(\mathbb{L}, \mathbb{S})$, we achieve more generality by replacing u in a model $\langle D, \mathbb{W}, t, u, s\rangle$ with a map $\trianglerighteq$ from $\mathbb{S}$ to the set of preorders over $\mathbb{W}$. [We write $\trianglerighteq_X$ for $\trianglerighteq(X)$, $X \in \mathbb{S}$.] In such a model $\langle D, \mathbb{W}, t, \trianglerighteq, s\rangle$, we evaluate $(\theta \succeq_X \psi)$ according to the following rule, in place of (15)e.

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

$$\varphi[\mathcal{M}, d] = \{w \in \mathbb{W} \mid s(w, \theta[\mathcal{M}, d]) \trianglerighteq_X s(w, \psi[\mathcal{M}, d])\}.$$

In what follows, we'll call the semantics based on (15)e′ *preorder logic*. The original semantics, based on (15)e, will be called *utility logic*. It is easy to see that utility logic is a special case of preorder logic (since assigning utilities to worlds preorders them). Also, it is straightforward to show that the formula φ in the proof of Proposition (23) is satisfied in a preorder model with uncountable domain D. Indeed, the following Löwenheim-Skolem Theorem holds for preorder models.

(25) PROPOSITION: Let $\langle D, \mathbb{W}, t, \trianglerighteq, s\rangle$ be a preorder model for a countable signature.

(a) If $\mathbb{W}$ is infinite, then for every infinite cardinal κ there is a preorder model $\mathcal{M}' = \langle D', \mathbb{W}', t', \trianglerighteq', s'\rangle$ such that $\mathsf{card}(\mathbb{W}') = \kappa$ and for every sentence φ,
$\mathcal{M} \models \varphi$ if and only if $\mathcal{M}' \models \varphi$.

(b) If D is infinite, then for every infinite cardinal κ there is a preorder model $\mathcal{M}' = \langle D', \mathbb{W}', t', \trianglerighteq', s'\rangle$ such that $\mathsf{card}(D') = \kappa$ and for every sentence φ,
$\mathcal{M} \models \varphi$ if and only if $\mathcal{M}' \models \varphi$.

Despite the greater generality of preorder logic, and the contrast between Propositions (25) and (23), the distinction between utility and preorder models is not discernible by formulas. Indeed:

(26) PROPOSITION: A formula θ is valid in the class of utility models if and only if it is valid in the class of preorder models.

Finally, the next proposition shows that the set of formulas which are valid in preorder models (and hence utility models, by the preceding proposition) is axiomatizable. We assume that the signature is effectively enumerable.

(27) PROPOSITION: The set of formulas which are valid in preorder models is effectively enumerable.

Proofs of Propositions (25), (26), and (27) are given in the Appendix 2. We have not investigated the quantified version of "generalized logic," introduced in Section 4 above.

7 Subclasses of utility models

For the remainder of the discussion, only utility models (introduced in Section 5.2) are at issue. (We leave preorder models to one side.)

7.1 Metricity

Many interesting properties of a model $\langle D, \mathbb{W}, t, u, s\rangle$ can formulated just in terms of $\mathbb{W}$ and s (the model's "frame"). For example, Osherson and Weinstein (2012) consider the following way to express the idea that s chooses "the nearest world."

(28) DEFINITION: A model $\langle D, \mathbb{W}, t, u, s\rangle$ is *metric* just in case there is a metric $d\colon \mathbb{W}\times\mathbb{W} \to \Re$ such that for all $w \in \mathbb{W}$ and $\emptyset \neq A \subseteq \mathbb{W}$, $s(w, A)$ is the unique d-closest member of A to w.

Note that a model is metric only if d-closest worlds exist (there are no chains of worlds ever d-closer to a given world). It is easy to see that in a metric model the set of worlds is countable. There are several properties of models that are implied by metricity, including the following two, articulated by Stalnaker (1968).

