# Preference based on reasons*

Daniel Osherson
Princeton University

Scott Weinstein
University of Pennsylvania

May 6, 2011

**Abstract**

We describe a logic of preference in which modal connectives reflect reasons to desire that a sentence be true. Various conditions on models are introduced and analyzed.

## 1 Introduction

Sometimes preferences are the result of identifiable reasons, as when you install a fire alarm for concern about safety. This suggests that studying reasons along with preference might illuminate both. An obstacle to such a project is the multitude of reasons for and against a given action that combine subliminally to yield a decision, as when you go ahead with the fire alarm despite the bother, cost, and false alerts. Unfortunately, the underlying calculus of reason aggregation seems largely hidden from introspection.

The centrality of reasons to action and rationality is nonetheless sufficient motive to persevere in their analysis despite the difficulty. The insights that can be achieved are illustrated in a recent paper by Dietrich and List (2009). These authors demonstrate a representation theorem relating choice to the respective bundles of reasons that apply to the options in play; the axioms needed for their result are remarkably weak. Dietrich and List's ground breaking work clarifies several issues, among them the significance of combining reasons (their analysis rests not on individual reasons but on sets of them). The present investigation attempts to fill in some additional detail about the same topic. Specifically, we advance a modal logic in which different reasons for a preference can be aggregated in various ways.

*We thank Horacio Arló-Costa for very helpful comments on an earlier draft. Contact: osherson@princeton.edu, weinstein@cis.upenn.edu

Our inquiry is preceded by several studies of the logic of preference, beginning with von Wright (1963). More contemporary work includes systems designed to elucidate the interaction between choice and epistemic possibility (see Lang et al., 2003; van Benthem et al., 2009). Of particular relevance is Liu (2008, Ch. 3). This work introduces "priorities" (which function like reasons in the present setting) that are ordered by importance, and integrated into a formal language of preference and belief. Several ways of extracting preferences from priorities are explored. The interplay of preferences and beliefs is also analyzed, along with the impact of updating belief and preference. Liu's work is closest to the approach taken here inasmuch as it develops a modal language and associated semantics. The conceptual framework is nonetheless different from ours, as will become clear presently. Another fruitful perspective on the integration of preferences issues from the graph-theoretic approach advanced in Andréka et al. (2002); different graphs represent alternative orderings of the alternatives in play, and might be considered separate reasons for choice among them. Within a yet different tradition, multi-attribute utility theory (Keeney and Raiffa, 1993) bears directly on reason aggregation through the combination of utilities based on separate dimensions. The theory has revealed exact conditions under which aggregation can proceed additively but it does not explore the logical structure of reasons and preference, as we shall do here.

To keep the present project manageable, conceptual issues about the nature of reasons and their role in rational discourse will be set aside. An entry to this literature is provided by Dietrich and List (2009), and sustained discussion is available in Pettit (2002). Of course, the reasons that come to mind are not necessarily those that govern choice (see, for example, Messick, 1985; Haidt, 2001). Our theory is indifferent to this distinction but it will be more natural to limit examples to conscious, effective reasons. The case of the fire alarm serves to convey the character of our theory. Specifically, we picture an agent who imagines a world *that resembles the actual one* but with a fire alarm, and another world (possibly his own) without one. The agent then compares the two worlds according to various utility scales (one that measures safety, another cost, and so forth), as well as a distinct utility scale that takes all the individual scales into account. Our formalism is designed to capture this picture.

We proceed by first introducing the language under investigation. Informal glosses for some of its formulas will clarify the ideas in play. Next the semantics of our logic is presented, followed by consideration of subclasses of models that meet various conditions. We then turn to decidability issues. A discussion of open questions is provided at the end.

## 2 Language

The present section introduces a family of modal languages, and discusses the intended meaning of the modality. A *language of reason-based preference* is determined by its *signature*, which consists of:

(a) a non-empty set $\mathbb{P}$ of propositional variables

(b) a nonempty collection $\mathbb{S}$ of nonempty subsets of $\mathbb{N}$ (the set $\{0, 1, \dots\}$ of natural numbers)

The language of reason-based preference 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:

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

Moreover, we rely on the following abbreviations.

$$\begin{array}{rcl}
(\varphi \vee \psi) & \text{for} & \neg(\neg\varphi \wedge \neg\psi) \\
(\varphi \to \psi) & \text{for} & (\neg\varphi \vee \psi) \\
(\varphi \leftrightarrow \psi) & \text{for} & ((\varphi \to \psi) \wedge (\psi \to \varphi)) \\
(\varphi \succeq_{1\dots k} \psi) & \text{for} & (\varphi \succeq_{\{1\dots k\}} \psi) \\
(\varphi \succ_X \psi) & \text{for} & (\varphi \succeq_X \psi) \wedge \neg(\psi \succeq_X \varphi) \\
(\varphi \approx_X \psi) & \text{for} & (\varphi \succeq_X \psi) \wedge (\psi \succeq_X \varphi) \\
(\varphi \preceq_X \psi) & \text{for} & (\psi \succeq_X \varphi) \\
(\varphi \prec_X \psi) & \text{for} & (\psi \succ_X \varphi) \\
\top & \text{for} & (p \to p) \\
\bot & \text{for} & \neg\top
\end{array}$$

The formula $\varphi \succ_1 \psi$ is to be understood along the following lines. Fix an agent $\mathcal{A}$ whose reasoning is at issue. Let $u_1$ be a utility scale that reflects some dimension of interest to $\mathcal{A}$. Then $\varphi \succ_1 \psi$ is true just in case:

> $\mathcal{A}$ envisions a situation in which $\varphi$ is true and that otherwise differs little from his actual situation (if $\varphi$ is already true then $\mathcal{A}$'s actual situation may well be the one he envisions). Likewise, $\mathcal{A}$ envisions a second situation that is like his actual situation except that $\psi$ is true. Finally, the utility according to $u_1$ of the first imagined situation exceeds that of the second.

In the fire alarm example, $\mathcal{A}$ envisions his home with a new fire alarm, but with the same furniture, cat and fireplace as before. Home with no fire alarm is the actual situation, hence especially easy to envision. If $u_1$ measures safety, and $p$ is "$\mathcal{A}$ will purchase a fire alarm" then $p \succ_1 \neg p$ holds inasmuch as the alarm improves safety. (Since $\top$ is also true in $\mathcal{A}$'s situation, $p \succ_1 \neg p$ is materially equivalent to $p \succ_1 \top$.) If $\mathcal{A}$ is short on cash, and $u_2$ reflects finances then $p \prec_2 \neg p$ is true, whereas the status of $p \succ_{1,2} \neg p$ depends on the manner in which utilities are aggregated (e.g., averaging, minimum, etc.). More generally, we allow preferences $\varphi \succ_X \psi$ between arbitrary formulas $\varphi, \psi$ in view of the (possibly multiple) utilities in $X \in \mathbb{S}$. The formula $\varphi \succ_X \psi$ thus represents $\mathcal{A}$'s preference for $\varphi$ over $\psi$ when $\mathcal{A}$ brings to mind just the reasons indexed in $X$. If $\bigcup \mathbb{S} \in \mathbb{S}$ then preference *tout court* for $\varphi$ over $\psi$ is represented by $\varphi \succ_{\bigcup \mathbb{S}} \psi$, that is, taking account of all reasons in play.

If our agent is presumed to be *moral* then reasons are meant, very roughly, to be *good* (at least, not *bad*). Morality will here be left unexplored, however. Instead, $\mathcal{A}$ is conceived as logically empowered but otherwise like the rest of us. Also notice how little any of this has to do with *reasons to believe* (except for odd cases like being rewarded for reaching genuine religious conviction). Only reasons for preference will be at issue. There is nonetheless one connection to belief that bears comment.

The appeal to situations that differ minimally from the actual one, except for satisfying a given formula, is familiar from well known theories of counterfactual conditionals (Stalnaker, 1968; Lewis, 1973). It thus risks bedevilment from a similar range of cases. Suppose, for example, that $p$ is "Winter ends a little earlier than last year." Then too many $p$-worlds offer themselves as alternatives to the actual world (since the set of shorter winters has no member closest to last year's winter). The present endeavor, however, may not be as vulnerable as the earlier one to such cases. For it here suffices that the reasoning agent bring to mind a cognitively salient situation that satisfies the formula in question (e.g., winter a week shorter), not necessarily the maximally similar one. Indeed, the agent may not be prepared to identify the maximally similar

$p$-world, or even to understand such an idea. Consistent with this relaxed attitude, to each consistent proposition our semantics assigns a world that represents life were the proposition true, where the choice of world may depend on the agent's current position. Some constraints on the choice will be examined, but otherwise the reasoning agent is on his own. We take all this to be a rough idealization of what happens in actual decision-making. One imagines an alternative situation that satisfies the proposition at issue, then evaluates it along various dimensions (i.e., utility scales).

The utility scales that determine the truth of modal formulas are intended to measure the impact on choice of specific considerations, e.g., cost, health, professional advancement. Because deliberation is assumed to transpire in a single mind (the agent's), aggregation of different scales into an overall value seems feasible; indeed, people do it all the time. For simplicity, the scales express *expected* utilities, that is, with probabilities already factored in. Thus, the safety improvements envisioned from installing a fire alarm already integrate the agent's confidence that the device will work as advertised.

Even when utility scales are kept separate, languages of reason-based preference allow interesting interactions. For an illustration, first observe that $\varphi \succ_i \top$ means (roughly) that the $u_i$-utility of the envisioned $\varphi$-world exceeds that of the actual world. Now consider:

$$(p \succ_1 \top) \succ_2 \top$$

This says that the agent has a $u_2$-reason for there being a $u_1$-reason in favor of $p$. For example, let $p$ be the assertion that you buy a low-power automobile. Let $u_2$-utility be pecuniary: $u_2(w_1) > u_2(w_2)$ iff you have more cash in $w_1$ compared to $w_2$. Let $u_1$-utility reflect personal safety: $u_1(w_1) > u_1(w_2)$ iff you incur less risk traveling in $w_1$ than in $w_2$. Then the formula asserts that it's in your financial interest that your buying a low-power automobile is in your safety interest — which might well be true inasmuch as low-power vehicles are cheaper.

We conclude this section with another illustration of the interaction of individual utility scales. Consider:

$$\neg q \succ_1 (p \succ_2 q)$$

This says that the agent $u_1$-prefers that $q$ be false rather than $u_2$-prefer $p$ over $q$. For example, let $q$ be the assertion that your brother runs for mayor, and let $p$ be that Miss Smith (no relation) also runs. Let $u_1$-utility measure family pride, and let $u_2$-utility measure political value to an ailing municipality. Then the formula asserts that from the point of view of family pride, you'd rather that your brother not run for mayor than that Miss Smith be the superior candidate.