(29) DEFINITION: Let model $\mathcal{M} = \langle D, \mathbb{W}, t, u, s\rangle$ be given.
(a) $\mathcal{M}$ is *reflexive* just in case for all $A \subseteq \mathbb{W}$ and $w \in A$, $s(w, A) = w$.
(b) $\mathcal{M}$ is *regular* just in case for all $A \subseteq B \subseteq \mathbb{W}$ and $w \in \mathbb{W}$, $s(w, B) \in A$ implies $s(w, A) = s(w, B)$.

These properties are explored in Osherson and Weinstein (2012). Here we focus on:

(30) Definition: A model $\langle D, \mathbb{W}, t, u, s\rangle$ is *transitive* just in case for all $A, B, C \subseteq \mathbb{W}$ with $A, B \neq \emptyset$, and $w_0 \in \mathbb{W}$, if $s(w_0, A \cup B) \in A$ and $s(w_0, B \cup C) \in B$ then $s(w_0, A \cup C) = s(w_0, A \cup B)$.

Exploiting our quantificational apparatus, we can write a formula that is true in all transitive models but not valid. We assume that the signature includes the predicate P. For notational ease, we suppress the X on $\approx_X$.

(31) Proposition: Let φ be the conjunction of the following formulas.
(a) $\forall xy(x \neq y \to (Px \not\approx Py))$
(b) $\forall xyz\,((x \neq y \wedge y \neq z \wedge x \neq z) \to ((((Px \vee Py) \approx Px) \wedge ((Py \vee Pz) \approx Py)) \to (Px \vee Pz) \approx Px))$
Then φ is invalid but valid in the class of transitive models.

The proposition can be viewed as expressing the transitivity of revealed preference, e.g., $(Px \vee Py) \approx Px$ says that Px is chosen from the mutually exclusive options Px, Py.

Proof: Let model $\mathcal{M} = \langle D, \mathbb{W}, t, u, s\rangle$, $w_0 \in \mathbb{W}$ and assignment d be given. Let $Px[\mathcal{M}, d] = A$, $Py[\mathcal{M}, d] = B$ and $Pz[\mathcal{M}, d] = C$. If any of $d(x), d(y), d(z)$ are identical or either A or B are empty then we are done. Otherwise, in the presence of (31)a, $(Px \vee Py) \approx Px$ and $(Py \vee Pz) \approx Py$ imply respectively that $s(w_0, A \cup B) \in A$ and $s(w_0, B \cup C) \in B$. So transitivity implies $s(w_0, A \cup C) = s(w_0, A \cup B)$ which entails $w_0 \in (Px \vee Pz) \approx (Px \vee Py)[\mathcal{M}, d]$. So the proposition follows by the transitivity of $\approx$ from $w_0 \in (Px \vee Py) \approx Px[\mathcal{M}, d]$. □

7.2 Beyond the frame

Rational agents might not be able to discriminate between isomorphic worlds. To formulate this idea, fix a signature $(\mathbb{L}, \mathbb{S})$, and let model $\mathcal{M} = \langle D, \mathbb{W}, t, u, s\rangle$ be given. We say that $v, w \in \mathbb{W}$ are *isomorphic* ($v \simeq w$) just in case there is a permutation h of D such that for all $Q \in \mathbb{L}$, h (applied component-wise) maps $t(v, Q)$ onto $t(w, Q)$.

(32) Definition: Model $\langle D, \mathbb{W}, t, u, s\rangle$ is *utility-invariant* just in case for all isomorphic $v, w \in \mathbb{W}$, $u_X(v) = u_X(w)$ for all $X \in \mathbb{S}$.

This is not a frame property because all components of the model are involved in its formulation. Validity in the utility-invariant models doesn't imply validity in the strict sense. Indeed, we have:

(33) Proposition: Let signature $(\mathbb{L}, \mathbb{S})$ be given with $\mathbb{L}$ finite, and distinct $X, Y \in \mathbb{S}$. Then there is invalid $\varphi \in \mathcal{L}(\mathbb{L}, \mathbb{S})$ that is valid in the class of utility-invariant models.

Proof: There is $\chi \in \mathcal{L}(\mathbb{L}, \mathbb{S})$ such that for all models $\mathcal{M} = \langle D, \mathbb{W}, t, u, s\rangle$, $\chi[\mathcal{M}] = \mathbb{W}$ iff $| D | = 2$. Hence, by the finitude of $\mathbb{L}$ and the presence of identity, there is closed, satisfiable $\psi \in \mathcal{L}(\mathbb{L}, \mathbb{S})$ such that for all models $\mathcal{M}$, if $w_1, w_2 \in \psi[\mathcal{M}]$ then $w_1 \simeq w_2$. Let the promised φ be:

$$(\psi \wedge (\psi \succ_X \top) \wedge (\psi \succ_Y \top)) \rightarrow ((\psi \wedge (\psi \succ_X \top)) \approx_X (\psi \wedge (\psi \succ_Y \top))).$$

We indicate why φ is invalid. The antecedent of φ is easily seen to be satisfiable, and a ψ-world satisfying $\psi \wedge (\psi \succ_X \top)$ need not be the same world that satisfies $\psi \wedge (\psi \succ_Y \top))$; and u_X may be chosen to be injective.

On the other hand, suppose that model $\mathcal{M} = \langle D, \mathbb{W}, t, u, s\rangle$ is utility-invariant and let $w_0 \in \mathbb{W}$. Suppose that the antecedent of φ is satisfiable in $\mathcal{M}$ (otherwise, we are done). Then $(\psi \wedge (\psi \succ_X \top)[\mathcal{M}] \neq \emptyset$ and $(\psi \wedge (\psi \succ_Y \top))[\mathcal{M}] \neq \emptyset$. So, let $w_1 = s(w_0, (\psi \wedge (\psi \succ_X \top))[\mathcal{M}])$ and $w_2 = s(w_0, \psi \wedge (\psi \succ_Y \top)))[\mathcal{M}])$. Then each of w_1, w_2 satisfies ψ so $w_1 \simeq w_2$. Hence $u_X(w_1) = u_X(w_2)$ by utility-invariance. □

8 Anonymity

Our next topic concerns the manner in which utilities are associated with formulas. First, a condition is exhibited that makes the utility of a conjunction depend on just the utilities of each conjunct separately. According to this condition the vocabulary appearing in a conjunct is not permitted to influence the utility of the conjunction; rather, the conjunct contributes its utility "anonymously." A second condition is then introduced that entails a similar kind of anonymity for the contribution of utility indexes 1 and 2 to the aggregated utility $\{1, 2\}$. The material in this section is inspired by the discussion in Krantz et al. (1971, §7.2).

8.1 Decomposing the utility of conjunctions

Let a signature $(\mathbb{L}, \mathbb{S})$ be given with predicate $P \in \mathbb{L}$. Conjunctive anonymity with respect to P is expressed by the following formula. (To lighten notation, we suppress $X \in \mathbb{S}$ in subscripts.)

(34) $\varphi \stackrel{\text{def}}{=} \forall xy((Px \approx Py) \rightarrow \forall z((Px \wedge Pz) \approx (Py \wedge Pz)))$

The next proposition gives the sense in which φ causes the utility of $Px \wedge Py$ to be a function (F) of the utilities of Px and Py.

(35) Proposition: Let model $\mathcal{M} = \langle D, \mathbb{W}, t, u, s\rangle$ be given with $w_0 \in \varphi[\mathcal{M}]$. Then there is a function $F : \Re^2 \rightarrow \Re$ such that for all assignments d with $Px \wedge Py[\mathcal{M}, d] \neq \emptyset$,

$$u(s(w_0, Px \wedge Py[\mathcal{M}, d])) = F(\, u(s(w_0, Px[\mathcal{M}, d])),\ u(s(w_0, Py[\mathcal{M}, d]))\,).$$

Proof: For numbers of the form $u(s(w_0, Px[\mathcal{M}, d]))$ and $u(s(w_0, Py[\mathcal{M}, d]))$ define:

(36) $F(\, u(s(w_0, Px[\mathcal{M}, d])),\; u(s(w_0, Py[\mathcal{M}, d]))\,) \stackrel{\text{def}}{=} u(s(w_0, Px \wedge Py[\mathcal{M}, d]))$.