## 3 Semantics

We now provide a formal semantics designed to capture the intuitive picture elaborated in the preceding section. Several preliminary concepts are needed. Fundamental is the choice of a nonempty set $\mathbb{W}$ to embody the imaginative possibilities ("worlds") available to an agent in the course of practical deliberation. Subsets of $\mathbb{W}$ are called *propositions*. As discussed above, given a nonempty proposition $A$ and a world $w$, an agent envisions a salient alternative to $w$ among the worlds in $A$. (If $w \in A$ then the "alternative" might be $w$ itself.) We formalize this idea as follows.

(1) DEFINITION: A *selection function $s$ over* $\mathbb{W}$ 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 $\emptyset \neq A \subseteq \mathbb{W}$, $s(w, A) \in A$.

Thus, $s(w, A)$ is a choice of world to represent $A$, where the choice depends on $w$. (The idea is that $s$ chooses a member of $A$ that is similar to $w$.)

Next, recall that each world can be evaluated according to various utility scales, each involving one or more dimensions of value. All the scales are 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 $u_X(w)$.

Let $\mathbb{P}$ be a nonempty set of propositional variables. Our last preliminary is the assignment of a subset of worlds 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 $w \in t(p)$ (and otherwise false). This is all we need to introduce models.

(4) DEFINITION: A *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 $\mathbb{P}$.

It remains to specify the proposition (set of worlds) expressed by a formula $\varphi$ in a model $\mathcal{M}$. This proposition is denoted $\varphi[\mathcal{M}]$, and defined inductively 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:

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

Observe 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 (intuitively) world $w$ satisfies $(\theta \succeq_X \psi)$ in $\mathcal{M}$ iff the world selected from $A$ as closest to $w$ has utility no less than that of the world selected from $B$ as closest to $w$. A word of caution: the existential requirement on the truth of $(\theta \succeq_X \psi)$ allows $\neg(\theta \succeq_X \psi)[\mathcal{M}] \neq (\theta \prec_X \psi)[\mathcal{M}]$. Indeed, if $\theta[\mathcal{M}] = \emptyset$ then $\neg(\theta \succeq_X \psi)[\mathcal{M}] = \mathbb{W}$ but $(\theta \prec_X \psi)[\mathcal{M}] = \emptyset$.

The following definition imports standard terminology and notation to the present context.

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

(a) $\mathcal{M}$ *satisfies* $\varphi$ just in case $\varphi[\mathcal{M}] \neq \emptyset$.

(b) $\varphi$ is *valid* in $\mathcal{M}$ just in case $\varphi[\mathcal{M}] = \mathbb{W}$.

(c) $\varphi$ is *valid* just in case $\varphi$ is valid in every model.

(d) $\varphi$ is *valid* in a given class $C$ of models just in case $\varphi$ is valid in every model of $C$.

We use related expressions (like "satisfiable") in the obvious way. It is noteworthy that our language allows expression of the global modality (see Blackburn et al. 2001, §2.1). Choose any $X \in \mathbb{S}$, and for $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ let:

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

Then unwinding clause (5)d of our semantic definition yields:

(8) 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 $\varphi[\mathcal{M}] \neq \emptyset$.

It follows from Proposition (8) that the axioms of S5 are valid for $\Box$ and $\Diamond$. Other valid formulas of our language include the following (proofs are easy). For all $X \in \mathbb{S}$, and $\varphi, \psi, \theta \in \mathcal{L}(\mathbb{P}, \mathbb{S})$:

$\models ((\varphi \succeq_X \psi) \wedge (\psi \succeq_X \theta)) \rightarrow (\varphi \succeq_X \theta)$

$\models (\Diamond\varphi \wedge \Diamond\psi) \rightarrow ((\varphi \succeq_X \psi) \vee (\psi \succeq_X \varphi))$

$\models (\Diamond\varphi \wedge \Diamond\psi) \leftrightarrow (\neg(\varphi \succeq_X \psi) \leftrightarrow (\psi \succ_X \varphi))$

$\models \neg(\bot \succeq_X \varphi)$ and $\models \neg(\varphi \succeq_X \bot)$

$\models \Diamond\varphi \rightarrow (\varphi \approx_X \psi)$ if $\varphi$ and $\psi$ are equivalent.

In the next section we introduce classes of structures which conform to several substantive hypotheses about selection and utility, and we explore the logical principles they validate.

## 4 Stronger theories

The present section advances some natural conditions on models. (Several of the conditions have been discussed within order-theoretic approaches to preference, for example, in Levi, 1986, Ch. 6.) For this section, let model $\mathcal{M} = (\mathbb{W}, s, u, t)$ have signature $(\mathbb{P}, \mathbb{S})$, and suppose that $p, q, r \in \mathbb{P}$.

### 4.1 Reflexivity

If a world $w$ satisfies a formula $\varphi$ then the "nearest" $\varphi$-world is intuitively $w$ itself. This condition is not imposed on selection functions by Definition (1) but can be added as follows.

(9) Definition: $\mathcal{M}$ is *reflexive* just in case for all $w \in \mathbb{W}$ and $A \subseteq \mathbb{W}$, if $w \in A$, then $s(w, A) = w$.

The formula exhibited in the following proposition illustrates the impact of reflexivity. It says that a given proposition is at least as good as the status quo or its negation is.

(10) Proposition: Let $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ be $(p \succeq_X \top) \vee (\neg p \succeq_X \top)$. Then $\varphi$ is invalid but valid in the class of reflexive models.

Proof: To verify the invalidity of $\varphi$, suppose that $\mathbb{W} = \{w_0, w_1, w_2\}$, $t(p) = \{w_0, w_1\}$, $s(w_0, \{w_0, w_1\}) = s(w_0, p[\mathcal{M}]) = w_1$, $s(w_0, \{w_2\}) = s(w_0, \neg p[\mathcal{M}]) = w_2$, $s(w_0, \mathbb{W}) = s(w_0, \top[\mathcal{M}]) = w_0$, and $u_X(w_0) > u_X(w_1), u_X(w_2)$. Then it is easy to see that $w_0 \notin \varphi[\mathcal{M}]$ hence $\varphi$ is not valid.

On the other hand, suppose that $\mathcal{M}$ is reflexive, and let $w_0 \in \mathbb{W}$. Then either $w_0 \in p[\mathcal{M}]$ or $w_0 \in \neg p[\mathcal{M}]$, say the former (the other case is parallel). By reflexivity, $s(w_0, p[\mathcal{M}] = w_0$. Likewise, $w_0 \in \top[\mathcal{M}] = \mathbb{W}$, so again by reflexivity, $s(w_0, \top[\mathcal{M}] = w_0$. Since $u_X(w_0) \geq u_X(w_0)$, $w_0 \in \varphi[\mathcal{M}]$. □

Reflexivity entails that some formulas are satisfied only by infinite models.

(11) Proposition: There is $\varphi \in \mathcal{L}$ such that $\varphi$ is satisfied by some infinite reflexive model but by no finite reflexive model.

Proof: Suppose that $X \in \mathbb{S}$, and let $\varphi$ be the conjunction of the following formulas.

(12) (a) $\Box(p \rightarrow (p \prec_X \neg p))$
(b) $\Box(\neg p \rightarrow (\neg p \prec_X p))$

It is easy to verify that $\varphi$ is satisfied by a model whose worlds form an $\omega$-sequence when ordered by $u_X$, and which alternate between satisfying $p$ and $\neg p$. On the other hand, suppose for a contradiction that $\varphi$ is satisfied by finite model $\mathcal{M} = (\mathbb{W}, s, u, t)$. Then some $w_0 \in \mathbb{W}$ has maximum $u_X$ utility. Suppose that $w_0$ satisfies $p$ (the other case is parallel). Then (12)a and Reflexivity imply that there is $w_1 \in \mathbb{W}$ satisfying $\neg p$ such that $u_X(w_0) < u_X(w_1)$. This contradicts the choice of $w_0$ as having maximum $u_X$ utility. □

### 4.2 Regularity

If you think that living in Boston is most similar to your current situation among the set of all addresses in New England then shouldn't you think that living in Boston is most similar to your current situation among the set of all addresses in Massachusetts? A similar principle is standardly applied to choice (Sen, 1971) even though its violation has been documented in several empirical studies (for example, Payne and Puto, 1982; Tentori et al., 2001). In the present setting, we are led to the following constraint on selection.

(13) DEFINITION: $\mathcal{M}$ is *regular* just in case for all $w \in \mathbb{W}$, nonempty $A \subseteq B \subseteq \mathbb{W}$, and $w_1 \in A$: If $s(w, B) = w_1$ then $s(w, A) = w_1$.

Regularity validates the formula appearing in the next proposition. An instance is this: If buying either a Ford or a Chevy makes more sense than buying a Toyota then either it makes more sense to buy a Ford than a Toyota, or it makes more sense to buy a Chevy than a Toyota (or both).

(14) PROPOSITION: Let $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ be $((p \vee q) \succ_X r) \rightarrow ((p \succ_X r) \vee (q \succ_X r))$. Then $\varphi$ is invalid but valid in the class of regular models.

PROOF: A counter model for $\varphi$ is easy to devise. To show validity in the regular models, suppose that $\mathcal{M}$ is regular, and let $w \in ((p \vee q) \succ_X r)[\mathcal{M}]$ be given. Then there are $w_1, w_2 \in \mathbb{W}$ with:

(15) (a) $w_1 = s(w, (p \vee q)[\mathcal{M}])$,

(b) $w_2 = s(w, r[\mathcal{M}])$, and

(c) $u_X(w_1) > u_X(w_2)$.

By (15)a, either $w_1 \in t(p)$ or $w_1 \in t(q)$, say the latter (the other case is parallel). Since $q[\mathcal{M}] \subseteq (p \vee q)[\mathcal{M}]$, it follows from regularity that $w_1 = s(w, q[\mathcal{M}])$. In view of (15)bc, $w \in (q \succ_X r)[\mathcal{M}]$. $\square$

The combination of reflexivity and regularity validates the following formula, which exhibits modal embedding.

(16) $((p \prec_1 \top) \succ_2 (q \prec_1 \top)) \rightarrow (\neg p \succ_2 \neg q)$

For an instance, suppose that $p, q$ represent plans for new shopping malls, and that $u_1, u_2$ measure their political and ecological interest, respectively. Then (16) asserts: If it is ecologically better for $p$ than for $q$ to politically backfire then abstaining from $p$ is ecologically better than abstaining from $q$.

(17) Proposition: Formula (16) is valid in the class of models that are reflexive and regular.

Proof: Let reflexive, regular model $\mathcal{M} = (\mathbb{W}, s, u, t)$ and $w \in \mathbb{W}$ be given. Suppose that:

(18) $w \in ((p \prec_1 \top) \succ_2 (q \prec_1 \top))[\mathcal{M}]$.

We must show:

(19) $w \in (\neg p \succ_2 \neg q)[\mathcal{M}]$.

By (18), there are $w_1, w_2 \in \mathbb{W}$ with:

(20) (a) $w_1 = s(w, (p \prec_1 \top)[\mathcal{M}])$,
(b) $w_2 = s(w, (q \prec_1 \top)[\mathcal{M}])$,
(c) $u_2(w_1) > u_2(w_2)$.

By reflexivity, it is easy to verify:

(21) (a) $(p \prec_1 \top)[\mathcal{M}] \subseteq \neg p[\mathcal{M}]$,
(b) $(q \prec_1 \top)[\mathcal{M}] \subseteq \neg q[\mathcal{M}]$.

So by (20)ab and (21), we have $\neg p[\mathcal{M}] \neq \emptyset$ and $\neg q[\mathcal{M}] \neq \emptyset$. Hence there are $w_1^*, w_2^* \in \mathbb{W}$ with:

(22) (a) $w_1^* = s(w, \neg p[\mathcal{M}])$,
(b) $w_2^* = s(w, \neg q[\mathcal{M}])$.

But by (20)a, (21)a, (22)a and regularity, $w_1^* = w_1$. Likewise, by (20)b, (21)b, (22)b and regularity, $w_2^* = w_2$. Thus, (20)c implies $u_2(w_1^*) > u_2(w_2^*)$ which together with (22) yields (19). □

Recall that a signature $(\mathbb{P}, \mathbb{S})$ is assumed as given.

(23) Definition:

(a) Let a collection $\mathbb{W}$ of worlds, a selection function $s$ over $\mathbb{W}$, and a utility function $u$ over $\mathbb{W}$ and $\mathbb{S}$ be given. We call the triple $(\mathbb{W}, s, u)$ a *frame*. The pair $(\mathbb{W}, s)$ is called a *partial frame*.

(b) Given a frame $(\mathbb{W}, s, u)$, we call $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ *valid in* $(\mathbb{W}, s, u)$ just in case for every truth-assignment $t$ over $\mathbb{W}$ and $\mathbb{P}$, $\varphi[\mathcal{M}] = \mathbb{W}$, where $\mathcal{M} = (\mathbb{W}, s, u, t)$.

(c) Given a partial frame $(\mathbb{W}, s)$, we call $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ *valid in* $(\mathbb{W}, s)$ just in case for every utility function $u$ over $\mathbb{W}$ and $\mathbb{S}$, and every truth-assignment $t$ over $\mathbb{W}$ and $\mathbb{P}$, $\varphi[\mathcal{M}] = \mathbb{W}$, where $\mathcal{M} = (\mathbb{W}, s, u, t)$.

It is clear that the regularity of a model depends only on its underlying partial frame. It is therefore natural to qualify a frame or partial frame as regular just in case all of its completions are regular. Furthermore, for a frame or partial frame $\mathcal{F}$, we call $\Sigma \subseteq \mathcal{L}(\mathbb{P}, \mathbb{S})$ valid in $\mathcal{F}$ if and only if all members of $\Sigma$ are valid in $\mathcal{F}$.

(24) Theorem: There is no $\Sigma \subseteq \mathcal{L}(\mathbb{P}, \mathbb{S})$ such that for all frames $\mathcal{F} = (\mathbb{W}, s, u)$, $\Sigma$ is valid in $\mathcal{F}$ if and only if $\mathcal{F}$ is regular.

Proof of (24) is deferred to Section 6.

(25) Theorem: Let $X \in \mathbb{S}$ be given. For all partial frames $\mathcal{F}$, $\mathcal{F}$ is regular if and only if

$$p \preceq_X q \rightarrow p \preceq_X (p \vee q)$$

is valid in $\mathcal{F}$.

Proof: The right-to-left direction is easy. For the other direction, let $\mathcal{F} = (\mathbb{W}, s)$ be a nonregular partial frame. We exhibit a utility function $u$ and a truth-assignment $t$ such that

$$p \preceq_X q \rightarrow p \preceq_X (p \vee q)[\,(\mathbb{W}, s, u, t)\,] \neq \mathbb{W}.$$

Since $\mathcal{F}$ is not regular, we may choose propositions $A, B \subseteq \mathbb{W}$ and $w \in \mathbb{W}$ such that $A \subseteq B$, $s(w, B) \in A$, but $s(w, B) \neq s(w, A)$. Let truth-assignment $t$ be such that $t(p) = A$ and $t(q) = B \backslash A$. Choose $u_X$ so that $u_X(s(w, A)) > u_X(s(w, B))$ but $u_X(s(w, A)) = u_X(s(w, B \backslash A))$. Then it is easy to see that

$$w \notin p \preceq_X q \rightarrow p \preceq_X (p \vee q)[\,(\mathbb{W}, s, u, t)\,].$$

$\square$

### 4.3 Lexicographic ordering

Let us consider a strengthened form of regularity.

(26) Definition: $\mathcal{M}$ is *lexicographic* just in case there is a well order $R$ of $\mathbb{W}$ such that for all $w \in \mathbb{W}$ and $A \subseteq \mathbb{W}$, $s(w, A)$ is the $R$-least member of $A$.

All lexicographic models are regular but not vice versa. The added constraint imposed by the lexicographic property validates some additional formulas.

(27) Proposition: Let $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ be $(p \succeq_X q) \rightarrow \Box(p \succeq_X q)$. Then $\varphi$ is false in some regular model but valid in the class of lexicographic models.

The box $\Box$ was defined in (7), above. The proof of (27) is elementary.

A generalization of lexicographic ordering may be defined as follows.

(28) Definition: A selection function $s$ over $\mathbb{W}$ is *proposition driven* just in case for all $w_1, w_2 \in \mathbb{W}$ and $\emptyset \neq A \subseteq \mathbb{W}$, $s(w_1, A) = s(w_2, A)$.

That is, proposition driven selection functions ignore their first arguments. Lexicographic ordering implies proposition drivenness; the next proposition shows the former to be a stronger condition than the latter.

(29) Proposition: There is a formula satisfiable in the class of proposition driven models but not in the class of lexicographic models.

Proof: Let $\varphi \in \mathcal{L}$ be the conjunction of the following formulas.

$$(p \vee q) \succ_X \top \qquad p \prec_X \top \qquad q \prec_X \top$$

It is easy to verify that no regular model satisfies $\varphi$ but that some proposition driven model does. Since lexicographic ordering implies regularity, the proposition follows immediately. $\Box$

Finally, we record the following fact.

(30) Proposition: A model is proposition driven and regular if and only if it is lexicographic.

Proof: We observed above the right-to-left direction. For the other direction we proceed as follows. Suppose $(\mathbb{W}, s, u, t)$ is regular and proposition driven. For proposition $P$ we write $s(P)$ for $s(w, P)$ (legitimate by proposition driven-ness). Define by transfinite recursion a well-ordering of $W$ as follows: for every ordinal $\alpha$ let $w_\alpha = s(W - \{w_\beta | \beta < \alpha\})$. [So $w_0 = s(W)$.] Now let $t$ be the following lexicographic selector. For every nonempty proposition $P$, $t(P) = w_\alpha$ where $\alpha$ is the least ordinal $\gamma$ such that $w_\gamma \in P$.

It suffices to show that for every nonempty proposition $P$, $s(P) = t(P)$. Let $t(P) = w_\alpha$. By our construction: $P \subseteq (W - \{w_\beta | \beta < \alpha\})$. By definition, $s(W - \{w_\beta | \beta < \alpha\}) = w_\alpha$. But $w_\alpha \in P$, so by the regularity of $s$, $s(P) = w_\alpha$. □

### 4.4 Proximity

Intuitively, a selection function applied to a world $w$ and nonempty proposition $A$ should pick a member $w_1$ of $A$ that is "near" or "similar" to $w$. One way to articulate this idea is to require that the two worlds differ minimally in the sets of propositional variables that each makes true. The following notation helps us formulate this idea. For $w \in \mathbb{W}$, let $t^{-1}(w) = \{p \in \mathbb{P} \mid w \in t(p)\}$. That is, $t^{-1}(w)$ is the set of propositional variables that $\mathcal{M}$ satisfies at $w$. For sets $S, T$, let $S \triangle T$ denote their symmetric difference $(S \setminus T) \cup (T \setminus S)$. Then the idea of selecting "nearby worlds" can be rendered as follows.

(31) Definition: $\mathcal{M}$ is *proximal* just in case the following condition is met, for all $w \in \mathbb{W}$ and all nonempty propositions $A \subseteq \mathbb{W}$.

If $s(w, A) = w_1$ then there is no $w_2 \in A$ such that $t^{-1}(w) \triangle t^{-1}(w_2) \subset t^{-1}(w) \triangle t^{-1}(w_1)$.

For example, suppose that $t^{-1}(w) = \{p, q\}$, $t^{-1}(w_1) = \{p, r\}$, and $t^{-1}(w_2) = \{p, q, r\}$. Let $A = \{w_1, w_2\}$. Then $s$ violates proximity if $s(w, A) = w_1$ since $t^{-1}(w) \triangle t^{-1}(w_2) = \{r\} \subset \{q, r\} = t^{-1}(w) \triangle t^{-1}(w_1)$.

In conjunction with regularity, proximity validates a formula reminiscent of the *sure thing principle* (Savage, 1954).

(32) Proposition: Let $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ be

$$(((p \wedge r) \succ_X (q \wedge r)) \wedge ((p \wedge \neg r) \succ_X (q \wedge \neg r))) \rightarrow ((p \succ_X q).$$

Then $\varphi$ is invalid in the class of regular and in the class of proximal models but valid in the class of models that are both regular and proximal.

An instance of $\varphi$ is the following. If one has better reason to vacation in Paris during a transport strike than to vacation in Rome during a transport strike, and if one has better reason to vacation in Paris with no transport strike than to vacation in Rome with no transport strike then one has better reason to vacation in Paris than in Rome.

PROOF OF PROPOSITION (32): Construction of the needed counter models is left for the reader. Suppose that $\mathcal{M}$ is regular and proximal with $w \in \mathbb{W}$. Either $w \in t(r)$ or $w \notin t(r)$; assume the former (the argument is parallel in the other case). There is nothing left to prove unless the following statements are true [since otherwise the left conjunct in the antecedent of $\varphi$ is false; see (5)d].

(33) (a) $t(p) \cap t(r) \neq \emptyset$
(b) $t(q) \cap t(r) \neq \emptyset$

By (33), $p[\mathcal{M}] \neq \emptyset$ and $q[\mathcal{M}] \neq \emptyset$. So let $w_1, w_2 \in \mathbb{W}$ be such that:

(34) (a) $w_1 = s(w, p[\mathcal{M}])$
(b) $w_2 = s(w, q[\mathcal{M}])$

Since $w \in t(r)$, (33)a, (34)a, and proximity imply $w_1 \in t(p) \cap t(r)$. Hence, $w_1 \in (p \wedge r)[\mathcal{M}] \subseteq p[\mathcal{M}]$, so regularity implies $w_1 = s(w, (p \wedge r)[\mathcal{M}])$. Likewise, $w_2 = s(w, (q \wedge r)[\mathcal{M}])$. From $(p \wedge r) \succ_X (q \wedge r)$ we infer $u_X(w_1) > u_X(w_2)$ which in view of (34) implies $p \succ_X q$. Thus $w \in \varphi[\mathcal{M}]$. □

Similar reasoning suffices to prove:

(35) PROPOSITION: Let $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ be

$$(p \wedge ((p \wedge q) \succ_X r)) \rightarrow (q \succ_X r).$$

Then $\varphi$ is invalid in the class of regular and in the class of proximal models but valid in the class of models that are both regular and proximal.