For all other numbers r_1, r_2, $F(r_1, r_2)$ is defined arbitrarily. We must show that F is a function. For this purpose, let variable q be given, and suppose that

(37) $u(s(w_0, Px[\mathcal{M}, d])) = u(s(w_0, Pq[\mathcal{M}, d]))$.

To finish the proof it suffices to show that

(38) $u(s(w_0, Px \wedge Py[\mathcal{M}, d])) = u(s(w_0, Pq \wedge Py[\mathcal{M}, d]))$,

the second argument of F being treated in the same way. It follows immediately from (37) that $w_0 \in (Px \approx Pq)[\mathcal{M}, d]$, hence by (34)

$$w_0 \in ((Px \wedge Py) \approx (Pq \wedge Py))[\mathcal{M}, d],$$

which implies (38). □

Observe that φ and Proposition (35) can be formulated with disjunction in place of conjunction — or with many other formulas. The proof proceeds in the same way.

8.2 Decomposing a complex utility index

Suppose for this section that the signature $(\mathbb{L}, \mathbb{S})$ contains unary $P \in \mathbb{L}$ along with $\{1\}, \{2\}, \{1, 2\} \in \mathbb{S}$. Define:

(39) $\varphi \stackrel{\text{def}}{=} \forall xy(\, ((Px \approx_1 Py) \wedge (Px \approx_2 Py)) \rightarrow (Px \approx_{1,2} Py)\,)$.

Then φ implies that the contributions of 1 and 2 to the complex utility index $\{1, 2\}$ can be separated then brought back together via a binary mapping on $\Re$. Specifically:

(40) Proposition: Let model $\mathcal{M} = \langle D, \mathbb{W}, t, u, s\rangle$ be given with $w_0 \in \varphi[\mathcal{M}]$. Then there is a function $F : \Re^2 \rightarrow \Re$ such that for all assignments d:

$$u_{1,2}(s(w_0, Px[\mathcal{M}, d])) = F(\, u_1(s(w_0, Px[\mathcal{M}, d])),\; u_2(s(w_0, Px[\mathcal{M}, d]))\,).$$

Proof: Call a pair $(p, q) \in \Re^2$ *critical* just in case there is an assignment d such that

(41) (a) $p = u_1(s(w_0, Px[\mathcal{M}, d]))$
(b) $q = u_2(s(w_0, Px[\mathcal{M}, d]))$.

Let $F : \Re^2 \rightarrow \Re$ be such that for any critical pair (p, q) as in(41), $F(p, q) = u_{1,2}(s(w_0, Px[\mathcal{M}, d]))$. The behavior of F on noncritical pairs is arbitrary. Suppose that for some assignment d':

(42) (a) $p = u_1(s(w_0, Px[\mathcal{M}, d'\,]))$
(b) $q = u_2(s(w_0, Px[\mathcal{M}, d'\,]))$.

To verify that F is a function, thereby completing the proof, we must show that

(43) $u_{1,2}(s(w_0, Px[\mathcal{M}, d\,])) = u_{1,2}(s(w_0, Px[\mathcal{M}, d'\,]))$.

Let y be a variable distinct from x, and let $d'' = d(d'(x)/y)$. From (41) and (42) we infer: $w_0 \in Px \approx_1 Py[\mathcal{M}, d''\,]$ and $w_0 \in Px \approx_2 Py[\mathcal{M}, d''\,]$. From (39) we then obtain $w_0 \in Px \approx_{1,2} Py[\mathcal{M}, d''\,]$ from which (43) is an immediate consequence. □

9 Arrow's theorem in the context of quantified preference logic

Finally, we illustrate how results in the theory of Social Welfare can be cast as constraints on the relation between the separate utility indexes $\{1\}\ldots\{k\}$ and their aggregate $\{1\ldots k\}$. For this purpose, we focus on Kenneth Arrow's classic theorem beginning with a review of its usual formulation (following Reny, 2001).

9.1 Review

Let A be a set of cardinality at least three. Let slo denote the set of strict linear orders (or *rankings*) on A, and let wlo be their weak counterparts. Fix a positive integer k. Members of slo are thought of as potential citizens in a community of size k. Each citizen expresses (rank order) preferences about the set A of agenda items (or "alternatives"). Any function from $\mathsf{slo}^k \to \mathsf{wlo}$ is called a *social welfare function*. For $C \in \mathsf{slo}^k$, the members of C are denoted by C_i. (C is a community of k citizens.)

Let f be a social welfare function, and consider four potential properties of f.

(44) Definition:
(a) (Universality): f is total.
(b) (Pareto efficiency): Let $a, b \in A$ and $C \in \mathsf{slo}^k$ be given. Suppose that for all $i \leq k$, a is ranked above b in C_i. Then a is ranked above b in $f(C)$.
(c) (Independence of irrelevant alternatives): Let $a, b \in A$ and $C, C' \in \mathsf{slo}^k$ be given. Suppose that for all $i \leq k$, a is ranked below b in C_i if and only if a is ranked below b in C'_i. Then a is ranked below b in $f(C)$ if and only if a is ranked below b in $f(C')$
(d) (Dictatorship): There is $i \leq k$ such that for all $C \in \mathsf{slo}^k$, $f(C) = C_i$.

(45) Theorem: Every social welfare function that satisfies Universality, Pareto efficiency, and Independence of irrelevant alternatives is dictatorial.

9.2 Reconstruction within preference logic

Let our signature include a monadic predicate P and utility indices $\{1\} \dots \{k\}$, $\{1 \dots k\}$. As usual, we abbreviate the index $\{i\}$ to just i. The language $\mathcal{L}(\mathbb{L}, \mathbb{S})$ is assumed to include distinct variables x, y, z possibly with subscripts, superscripts and primes. The formulas defined below are meant to recapitulate the four properties in Definition (44). We consider $m \geq 3$ agenda items.

Universality. Fix m variables $x^1 \dots x^m$ where $m \geq 3$. For variables $x_1 \dots x_m$, let $\chi(x_1 \dots x_m)$ be the formula that says that each of $x_1 \dots x_m$ is equal to exactly one of $x^1 \dots x^m$.

(46) Definition: (Universality): Let ψ be the conjunction of

$$\begin{gathered}\chi(x_1 \dots x_m) \wedge (Px_1 \succ_1 Px_2) \wedge (Px_2 \succ_1 Px_3) \wedge \cdots \wedge (Px_{m-1} \succ_1 Px_m) \\ \vdots \\ \chi(x_1 \dots x_m) \wedge (Px_1 \succ_k Px_2) \wedge (Px_2 \succ_k Px_3) \wedge \cdots \wedge (Px_{m-1} \succ_k Px_m)\end{gathered}$$

Let φ_{univ} be the universal closure of $\Diamond\psi$.

That is, each conjunct of ψ imposes a complete $\succ_i$-ordering on the Px_j where $1 \leq i \leq k$ and $1 \leq j \leq m$. So φ_{univ} is true in a model $\langle D, \mathbb{W}, t, u, s \rangle$ just in case every community is realized in some $w \in W$.

Pareto efficiency.

(47) Definition: Let φ_{pareto} be the universal closure of $\Box((Px \succ_1 Py \wedge \cdots \wedge Px \succ_k Py) \to Px \succ_{\{1 \dots k\}} Py)$.

Independence of irrelevant alternatives. Fix two variables x, y. For variables x', y', let $\psi(x', y')$ be the formula that says that each of x', y' is equal to exactly one of x, y.