For an instance of this formula, suppose that you have a greater gustatory interest in ham and eggs than oatmeal. Then if you already have ham, you'll be more interested in eggs than oatmeal.

### 4.5 Extensionality, saturation, and perfection

We next consider the relation between worlds and the propositional variables they satisfy. The following condition requires that distinct worlds don't make the same variables true.

(36) Definition: $\mathcal{M}$ is *extensional* just in case for all $w_1, w_2 \in \mathbb{W}$, $\{v \in \mathbb{P} \mid w_1 \in t(v)\} = \{v \in \mathbb{P} \mid w_2 \in t(v)\}$ implies $w_1 = w_2$.

Observe that every proximal, extensional model is reflexive. If every subset of variables inhabits some world, the model may be called "saturated."

(37) Definition: $\mathcal{M}$ is *saturated* just in case for all $T \subseteq \mathbb{P}$ there is $w \in \mathbb{W}$ with $\{v \in \mathbb{P} \mid w \in t(v)\} = T$.

(38) Definition: $\mathcal{M}$ is *perfect* just in case $\mathcal{M}$ is both extensional and saturated.

In a perfect model, $\mathbb{W}$ can be identified with the power set of $\mathbb{P}$. The combination of perfection and proximity has consequences for the "contraposition" of reasons, as in $(p \succ_X q) \rightarrow (\neg q \succ_X \neg p)$. This formula is plausible at first sight; for it seems that if $p$ is $u_X$-superior to $q$ then $u_X$ also favors $q$ rather than $p$ *failing* to hold. Thus, keeping a promise is morally superior to teasing the infirm hence not teasing the infirm should be morally superior to not keeping a promise, which it is. Closer inspection, however, reveals that only a weaker form of contraposition can be maintained.

(39) Proposition: Let $C$ be the class of perfect and proximal models. Then $(p \succ_X q) \rightarrow (\neg q \succ_X \neg p)$ is not valid in $C$. However, $((\neg p \wedge \neg q) \rightarrow (p \succ_X q))$ is valid in a given model of $C$ iff $((p \wedge q) \rightarrow (\neg q \succ_X \neg p))$ is valid in the same model.

Proof: We demonstrate the left-to-right direction in the second part of the proposition. Let $\mathcal{M} \in C$ be given, and suppose that:

(40) $(\neg p \wedge \neg q) \rightarrow (p \succ_X q)$ is valid in $\mathcal{M}$.

By saturation, let $w \in t(p) \cap t(q)$. By saturation again, there are $w_1, w_2 \in \mathbb{W}$ with:

(41) (a) $w_1 = s(w, \neg q[\mathcal{M}])$, and
(b) $w_2 = s(w, \neg p[\mathcal{M}])$.

To complete the proof it suffices to show that:

(42) $u_X(w_1) > u_X(w_2)$.

By (41), proximity, and perfection:

(43) (a) $w_1$ satisfies the same variables as $w$, except for $q$.
(b) $w_2$ satisfies the same variables as $w$, except for $p$.

By perfection, there is $w^* \in \mathbb{W}$ that satisfies the same subset of $\mathbb{P}$ as $w$ except for $p, q$. That is, $w^*$ falsifies $p$ and $q$ but otherwise agrees with $w$. Hence by (40), $w^* \in (p \succ_X q)[\mathcal{M}]$. So there are $w_1', w_2' \in \mathbb{W}$ with:

(44) (a) $w_1' = s(w^*, p[\mathcal{M}])$,
(b) $w_2' = s(w^*, q[\mathcal{M}])$, and
(c) $u_X(w_1') > u_X(w_2')$.

By (44)ab, proximity, and perfection:

(45) (a) $w_1'$ satisfies the same variables as $w^*$, except for $p$.
(b) $w_2'$ satisfies the same variables as $w^*$, except for $q$.

From (43), (45), and perfection, $w_1 = w_1'$ and $w_2 = w_2'$. Therefore, (42) follows from (44)c. □

### 4.6 Conditions on the utility function

We now consider different ways that utilities can be combined. This topic is at the heart of the relation between reasons and preference. For as noted earlier, we conceive preference for $\varphi$ over $\psi$ to be represented by $\varphi \succ_{\bigcup \mathbb{S}} \psi$, that is, taking account of all reasons in play. (Here it is assumed that $\bigcup \mathbb{S} \in \mathbb{S}$.) We start with the most basic condition on utility-aggregation, namely, that $u_X$ depends on just the $u_i$ indexed by $X$.

(46) DEFINITION: Let finite $X \in \mathbb{S}$ be given. $\mathcal{M}$ is *local for $X$* just in case:

(a) for all $i \in X$, $\{i\} \in \mathbb{S}$, and
(b) there is a function $g$ from finite subsets of $\Re$ to $\Re$ such that for all $w \in \mathbb{W}$, $u_X(w) = g(\{u_i(w) \mid i \in X\})$.

In this case, we call $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ *g-valid* if $\varphi$ is true in the class of models for which $u_X$ is computed via $g$.

For example, locality prevents $u_{\{1,2\}}(w)$ from depending on $u_3(w)$. It is easy to see that the following formula is valid in the class of $\{1,2\}$-local models but false in some non$\{1,2\}$-local model.

$$((p \approx_1 p') \wedge (q \approx_1 q') \wedge (p \approx_2 p') \wedge (q \approx_2 q')) \rightarrow ((p \approx_{\{1,2\}} q) \leftrightarrow (p' \approx_{\{1,2\}} q'))$$

Candidates for $g$ in Definition (46) include:

$$u_X(w) = \begin{array}{|l|l|} \hline \text{average}\{u_i(w) \mid i \in X\} & \text{median}\{u_i(w) \mid i \in X\} \\ \hline \text{minimum}\{u_i(w) \mid i \in X\} & \text{maximum}\{u_i(w) \mid i \in X\} \\ \hline \end{array}$$

Formulas separate some of these locality classes. For example, the following schema is average-valid but neither min- nor max-valid with respect to $\{i, j\}$.

$$((\varphi \succ_i \psi) \wedge (\varphi \approx_j \psi)) \rightarrow (\varphi \succ_{\{i,j\}} \psi)$$

To see that the schema is not min-valid, take $u_j$ to assign identical numbers to all worlds, much smaller than the numbers that $u_i$ assigns. Do the reverse for a counter-model to max-validity.

Next is a schema that is min-valid and max-valid but not average-valid.

$$(\varphi \approx_{\{i,j,k\}} \psi) \rightarrow ((\varphi \approx_{\{i,j\}} \psi) \vee (\varphi \approx_{\{i,k\}} \psi) \vee (\varphi \approx_{\{j,k\}} \psi))$$

For a counter-model to the formula with respect to averaging, let $w_1, w_2$ be the worlds attained through $\varphi, \psi$, respectively, and let the $i, j, k$ utilities be given in the accompanying table.

| | $w_1$ | $w_2$ |
|---|---|---|
| $i$ | 2 | 0 |
| $j$ | 2 | 3 |
| $k$ | 2 | 3 |

Observe that utility aggregation has so far been monotonic in the following sense.

(47) Definition: Let $X = \{x_1 \dots x_n\} \in \mathbb{S}$ be given, where also $\{x_1\}, \dots, \{x_n\} \in \mathbb{S}$. A model $\mathcal{M} = (\mathbb{W}, s, u, t)$ is *monotone for X* just in case for all $\varphi, \psi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$,

$$((\varphi \succ_{x_1} \psi) \wedge \cdots \wedge (\varphi \succ_{x_n} \psi)) \rightarrow (\varphi \succ_X \psi)$$

is valid in $\mathcal{M}$.

The four functions discussed above are consistent with monotonicity but it is easy to imagine circumstances in which non-monotonic aggregation takes place. For example, you might prefer to spend time with people of luxuriant life style (they're more fun), encoded in $u_1$, and also prefer people who espouse asceticism and self-restraint (they're more admirable), encoded in $u_2$. The two utility functions considered individually might order Jim above Jack as dinner partners but $u_{1,2}$ will reverse the preference if it is sensitive to Jim's hypocrisy.

Finally, we formalize one kind of independence between propositional variables (analogous to concepts evoked in the theory of conjoint measurement, Krantz et al., 1971, §6.1.4). This condition figures in the undecidability theorem presented in Section 5.

Let model $\mathcal{M} = (\mathbb{W}, s, u, t)$ for signature $(\mathbb{P}, \mathbb{S})$ be given, with $w_1, w_2 \in \mathbb{W}$. For $z \in \mathbb{P}$, $w_1$ and $w_2$ are said to be *z-variants* just in case $t$ makes the same variables true in $w_1$ and $w_2$, except for $z$, which is true in just one of $w_1, w_2$. For example, suppose that $\mathbb{P} = \{p, q, r\}$. If both $w_1, w_2$ fall into $t(p)$ but not $t(r)$, and $w_1$ but not $w_2$ falls in $t(q)$, then $w_1$ and $w_2$ are $q$-variants. According to the following definition, selection of a $z$-variant from a given world changes utility by a fixed amount.

(48) DEFINITION: Let model $\mathcal{M} = (\mathbb{W}, s, u, t)$ be given with $z \in \mathbb{P}$. Then $\mathcal{M}$ satisfies *z-independence* just in case for all $X \in \mathbb{S}$ there is $\alpha_X \in \Re$ such that for all $w \in \mathbb{W}$ and $\emptyset \neq A \subseteq \mathbb{W}$, if $w$ and $s(w, A)$ are $z$-variants then $\mid u_X(w) - u_X(s(w, A)) \mid = \alpha_X$. We call $\mathcal{M}$ *weakly independent* iff $\mathcal{M}$ satisfies $p$-independence and $q$-independence for distinct $p, q \in \mathbb{P}$.

The following fact (easily demonstrated) illustrates the impact of independence.

(49) PROPOSITION: Let $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ be

$$(p \wedge q) \rightarrow ((\top \approx_X \neg p) \leftrightarrow (\top \approx_X (\neg p \wedge q)).$$

Then $\varphi$ is valid in the class of models that are reflexive, proximal, saturated and $p$-independent, but not valid in the class of models that are reflexive, proximal and saturated.

For an instance of $\varphi$ in (49), suppose that the *Atlanta Thrashers* won last night's hockey match ($p$) and there is milk in your coffee ($q$). Then $\varphi$ says that you are indifferent about the *Thrashers* victory if and only if you are indifferent about the *Thrashers* victory with milk (still) in your coffee. (For the illustration, we assume that $X = \bigcup \mathbb{S}$.)

### 4.7 Discernibility and Stalnaker conditionals

The present section considers models in which each world is uniquely valuable.

(50) DEFINITION: Model $\mathcal{M} = (\mathbb{W}, s, u, t)$ is *discernible* just in case there are no two worlds $v, w \in \mathbb{W}$ such that for every $X \in \mathbb{S}$, $u_X(v) = u_X(w)$.

That is, distinct worlds in a discernible model don't agree on all utility scales. The next proposition is immediate.

(51) PROPOSITION: Let signature $(\mathbb{P}, \{X, Y\})$ be given. Then the invalid schema

$$\neg(\theta \approx_X \neg\theta) \vee \neg(\theta \approx_Y \neg\theta)$$

is valid in the class of discernible models.

Discernibility is connected to the following question. Can $\mathcal{L}$ express the idea that $\psi$ is true in the world that comes to mind when envisioning $\varphi$? If so then our logic can represent the conditional "if $\varphi$ then $\psi$" in something like the sense introduced by Stalnaker (1968). We investigate the matter via the following definition.

(52) DEFINITION: Let model $\mathcal{M} = (\mathbb{W}, s, u, t)$ and $\varphi, \psi \in \mathcal{L}$ be given. We write $(\varphi \rhd \psi)[\mathcal{M}]$ for the set of $w \in \mathbb{W}$ such that $s(w, \varphi[\mathcal{M}]) \in \psi[\mathcal{M}]$. (If $\varphi[\mathcal{M}]$ is empty then so is $(\varphi \rhd \psi)[\mathcal{M}]$.)

Thus, $(\varphi \rhd \psi)$ is true at $w \in \mathbb{W}$ if the world selected at $w$ from $\varphi[\mathcal{M}]$ satisfies $\psi$. The next proposition shows that some instances of $(\varphi \rhd \psi)$ are inexpressible in $\mathcal{L}$.

(53) PROPOSITION: Let signature $(\mathbb{P}, \mathbb{S}) = (\{p\}, \{\{i\}\})$ be given. Then there is no $\theta \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ such that $\theta[\mathcal{M}] = (\top \rhd p)[\mathcal{M}]$ for all models $\mathcal{M}$ of signature $(\mathbb{P}, \mathbb{S})$.

PROOF: It suffices to exhibit models $\mathcal{M}_1$, $\mathcal{M}_2$ (of the foregoing signature) such that

(54) (a) $\chi[\mathcal{M}_1] = \chi[\mathcal{M}_2]$ for all $\chi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$, and
(b) $(\top \rhd p)[\mathcal{M}_1] \neq (\top \rhd p)[\mathcal{M}_2]$.

Let $\mathcal{M}_1 = (\mathbb{W}, s_1, u, t)$ and $\mathcal{M}_2 = (\mathbb{W}, s_2, u, t)$, with

$$\begin{aligned}
&\mathbb{W} = \{w_1, w_2\} \\
&u_i(w_1) = u_i(w_2) \\
&t(p) = \{w_1\} \\
&s_1(w_1, \{w_1, w_2\}) = s_1(w_2, \{w_1, w_2\}) = w_1, \text{ and} \\
&s_2(w_1, \{w_1, w_2\}) = s_2(w_2, \{w_1, w_2\}) = w_2.
\end{aligned}$$

That is, $\mathcal{M}_1$, $\mathcal{M}_2$ differ only in selection function. [Of course, $s(w, \{w\}) = w$ for any selection function $s$.] A simple induction on the complexity of formulas verifies (54)a, relying on $u_i(w_1) = u_i(w_2)$ for the modal connective. For (54)b, it is easy to verify that $(\top \rhd p)[\mathcal{M}_1] = \{w_1, w_2\}$ whereas $(\top \rhd p)[\mathcal{M}_2] = \emptyset$. □

The models evoked in the foregoing proof are regular, hence even assuming regularity does not allow $(\top \rhd p)$ to be represented by a formula. The models are not discernible, however. Adding this property does the trick, at least in the case of finitely many utility indexes.

(55) Proposition: Let signature $(\mathbb{P}, \mathbb{S})$ with $\mathbb{S}$ finite be given. Let $C$ be the class of regular, discernible models. Then for all $\mathcal{M} \in C$ and $\varphi, \psi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$, $(\varphi \rhd \psi)[\mathcal{M}] = \chi[\mathcal{M}]$ where $\chi$ is the conjunction over indices $X \in \mathbb{S}$ of $(\varphi \approx_X (\varphi \wedge \psi))$.

Proof: Left to right, $w \in (\varphi \rhd \psi)[\mathcal{M}] \Rightarrow s(w, \varphi[\mathcal{M}]) \in \psi[\mathcal{M}] \Rightarrow$ (by regularity) $s(w, \varphi[\mathcal{M}]) = s(w, (\varphi \wedge \psi)[\mathcal{M}]) \Rightarrow w \in (\varphi \approx_X (\varphi \wedge \psi))[\mathcal{M}]$ for all $X \in \mathbb{S}$. Right to left, $w \in (\varphi \approx_X (\varphi \wedge \psi))[\mathcal{M}]$ for all $X \in \mathbb{S} \Rightarrow$ (by discernibility) $s(w, \varphi[\mathcal{M}]) = s(w, (\varphi \wedge \psi)[\mathcal{M}]) \Rightarrow s(w, \varphi[\mathcal{M}]) \in \psi[\mathcal{M}] \Rightarrow w \in (\varphi \rhd \psi)[\mathcal{M}]$. □

A simple modification of the proof of Proposition (53) reveals that the finiteness assumption in Proposition (55) is essential.

## 5 Decidability and Compactness

The present section offers four theorems about the compactness and decidability of satisfiability (hence, about the decidability of validity as well). For this purpose, we fix a signature $(\mathbb{P}, \mathbb{S})$ in which $\mathbb{P}$ is an initial segment of $\mathbb{N}$, and $\mathbb{S}$ is a set of finite subsets of $\mathbb{N}$. The first theorem concerns satisfiability with respect to the class of all models.

(56) Theorem: The set of satisfiable formulas of $\mathcal{L}(\mathbb{P}, \mathbb{S})$ is decidable.

Adjustments to the proof of Theorem (56) verify the following corollaries.

(57) Corollary: If a formula of $\mathcal{L}(\mathbb{P}, \mathbb{S})$ is satisfiable then it is satisfied in a finite model (that is, in a model with finitely many worlds).

(58) Corollary: The set of formulas of $\mathcal{L}(\mathbb{P}, \mathbb{S})$ that are satisfiable in the class of reflexive models is decidable.

Corollary (57) may be contrasted with Proposition (11), stating that some formulas can be satisfied by a reflexive model only if the model contains infinitely many worlds.

The second theorem affirms *un*decidability in the class of reflexive, discernible models that satisfy weak independence. Reflexivity, discernibility, and weak independence were introduced in Definitions (9), (50), and (48).

(59) Theorem: There is no decision procedure for the set of formulas of $\mathcal{L}(\mathbb{P}, \mathbb{S})$ that are satisfiable in the class of models that are reflexive, discernible, and weakly independent. In fact, this set of formulas is co-r.e.-complete.

The third and last theorem bears on lexicographic ordering in the sense of Definition (26), and on proposition drivenness in the sense of Definition (28).

(60) Theorem: The set of formulas of $\mathcal{L}(\mathbb{P}, \mathbb{S})$ that are satisfiable in the class of lexicographic models is decidable, as is the set of formulas that are satisfiable in the class of proposition driven models. Indeed, both sets of formulas are **NP**-complete.

The final theorem affirms that satisfiability with respect to the class of all models is countably compact. We call a collection $\Sigma \subseteq \mathcal{L}$ of formulas "satisfiable" just in case there is a model $\mathcal{M}$ that satisfies every member of $\Sigma$ at a common point, that is, just in case:

$$\bigcap\{\varphi[\mathcal{M}] \mid \varphi \in \Sigma\} \neq \emptyset.$$

(61) Theorem: Suppose that signature $(\mathbb{P}, \mathbb{S})$ is countable, and let $\Sigma \subseteq \mathcal{L}(\mathbb{P}, \mathbb{S})$ be given. Then $\Sigma$ is satisfiable if and only if every finite subset of $\Sigma$ is satisfiable.

On the other hand, if either $\mathbb{P}$ or $\mathbb{S}$ is uncountable then compactness breaks down. Proofs of the theorems are provided in the Appendixes.

## 6 Generalized frames and models

In our theory, $\varphi \succeq_X \psi$ can be understood as asserting that $u_X$ assigns 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, picked out as a function of the world at which the formula is evaluated. A natural generalization is to compare the value of propositions directly, without recourse to individual worlds as representatives. We explore this idea in the present section, proving Theorem (24) along the way. Let $(\mathbb{P}, \mathbb{S})$ be our background signature, and recall that a *total preorder* is transitive, connected, and reflexive over its domain.

(62) 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 $\mathbb{W}$. We call the pair $(\mathbb{W}, v)$ a *generalized frame.*

(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.*

Intuitively, a value-ordering arranges propositions by utility, relative to index $X \in \mathbb{S}$ and vantage point $w \in \mathbb{W}$. An easy adaptation of the proof of Theorem (61) shows that satisfaction in the class of generalized models is compact for arbitrary signatures.

Let $(\mathbb{W}, s, u)$ be a frame in the sense of Definition (23); 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 before $B$ iff $u_X(w_A) \geq u_X(w_B)$ where $w_A = s(w, A)$ and $w_B = s(w, B)$. We call $(\mathbb{W}, v)$ *the generalized frame induced by* $(\mathbb{W}, s, u)$.

Given a model $(\mathbb{W}, s, u, t)$, $w \in \mathbb{W}$, and nonempty $A \subseteq \mathbb{W}$, there is $w_0 \in \mathbb{W}$ with $u_X(s(w, A)) = u_X(s(w, \{w_0\}))$, namely, $w_0 = s(w, A)$. So we have:

(63) Lemma: Let value-ordering $v$ be induced by frame $(\mathbb{W}, s, u)$. Then for all $w \in \mathbb{W}$ and $X \in \mathbb{S}$, every equivalence class in $v(w, X)$ contains a singleton set.

The following proposition follows immediately.

(64) Proposition: Let $\mathbb{W}$ contain at least two worlds. Let value-ordering $v$ be such that for some $w \in \mathbb{W}$ and $X \in \mathbb{S}$, either

(a) $v(w, X)$ refines $\subset$ over the field of nonempty subsets of $\mathbb{W}$, or

(b) $v(w, X)$ is a strict linear order over the nonempty subsets of $\mathbb{W}$.

Then $(\mathbb{W}, v)$ is not induced by any frame $(\mathbb{W}, s, u)$.

The semantics of generalized models is given by Definition (5) with the following substitution. 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.