(48) Definition: Let φ_{iia} be the universal closure of the formula

$$\begin{aligned}(\psi(x_1, y_1) \wedge \cdots \wedge \psi(x_k, y_k)) \to \\ (\Diamond((Px_1 \succ_1 Py_1) \wedge \cdots \wedge (Px_k \succ_k Py_k) \wedge (Px \succ_{\{1 \dots k\}} Py)) \to \\ \Box(((Px_1 \succ_1 Py_1) \wedge \cdots \wedge (Px_k \succ_k Py_k)) \to Px \succ_{\{1 \dots k\}} Py)).\end{aligned}$$

Then φ_{iia} is expresses that $\succ_{\{1 \dots k\}}$ has the property of independence of irrelevant alterantives (with respect to formulas of the form Pv).

Dictatorship.

(49) Definition: Let φ_{dict} be the disjunction of the following formulas.

$$\forall x_1 \dots x_m \Box(((Px_1 \succ_1 Px_2) \wedge (Px_2 \succ_1 Px_3) \dots (Px_{m-1} \succ_1 Px_m)) \leftrightarrow ((Px_1 \succ_{\{1\dots k\}} Px_2) \wedge (Px_2 \succ_{\{1\dots k\}} Px_3) \dots (Px_{m-1} \succ_{\{1\dots k\}} Px_m)))$$
$$\vdots$$
$$\forall x_1 \dots x_m \Box(((Px_1 \succ_k Px_2) \wedge (Px_2 \succ_k Px_3) \dots (Px_{m-1} \succ_k Px_m)) \leftrightarrow ((Px_1 \succ_{\{1\dots k\}} Px_2) \wedge (Px_2 \succ_{\{1\dots k\}} Px_3) \dots (Px_{m-1} \succ_{\{1\dots k\}} Px_m)))$$

Then φ_{dict} asserts that one of the individual indexes reveals the collective preference. Notice that dictatorship extends beyond the particular m-tuple that we might wish to fix at the start of the discussion. The (unique) dictator described in (49) determines all preferences.

9.3 Arrow's theorem revisited

The next theorem follows easily from the definitions above along with any proof of Arrow's Theorem.

(50) Theorem: In quantified preference logic:

$$\{\varphi_{univ}, \varphi_{pareto}, \varphi_{iia}\} \models \varphi_{dict}.$$

Appendix 1: Proof of Theorem (10)

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,

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

The proof of (51) will be based on a canonical model construction. In order to explain the construction we require the notion of *modal depth.*

(52) Definition: We define $\mu(\varphi)$, the modal depth of φ, by recursion on $\varphi \in \mathcal{L}(\mathbb{L}, \mathbb{S})$ 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 $\varphi \in \mathcal{L}(\mathbb{L}, \mathbb{S})$ 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 Σ is a set of formulas, we let $\nu(\Sigma) = \{\Box\varphi \mid \Box\varphi \in \Sigma\}$. If Σ and Σ' are sets of formulas, we say Σ is *compatible* with Σ' just in case $\nu(\Sigma) = \nu(\Sigma')$.

A set of formulas Σ is *consistent* just in case $\bot$ is not O-derivable from Σ; a set of formulas Σ is *maximally consistent* just in case it is consistent and no proper extension of it is consistent. We say a set of formulas Γ is *n-maximally consistent* if and only if there is a maximally consistent set Σ 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.

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

For each $n, m \geq 0$ and n-mcs Σ, 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 Σ. Given n-mcs Σ, let $\Xi^n(\Sigma)$ be the family of n-mcs's which are compatible with Σ. The collection of worlds $\mathbb{W}^{n,m}$ of $\mathcal{M}^{n,m}(\Sigma)$ is $\Xi^n(\Sigma) \times \{0, \ldots, m\}$. In order to specify the remaining components of $\mathcal{M}^{n,m}(\Sigma)$, we fix an n-mcs Σ_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 φ 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 φ 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 φ of modal depth 0, $w \in \varphi[\mathcal{M}^n_0]$ if and only if $\varphi \in w$. This follows immediately from (53), 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 $\leq j$,

(54) $w \in \varphi[\mathcal{M}^n_j]$ 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 Ω^n_j defined as follows.

(55) For all φ and ψ 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

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

In order to establish (56)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 φ, ψ, and θ are formulas of modal depth $\leq j$ and $\varphi[\mathcal{M}^n_j] = \psi[\mathcal{M}^n_j]$, then

(57) $\langle \varphi[\mathcal{M}^n_j], \theta[\mathcal{M}^n_j] \rangle \in v^n_{j+1}(w)$ 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 φ and θ and ψ and θ reversed. So suppose that

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

It follows at once from (58), (54), and (53), recalling the fact that every $w' \in \mathbb{W}^n$ is an n-mcs, that

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

Let χ be the conjunction of the formulas in $\nu(w)$. It follows from (59) and the definition of $\mathbb{W}^n$ that

(60) $\chi \rightarrow (\varphi \leftrightarrow \psi)$ is a theorem of O,

for otherwise, there would be an n-mcs $w' \in \mathbb{W}^n$ with $\neg(\varphi \leftrightarrow \psi) \in w'$ contradicting (59). Since the theorems of O are closed under necessitation, (60) implies that

(61) $\Box(\chi \rightarrow (\varphi \leftrightarrow \psi))$ is a theorem of 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 χ is a "boxed" formula, it follows from (53) that

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

It follows from (61), (62), and (53), and the S5 modal principle

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

that

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

But then, by (53), (63), Axiom (8)c and the fact that w is an n-mcs,

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

Therefore v^n_{j+1} is well-defined, since (57) follows directly from (64) and (55).

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

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

Toward establishing condition (65)a, we show that if A is in the field of $v^n_{j+1}(w)$, then $A \neq \emptyset$. So suppose that

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

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 (54), 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 χ be the conjunction of the formulas in $\nu(w)$. Arguing as we did for (63), we may conclude that $(\chi \rightarrow \neg\varphi)$ is a theorem of O, and thence that $\Box\neg\varphi \in w'$ for all $w' \in \mathbb{W}^n$. It follows immediately by (53) that

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

On the other hand, it is a direct consequence of (55) and (66) that

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

It follows from (53), (68), and the right-to-left direction of Axiom (8)b that

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

But (69) contradicts (67), thereby establishing that $\varphi[\mathcal{M}^n_{j+1}] \neq \emptyset$.

In order to establish (65)b, suppose that φ, ψ, and θ are formulas of modal depth $\leq j$, $w \in \mathbb{W}^n$ and that

(70) $\langle \varphi[\mathcal{M}_j^n], \psi[\mathcal{M}_j^n] \rangle \in v_{j+1}^n(w)$ and $\langle \psi[\mathcal{M}_j^n], \theta[\mathcal{M}_j^n] \rangle \in v_{j+1}^n(w)$.

It follows immediately from (70) and (55) that

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

Therefore, by Axiom (8)a and (53),

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

Hence, by (72) and (55)

(73) $\langle \varphi[\mathcal{M}_j^n], \theta[\mathcal{M}_j^n] \rangle \in v_{j+1}^n(w)$.

We leave the argument for (65)c to the reader – it is virtually the same as the argument for (65)b, using the left-to-right direction of Axiom (8)b in place of Axiom (8)a.

We now verify (56)b. Note that by (56)a, for every φ with $\mu(\varphi) \leq j+1$, $\varphi[\mathcal{M}_{j+1}^n]$ is a well-defined. It is clear from (55) and the choice of v_0^n as the empty partial function that for all $0 \leq i \leq j$ and all $w \in \mathbb{W}^n$, $v_i^n(w) \subseteq v_{i+1}^n(w)$. It follows at once that

(74) for all φ of modal depth $\leq j$, $\varphi[\mathcal{M}_j^n] = \varphi[\mathcal{M}_{j+1}^n]$.