(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:

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

(65) Proposition: Let $X \in \mathbb{S}$ be given. There are $\varphi_1, \varphi_2 \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ such that

(a) $\varphi_1$ is valid in a generalized frame $(\mathbb{W}, v)$ if and only if for all $w \in \mathbb{W}$, $v(w, X)$ refines $\subset$ over the nonempty subsets of $\mathbb{W}$;

(b) $\varphi_2$ is valid in a generalized frame $(\mathbb{W}, v)$ if and only if for all $w \in \mathbb{W}$, $v(w, X)$ is a strict linear order over the nonempty subsets of $\mathbb{W}$.

Proof: It is easy to verify the proposition with the following choices of $\varphi_1, \varphi_2 \in \mathcal{L}(\mathbb{P}, \mathbb{S})$, respectively.

$$(\Box(p \rightarrow q) \wedge \neg\Box(q \rightarrow p) \wedge \Diamond p) \rightarrow p \prec_X q$$

$$\neg\Box(p \leftrightarrow q) \rightarrow ((p \prec_X q) \vee (q \prec_X p))$$

$\Box$

Propositions (64) and (65) exhibit characterizable classes of generalized frames none of which can be induced by any frame. The following lemma presents a fundamental connection between a frame and the generalized frame it induces. It follows from a simple induction on the complexity of formulas.

(66) Lemma: Let $\mathcal{F}_1 = (\mathbb{W}, s_1, u_1)$ and $\mathcal{F}_2 = (\mathbb{W}, s_2, u_2)$ be frames, and suppose that the generalized frame induced by $\mathcal{F}_1$ is identical to the generalized frame induced by $\mathcal{F}_2$. Then for all $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ and all truth assignments $t$, $\varphi[\,(\mathbb{W}, s_1, u_1, t)\,] = \varphi[\,(\mathbb{W}, s_2, u_2, t)\,]$.

Theorem (24) is a corollary to Lemma (66).

Proof of Theorem (24): It follows immediately from Lemma (66) that if $u(w, X)$ is a constant function for all $X \in \mathbb{S}$ and $w \in \mathbb{W}$, then for all selection functions $s_1$ and $s_2$, all truth assignments $t$, and all $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$, $\varphi[\,(\mathbb{W}, s_1, u, t)\,] = \varphi[\,(\mathbb{W}, s_2, u, t)\,]$. Theorem (24) now follows by choosing $s_1$ and $s_2$ so that the frame $(\mathbb{W}, s_1, u)$ is regular and the frame $(\mathbb{W}, s_2, u)$ is not. $\square$

## 7 Discussion

The foregoing investigation raises many questions and avenues for further research. We indicate some directions.

### 7.1 Utility

Suppose that distinct $\{i\}, \{j\}, \{k\} \in \mathbb{S}$. For all $\varphi, \theta \in \mathcal{L}(\mathbb{P}, \mathbb{S})$, let:

$$(\varphi \,\mathtt{V}\, \theta) \stackrel{\text{def}}{=} ((\varphi \succ_i \theta) \wedge (\varphi \succ_j \theta)) \vee ((\varphi \succ_i \theta) \wedge (\varphi \succ_k \theta)) \vee ((\varphi \succ_j \theta) \wedge (\varphi \succ_k \theta))$$

Then $\varphi \,\mathtt{V}\, \theta$ is true if a majority of the utility scales $i, j, k$ are favorable to $\varphi$ compared to $\theta$. Observe that $((\varphi \,\mathtt{V}\, \theta) \wedge (\theta \,\mathtt{V}\, \psi)) \rightarrow (\varphi \,\mathtt{V}\, \psi)$ (transitivity) is not guaranteed in a given model inasmuch as the utility scales $u_i, u_j, u_k$ might embody a voting cycle (see Johnson, 1998). Therefore, $\mathtt{V}$ cannot itself be represented by a utility scale. The following matter thus merits exploration.

(67) Open Question: Suppose that $\{i, j, k\} \in \mathbb{S}$. Under what conditions does $\varphi \,\mathtt{V}\, \theta$ imply $\varphi \succ_{i,j,k} \theta$, and vice versa?

The voting operator $\mathtt{V}$ might best be analyzed in the context of a generalization of our approach to utility. Instead of utility scales corresponding to each $X \in \mathbb{S}$, we may posit relations $R_X \subseteq \mathbb{W} \times \mathbb{W}$. In this set up, $\theta \succeq_X \psi$ is true at $w \in \mathbb{W}$ just in case $(s(w, \theta[\,\mathcal{M}\,]), s(w, \psi[\,\mathcal{M}\,])) \in R_X$. Such relations $R_X$ could vary in their order-theoretic properties (e.g., transitivity) as well as in their connection to relations $R_i$ with $i \in X$. This perspective might allow the remarkable results developed in Andréka et al. (2002), about combining preference relations, to shed light on the logic of reasons.

Questions also remain about the classes of utility functions defined in Section 4.6. Can any of them be uniquely characterized by a set of formulas? Even the less ambitious problem of separating utility functions is currently unresolved. For example, the following question was left open.

(68) Open Question: Is there $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ that is minimum-valid but not maximum-valid (and vice versa)?

### 7.2 Selection

Additional conditions on selection functions remain to be investigated. Here is an example.

(69) Definition: A selection function $s$ over $\mathbb{W}$ is *metrizable* 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 $w$.

A weaker version of the same idea would allow $s(w, A)$ to be among the $d$-closest members of $A$ to $w$. In both cases, $s$ is metrizable only if such $d$-closest worlds exist (no chains of worlds ever $d$-closer to $w$).

(70) Open Question: Is there an invalid formula (or, invalid in an interesting class of models) that is valid in the class of models with metrizable selection functions?

Observe that metrizability opens the door to a family of natural conditions linking selection to utility. For example, *continuity* might be imposed, requiring small changes in utility for small steps via selection. [Another kind of link between selection and utility is described in Definition (48), above.]

A different path to a richer concept of selection allows more than one world to be "nearest" to a target, without evoking distance explicitly. To embody this idea, we replace Definition (1) with the following.

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

Selection functions in the original sense of Definition (1) can now be seen as the special case in which only singleton sets are returned. To satisfy a formula $\theta \succeq_X \psi$ in the context of a wide selection function, we may require that some nearby $\theta$-world is weakly $X$-better than some nearby $\psi$-world, or that all of them are, etc. The consequences of these options have yet to be explored.

### 7.3 Intensionality

It is well known that reasons for choices are often constructed "on the fly," as a function of the character and context of the question posed. For example, pricing a lottery draws attention to its dollar payoff at the expense of probability (Tversky et al., 1988). As a result, one lottery may be preferred to another in binary choice yet sold at a lower price (Grether and Plott, 1979; for other examples, see Shafir et al., 1993 and the anthology Lichtenstein and Slovic, 2006). We here consider a logical counterpart to contextual influence; specifically, we allow a selection function to be sensitive to the formula that calls for its use. For this purpose, we modify Definition (4) from Section 3 as follows.

(72) DEFINITION: An *intensional model* $\mathcal{M}$ for a signature $(\mathbb{P}, \mathbb{S})$ is just like a model in the original sense except that its selection function is a (potentially partial) function from $\mathbb{W} \times \mathcal{L}(\mathbb{P}, \mathbb{S})$ to $\mathbb{W}$.

As a special case, $s(w, \varphi)$ might be computed as $s(w, \varphi[\mathcal{M}])$ in the familiar way, hence models in the original (extensional) sense are a special kind of intensional model. Properly intensional models make $\varphi$ visible to $s$, thus allowing $\varphi$ to influence the choice from $\mathbb{W}$. To reconstruct the rest of our original semantics in the intensional setting, it suffices to replace the last clause of Definition (5) with:

(d) If $\varphi$ has the form $(\theta \succeq_X \psi)$ for $X \in \mathbb{S}$, then $\varphi[\mathcal{M}]$ equals the set of $w \in \mathbb{W}$ such that $s(w, \theta)$ and $s(w, \psi)$ are both defined, and $u_X(s(w, \theta)) \geq u_X(s(w, \psi))$.

Intensional models are exposed to the usual perils. For example, unless further conditions are imposed, nothing prevents different selections in the context of $(p \wedge q)$ compared to $(q \wedge p)$. At the limit, the following properties ensure that a given intensional model $\mathcal{M}$ is elementarily equivalent to some extensional one.

(a) $s(w, \varphi)$ is undefined if $\varphi[\mathcal{M}] = \emptyset$; otherwise:

(b) $s(w, \varphi) \in \varphi[\mathcal{M}]$, and

(c) $s(w, \varphi) = s(w, \psi)$ if $\varphi[\mathcal{M}] = \psi[\mathcal{M}]$.

But softer conditions seem required to model human reasoners. Finding elegant but descriptively revealing ways to manage intensional selection strikes us as a major theoretical challenge.

### 7.4 Analysis of model classes

For every class $C$ of models (e.g., the regular models), the following questions arise.

(a) Is the set of formulas satisfiable in $C$ decidable? (Hence, is the set of formulas valid in $C$ decidable?)

(b) If the answer to (a) is affirmative, what is the computational complexity of the set of formulas satisfiable in $C$?

(c) Is there $\Gamma \subset \mathcal{L}(\mathbb{P}, \mathbb{S})$ such that $\mathcal{M} \in C$ iff $\Gamma$ is valid in $\mathcal{M}$?

(d) Can the set of formulas that are valid in $C$ be finitely axiomatized with respect to a natural set of inference rules?

The facts reported in Section 5 barely begin to respond to Queries (a) and (b) while leaving (c) and (d) untouched. The same queries arise for extensions of our language, for example, to formulas $(\varphi \rhd \psi)$ interpreted as Stalnaker conditionals via Definition (52) [Section 4.7].

### 7.5 Updating

Suppose you live in a model $\mathcal{M} = (\mathbb{W}, s, u, t)$ but wish to take on board $\varphi \in \mathcal{L}$ as an assumption. We take this to mean that $\varphi$ will be made true in all worlds of some successor model $\mathcal{M}' = (\mathbb{W}', s', u', t')$ that is the natural $\varphi$-*update* to $\mathcal{M}$. (Updating is analyzed from a graph-theoretic perspective in Andréka et al., 2002; van Benthem and Liu, 2007.)