Hence, by (54) and (74), it follows at once that in order to prove (56)b, we need only show that for every $w \in \mathbb{W}^n$ and every formula φ, if $\mu(\varphi) = j+1$, then

(75) $w \in \varphi[\mathcal{M}_{j+1}^n]$ 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 (53) and the fact that all $w \in \mathbb{W}^n$ are n-mcs's, in order to establish (75), it suffices to show that for all ψ and θ with $\mu(\psi), \mu(\theta) \leq j$,

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

But (76) is an immediate consequence of (55). This concludes the construction of the partial generalized model $\mathcal{M}_n^n$. By (56)b, it has the "canonical model property"

(77) for all φ 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 (77) that $\mathcal{M}^n$ satisfies Σ_0. Since every formula φ is contained in an n-mcs for some n, this concludes the proof of (51).

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 Ω^n_{n-1} induced by $\mathcal{M}$, for this will establish that every consistent φ is satisfied by some basic model. It is easy to see that no matter how m is chosen,

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

Let Π be the set of pre-orderings of Ω^n_{n-1}, and choose $m \geq \mathsf{card}(\Pi) \cdot \mathsf{card}(\Omega^n_{n-1})$. It then follows from (78) that there is a function $f : \Pi \times \Omega^n_{n-1} \mapsto \mathbb{W}^n$ such that

(79) (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 (79) that we may define u in such a way that

(80) 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.

(81) 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 (80) and (81) 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 Ω^n_{n-1} induced by $\mathcal{M}$. □

Appendix 2: Proofs of Propositions (25), (26), and (27)

All three proofs elaborate a construction that appears in the demonstration of Theorem (55) in Osherson and Weinstein (2012). Specifically, the earlier construction can be adapted to show that there is an effective translation from sentences $\varphi \in \mathcal{L}(\mathbb{L}, \mathbb{S})$ to formulas $\varphi^\dagger(x)$ of first-order logic, and a map from preorder models $\mathcal{M} = \langle D, \mathbb{W}, t, \trianglerighteq, s \rangle$ to relational structures $\mathcal{F}_\mathcal{M}$ such that

(82) $w \in \varphi[\mathcal{M}]$ iff $\mathcal{F}_\mathcal{M} \models \varphi^\dagger[w]$.

Moreover, assuming that $(\mathbb{L}, \mathbb{S})$ is recursive, there is a recursively axiomatizable first-order theory T in the signature of $\mathcal{F}_\mathcal{M}$ such that

(83) for every preorder model $\mathcal{M}$, $\mathcal{F}_\mathcal{M} \models T$

and

(84) for every first-order structure A, if $A \models T$, then for some preorder model $\mathcal{M}$, $A = \mathcal{F}_{\mathcal{M}}$.

Proposition (27) now follows from the completeness theorem for first-order logic, since (82), (83), and (84) imply that $\varphi \in \mathcal{L}(\mathbb{L}, \mathbb{S})$ is valid in preorder logic if and only if $\forall x \varphi^{\dagger}(x)$ is a consequence of T. In like fashion, Proposition (25) follows from the Löwenheim-Skolem Theorem for first-order logic. Proposition (26) now follows immediately, since every countable preorder model is induced by a corresponding utility model, a consequence of the fact that the rational numbers are universal among countable linear orders. $\square$

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.

S. O. Hansson. A new semantical approach to the logic of preference. *Erkenntnis*, 31(1):1–42, 1989.

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

D. H. Krantz, R. D. Luce, P. Suppes, and A. Tversky. *Foundations of Measurement*, volume I. Academic Press, New York, 1971.

S. Kripke. The undecidability of monadic modal quantification theory. *Zeitschr. f. math. Logik und Grundlagen d. Math.*, 8:113–116, 1962.

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. *The Review of Symbolic Logic*, 5 (1):122–147, 2012.

Philip J. Reny. Arrow's Theorem and the Gibbard-Satterthwaite Theorem: A Unified Approach. *Economics Letters*, 70:99–105, 2001.

N. Rescher. Semantic foundations for a the logic of preference. In *The Logic of Decision and Action*. University of Pittsburgh Press, 1967.

R. Stalnaker. A theory of conditionals. In N. Rescher, editor, *Studies in logical theory*. Blackwell, Oxford UK, 1968.

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.