If $\varphi$ is boolean, updating $\mathcal{M}$ seems easy: set $\mathbb{W}' = \{w \in \mathbb{W} \mid w \models \varphi\}$, and let $s'$, $u'$, $t'$ be the obvious reducts of $s$, $u$, and $t$ to $\mathbb{W}'$. (Updating in this sense is not defined if $\mathbb{W}' = \emptyset$.) But if $\varphi$ has a modal connective, matters are not straightforward. Consider the following choice for $\mathcal{M}$, where $(\mathbb{P}, \mathbb{S}) = (\{p, q\}, \{\{i\}\})$.

$$\mathbb{W} = \{w_1, w_2, w_3, w_4\}$$

$$t(p) = \{w_2, w_4\} \quad t(q) = \{w_3\}$$

$$u_i(w_4) < u_i(w_3) < u_i(w_2) < u_i(w_1)$$

$$s(w_1, p[\mathcal{M}]) = w_2$$
$$s(w_1, q[\mathcal{M}]) = w_3$$

$$s(w_2, p[\mathcal{M}]) = s(w_3, p[\mathcal{M}]) = s(w_4, p[\mathcal{M}]) = w_4$$
$$s(w_2, q[\mathcal{M}]) = s(w_3, q[\mathcal{M}]) = s(w_4, q[\mathcal{M}]) = w_3$$

For $\varphi := p \prec_i q$ to be true throughout $\mathcal{M}'$, it suffices to remove $w_1$ from $\mathbb{W}$. But since $s(w_1, \{w_4\})$ must equal $w_4$, it is easy to verify that removing $w_2$ from $\mathbb{W}$ also suffices for the same purpose. Updating in the general case thus requires choice among successor models, in a sense familiar from the theory of belief revision (Gärdenfors, 1988). Investigation of the matter might usefully address the following issue. Given a proposed updating operator $\ddagger$ and a class $C$ of models with (say) the regularity property, for which $\varphi \in \mathcal{L}$ (if any) is $\{\mathcal{M} \ddagger \varphi \mid \mathcal{M} \in C\}$ guaranteed to be regular?

## Appendix: Proof of Theorem (56)

To demonstrate that the set of satisfiable formulas of $\mathcal{L}(\mathbb{P},\mathbb{S})$ is decidable, we apply the well-known "method of mosaics" (see Blackburn et al. 2001, §6.4). We carry out the construction in some detail. Proofs of the other two theorems will be more summary in character.

Let $\theta \in \mathcal{L}(\mathbb{P},\mathbb{S})$ be given. Let $\Sigma$ be the collection of subformulas of $\theta$, and let $Z$ be the set of utility indices that appear in $\theta$. We close $\Sigma$ under one application of negation, followed by one application of $\leftrightarrow$, followed by one application of negation, followed by one application of $\Diamond$, followed by one application of negation. The resulting set of formulas will be called $\Omega$. We say that $\Delta \subseteq \Omega$ is a *Hintikka set* (abbreviated *H-set*) if and only if

(73) (a) for every $\neg\varphi \in \Omega$, $\varphi \in \Delta$ iff $\neg\varphi \notin \Delta$, and

(b) for every $(\varphi \wedge \psi) \in \Omega$, $(\varphi \wedge \psi) \in \Delta$ iff both $\varphi \in \Delta$ and $\psi \in \Delta$.

We let $\Xi$ be the collection of all H-sets. Note that if $n$ is the length of $\theta$, then the size $c$ of $\Omega$ (and thus of every H-set) is $O(n^2)$. Therefore, the size $d$ of $\Xi$ is $O(2^{n^2})$. For the purposes of the next definition, we establish the notational convention that if $f$ is the graph of a partial function, we write $f(a)$ for the $b$ such that $\langle a, b\rangle \in f$, when $a$ is in the domain of $f$. A *brick* is a triple $\langle \Delta, \sigma, \{\upsilon_X \mid X \in Z\}\rangle$ where

(74) (a) $\Delta$ is an H-set;

(b) $\sigma$ is the graph of a partial function from $\Sigma$ into $\Xi$ such that $\varphi \in \sigma(\varphi)$ for every $\varphi \in \Sigma$ on which $\sigma$ is defined;

(c) for each $X \in Z$, $\upsilon_X$ is a function from $\sigma$ to $\{i \mid 1 \leq i \leq \mathsf{card}(\sigma)\}$;

(d) if $\Diamond\varphi \in \Delta$, then for some $\Delta' \in \mathsf{range}(\sigma)$, $\varphi \in \Delta'$;

(e) if $\Diamond\varphi \notin \Delta$, then for all $\Delta' \in \mathsf{range}(\sigma) \cup \{\Delta\}$, $\varphi \notin \Delta'$;

(f) $\Box(\varphi \leftrightarrow \psi) \in \Delta$ if and only if $\sigma(\varphi) = \sigma(\psi)$;

(g) $(\varphi \preceq_X \psi) \in \Delta$ iff $\upsilon_X(\langle \varphi, \sigma(\varphi)\rangle) \leq \upsilon_X(\langle \psi, \sigma(\psi)\rangle)$.

Let $z$ be the size of $Z$. Note that the number $b$ of bricks is $O(d^{c+1} \cdot c^{cz})$.

If $\beta$ is a brick, we write $\beta_1, \beta_2$, and $\beta_3$ for the first, second, and third coordinates of $\beta$. A set $B$ of bricks is a *mosaic* if and only if

(75) (a) for all $\beta, \beta' \in B$, $\{\varphi \mid \Diamond\varphi \in \beta_1\} = \{\varphi \mid \Diamond\varphi \in \beta'_1\}$, and

(b) for all $\beta \in B$ and for all $\Delta \in \mathsf{range}(\beta_2)$ there is a $\beta' \in B$ such that $\beta'_1 = \Delta$.

A set $B$ of bricks is a mosaic for $\theta \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ if and only if $B$ is a mosaic and for some $\beta \in B$, $\theta \in \beta_1$. Note that the number of mosaics is $O(2^b)$ and that it is decidable in time polynomial in the size of a set $B$ of bricks whether $B$ is a mosaic. It follows that the decision problem "Does there exist a mosaic for $\theta$" is in $\mathsf{NTIME}(b)$. Theorem (56) is thus a corollary to the following.

(76) PROPOSITION: For every $\theta \in \mathcal{L}(\mathbb{P}, \mathbb{S})$, $\theta$ is satisfiable if and only if there is a mosaic for $\theta$.

To prove the left to right direction of (76), let satisfiable $\theta \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ be given. For notational convenience we assume that $Z = \{X\}$. The generalization to multiple utility indices is routine. Let model $\mathcal{M} = (\mathbb{W}, s, u, t)$ satisfy $\theta$ and suppose that $w_0 \in \theta[\mathcal{M}]$. For each $w \in \mathbb{W}$, let $\Delta_w = \{\varphi \in \Omega \mid w \in \varphi[\mathcal{M}]\}$. Note that for every $w \in \mathbb{W}$, $\Delta_w$ is an H-set. Now for each $w \in \mathbb{W}$, we construct a brick $\beta_w = \langle \Delta_w, \sigma_w, \upsilon_w \rangle$, where for each $\varphi \in \Sigma$,

$$\sigma_w(\varphi) = \Delta_{s(w, \varphi[\mathcal{M}])}$$

and for all $\varphi, \psi \in \Delta_w$

$$\upsilon_w(\langle \varphi, \sigma_w(\varphi) \rangle) \leq \upsilon_w(\langle \psi, \sigma_w(\psi) \rangle) \text{ iff } u(s(w, \varphi[\mathcal{M}])) \leq u(s(w, \psi[\mathcal{M}])).$$

Let $B = \{\beta_w \mid w \in \mathbb{W}\}$. It is easy to verify that $B$ is a mosaic for $\theta$.

For the right to left direction of (76), suppose that $B$ is a mosaic for $\theta$. We show how to use the bricks of $B$ to construct a tree-like infinite model $\mathcal{M} = (\mathbb{W}, s, u, t)$, the root world $w_0$ of which satisfies $\theta$. The tree underlying the model is both node-labelled and edge-labelled. We call the node labels "brick-labels" and we call the edge labels "selector-labels". The construction of the tree proceeds by induction; at stage $i$, we construct the worlds of depth $i$.

At stage 0 we introduce a world $w_0$ and label $w_0$ with some brick $\beta \in B$ such that $\theta \in \beta_1$ (such a brick-label exists, since $B$ is a mosaic for $\theta$).

Let $W_n$ be the set of worlds constructed at stage $n$. At stage $n+1$ we proceed as follows. For each $w \in W_n$ we construct the children of $w$ as follows. Let $\beta_w = \langle \Delta_w, \sigma_w, \upsilon_w \rangle$ be the brick-label of $w$. For each $\Delta \in \mathsf{range}(\sigma_w)$ we introduce a child $w'$ of $w$; we label the edge from $w$ to $w'$ with selector-label $\sigma_w^{-1}[\Delta](= \{\varphi \mid \sigma_w(\varphi) = \Delta\})$ and we choose as the the brick-label of $w'$ some brick $\beta \in B$ with $\beta_1 = \Delta$ (by (75)b, such a brick exists since $B$ is a mosaic). This completes the definition of $W_{n+1}$. We let $\mathbb{W} = \bigcup_n W_n$. Next we define $u$ and $t$.

(a) For every $p \in \mathbb{P}$, $t(p) = \{w \mid p \in (\beta_w)_1\}$.

(b) For every $w' \in \mathbb{W}$ $u(w') = \upsilon_w(\Delta)$, where $w$ and $\Delta$ are the unique world and Hintikka set such that $w'$ is a child of $w$ and $(\beta_{w'})_1 = \Delta$.

To complete the definition of the model $\mathcal{M}$ we must specify the selection function $s$. Recall that $\Sigma$ is the set of subformulas of $\theta$. For nonempty subsets $T$ of $\mathbb{W}$ not of the form $\varphi[\mathcal{M}]$ for some $\varphi \in \Sigma$, and for each $w$, we choose $s(w, T)$ to be an arbitrary element of $T$. For $\varphi \in \Sigma$ we wish to define $s$ so that for every $w \in \mathbb{W}$, $s(w, \varphi[\mathcal{M}])$ is the unique child $w'$ of $w$ such that $\varphi$ is an element of the selector-label of the edge from $w$ to $w'$, provided that $\varphi[\mathcal{M}]$ is nonempty. Since, in general, $\varphi[\mathcal{M}]$ depends on $s$, we will define $s$ on sets of the form $\varphi[\mathcal{M}]$ by recursion on the logical complexity of $\varphi$. Simultaneously, we will prove by induction on logical complexity that for every $w \in \mathbb{W}$ and $\varphi \in \Sigma$,

$$w \in \varphi[\mathcal{M}] \quad \text{iff} \quad \varphi \in (\beta_w)_1$$

thereby completing the proof of the theorem.

Basis: It follows immediately from the definition of $t$ that for every $p \in \Sigma$ and $w \in \mathbb{W}$:

(77) $w \in p[\mathcal{M}] \quad \text{iff} \quad p \in (\beta_w)_1$.

Now, for each $p \in \Sigma$ and $w \in \mathbb{W}$, we let $s(w, p[\mathcal{M}])$ be the unique child $w'$ of $w$ such that $p$ is an element of the selector-label of the edge from $w$ to $w'$, provided that $p[\mathcal{M}]$ is nonempty. In order to secure the legitimacy of this definition we need to verify that

(78) (a) $s(w, p[\mathcal{M}]) \in p[\mathcal{M}]$;

(b) $p[\mathcal{M}] = \emptyset$ iff no selector-label of an edge exiting $w$ contains $p$;

(c) for all $q \in \Sigma$, $p[\mathcal{M}] = q[\mathcal{M}]$ iff $p$ and $q$ are contained in exactly the same selector-labels of edges exiting $w$ (that is, either none of them, or a unique one containing both).

Condition (78)a follows immediately from (77). In order to establish (78)b, suppose first that $p[\mathcal{M}] = \emptyset$. By (77), it follows that for all $w' \in \mathbb{W}$, $p \notin (\beta_{w'})_1$. Hence, for no child $w'$ of $w$ is $p \in (\beta_{w'})_1$. Therefore, no selector-label of an edge exiting $w$ contains $p$. For the converse, suppose that no selector-label of an edge exiting $w$ contains $p$. Note that since $p \in \Sigma$, $\neg\Diamond p \in \Omega$. It follows from (74)d that $\Diamond p \notin (\beta_w)_1$, and thence from (75)a, that for every $w' \in \mathbb{W}$, $\Diamond p \notin (\beta_{w'})_1$. Hence, by (73)a, for every $w' \in \mathbb{W}$, $\neg\Diamond p \in (\beta_{w'})_1$. Therefore, by (74)e and (73)a, for every $w' \in \mathbb{W}$, $p \notin (\beta_{w'})_1$. Hence, by (77), $p[\mathcal{M}] = \emptyset$.

In order to establish (78)c, suppose first that $q \in \Sigma$ and $p[\mathcal{M}] = q[\mathcal{M}]$. By (77), we may conclude that for all $w' \in \mathbb{W}$, $p \in (\beta_{w'})_1$ iff $q \in (\beta_{w'})_1$; the RHS of (78)c now follows immediately from the definition of $\mathcal{M}$ (the argument parallels that for the left to right direction of (78)b above). Finally, suppose that $p$ and $q$ are contained in exactly the same selector-labels of edges exiting $w$. We may suppose that this set is nonempty, for otherwise the result follows from (78)b. It follows at once that $\sigma_w(p) = \sigma_w(q)$. Note that since $p, q \in \Sigma$, $\Box(p \leftrightarrow q) \in \Omega$; we may then conclude, by (74)f, that $\Box(p \leftrightarrow q) \in (\beta_w)_1$. Hence, by (75)a, for all $w' \in \mathbb{W}$, $\Box(p \leftrightarrow q) \in (\beta_{w'})_1$. But then, by (73)ab and (74)e, for all $w' \in \mathbb{W}$, $p \in (\beta_{w'})_1$ iff $q \in (\beta_{w'})_1$. We may conclude, by (77), that $p[\mathcal{M}] = q[\mathcal{M}]$.

*Induction Hypothesis*: Suppose that for all $w \in \mathbb{W}$,

(79) (a) $w \in \varphi[\mathcal{M}]$ iff $\varphi \in (\beta_w)_1$;
(b) $w \in \psi[\mathcal{M}]$ iff $\psi \in (\beta_w)_1$;
(c) $s(w, \varphi[\mathcal{M}])$ and $s(w, \psi[\mathcal{M}])$ are determined.

Induction Step: It follows immediately from (73)a,b and (79)a,b that for all $w \in \mathbb{W}$,

$$w \in (\varphi \wedge \psi)[\mathcal{M}] \quad \text{iff} \quad (\varphi \wedge \psi) \in (\beta_w)_1$$

and

$$w \in (\neg\varphi)[\mathcal{M}] \quad \text{iff} \quad (\neg\varphi) \in (\beta_w)_1.$$

It remains to show that

$$w \in (\varphi \preceq \psi)[\mathcal{M}] \quad \text{iff} \quad (\varphi \preceq \psi) \in (\beta_w)_1.$$

Suppose first that $w \in (\varphi \preceq \psi)[\mathcal{M}]$. Then, $s(w, \varphi[\mathcal{M}])$ and $s(w, \psi[\mathcal{M}])$ are both defined and $v_w(s(w, \varphi[\mathcal{M}])) = u(s(w, \varphi[\mathcal{M}])) \leq u(s(w, \psi[\mathcal{M}])) = v_w(s(w, \psi[\mathcal{M}]))$. It follows at once, by (74)g and (79)a,b, that $(\varphi \preceq \psi) \in (\beta_w)_1$. Suppose, on the other hand, that $w \notin (\varphi \preceq \psi)[\mathcal{M}]$. Then either at least one of $\varphi[\mathcal{M}]$ or $\psi[\mathcal{M}]$ is empty, or

$\upsilon_w(s(w,\psi[\mathcal{M}])) = u(s(w,\psi[\mathcal{M}])) < u(s(w,\varphi[\mathcal{M}])) = \upsilon_w(s(w,\varphi[\mathcal{M}]))$. In either case, it follows from (79)a,b, (73)a, and (74)g, that $(\varphi \preceq \psi) \notin (\beta_w)_1$.

The extension of the definition of $s$ to $(\varphi \wedge \psi)[\mathcal{M}]$, $(\neg\varphi)[\mathcal{M}]$, and $(\varphi \preceq \psi)[\mathcal{M}]$ is justified exactly as in the basis of the induction. □

The above argument may be adapted to establish Corollaries (57) and (58).

Proof of Corollary (57):

We modify the construction of $(\mathbb{W}, s, u, t)$ in the argument for the right to left direction of Proposition (76) to build a finite model satisfying $\theta$ from a mosaic $B$ for $\theta$. Let $\mathbb{W} = \bigcup_n W_n$ be the set of worlds constructed in the proof above, and let $n$ be the first stage such that for every $w \in \mathbb{W}$ there is an $m < n$ and a $w' \in W_m$ such that $\beta_w = \beta_{w'}$. We now close the construction of $(\mathbb{W}, s, u, t)$ at stage $n+1$ by choosing the children of each $w \in W_n$ to be suitable worlds in $\bigcup_{m \leq n} W_m$ that satisfy the conditions in the construction of $(\mathbb{W}, s, u, t)$. □

Proof of Corollary (58):

We modify the definition of brick as follows. A *reflexive brick* $B = \langle \Delta, \sigma, \upsilon \rangle$ is a brick that satisfies the following additional condition:

(80) for all $\varphi \in \Delta$, $s(\varphi) = \Delta$.

A *reflexive mosaic* is a mosaic composed of reflexive bricks; a reflexive mosaic for $\theta$ is defined similarly. Corollary (58) follows from the next proposition.

(81) Proposition: For every $\theta \in \mathcal{L}(\mathbb{P}, \mathbb{S})$, $\theta$ is satisfiable in a reflexive model if and only if there is a reflexive mosaic for $\theta$.

The proof of Proposition (81) is a straightforward adaptation of the proof of Proposition (76). The only subtlety is that in the definition of the tree-like model $(\mathbb{W}, s, u, t)$ we can no longer define $u$ in advance, but must define it by recursion following the recursive construction of the tree. We extend the definition of $u$ to the children of a world $w$ by choosing rational values for the children in such a way as to establish an isomorphism with the order induced by $\upsilon_w$, remembering that $w$ will be chosen as a child of $w$ in the obvious way, so that $u$ must retain the value for $w$ that was determined at stage $n$. □

## Appendix: Proof of Theorem (59)

Let $\Sigma$ be the set of formulas of $\mathcal{L}(\mathbb{P}, \mathbb{S})$ that are satisfiable in the class of models that are reflexive, discernible, and weakly independent. We present a computable many-one reduction of the tiling problem to $\Sigma$, which thereby shows that $\Sigma$ is undecidable. (For the undecidability of the tiling problem, see Lewis and Papadimitriou 1981, §6.5; and for reductions of the tiling problem to decision problems in modal logic see, Blackburn et al. 2001, §6.5)

An instance of the tiling problem is given by a tiling system $\mathsf{D} = \langle D, d_0, H, V\rangle$, where $D$ is a finite set, $d_0 \in D$, and $H, V \subseteq D \times D$. A tiling by $\mathsf{D}$ is a function $f : \mathbb{N} \times \mathbb{N} \mapsto D$ such that

(82) (a) $f(0, 0) = d_0$;

(b) $\langle f(m, n), f(m + 1, n)\rangle \in H$, for all $m, n \in \mathbb{N}$;

(c) $\langle f(m, n), f(m, n + 1)\rangle \in V$, for all $m, n \in \mathbb{N}$.

Given a tiling system $\mathsf{D}$, we effectively construct a formula $\varphi_{\mathsf{D}}$ such that there is a tiling by $\mathsf{D}$ if and only if $\varphi_{\mathsf{D}}$ is satisfied by some model $\mathcal{M}$ which is reflexive, discernible and weakly independent. The signature $(\mathbb{P}, \mathbb{S})$ for $\varphi_{\mathsf{D}}$ has $\mathbb{P} = \{p, q, r\} \cup \{p_d \mid d \in D\}$, and $\mathbb{S} = \{X, Y\}$. The formula $\varphi_{\mathsf{D}}$ will be the conjunction of the following conditions:

(a) $\gamma$, a formula that imposes a grid-like structure on a subset of the worlds of any structure that satisfies it;

(b) $\pi$, a formula that associates a unique tile to each world;

(c) $\chi$, a formula requiring that the placement of tiles on the grid-like substructure is compatible with the tiling conditions; and

(d) $\delta$, a formula insuring that $d_0$ is placed at the origin of the grid-like substructure.

The formula $\gamma$ is the conjunction of $\Diamond r$ with all the sentences in the accompanying table. To explain, let reflexive, discernible and weakly independent model $\mathcal{M} = (\mathbb{W}, s, u, t)$ satisfy $\gamma$. Suppose $w_{0,0} \in \mathbb{W}$ satisfies $p \wedge q \wedge r$. We may then view $w_{0,0}$ as the origin of a grid whose horizontal axis is labelled by an infinite sequence $x_0, x_1, \ldots$ of $X$-utilities and whose vertical axis is labelled with an infinite sequence $y_0, y_1, \ldots$ of $Y$-utilities. The even columns of the grid consist of worlds that satisfy $p \wedge r$ while the odd columns consist of worlds that satisfy $\neg p \wedge r$. The even and odd rows are similarly distinguished by $q \wedge r$ and $\neg q \wedge r$. The formulas in the first rectangle of the table push us horizontally across even rows of the grid from even columns to odd columns (first formula) and from odd to even columns (second formula); the third and fourth formulas maintain motion along the same even row. The formulas in the other three rectangles of the table have similar effect thereby constructing the grid along odd rows and even and odd columns.

| |
|---|
| $\Box((p \wedge q \wedge r) \rightarrow ((p \wedge q \wedge r) \prec_X (\neg p \wedge q \wedge r)))$ |
| $\Box((\neg p \wedge q \wedge r) \rightarrow ((\neg p \wedge q \wedge r) \prec_X (p \wedge q \wedge r)))$ |
| $\Box((p \wedge q \wedge r) \rightarrow ((p \wedge q \wedge r) \approx_Y (\neg p \wedge q \wedge r)))$ |
| $\Box((\neg p \wedge q \wedge r) \rightarrow ((\neg p \wedge q \wedge r) \approx_Y (p \wedge q \wedge r)))$ |
| $\Box((p \wedge q \wedge r) \rightarrow ((p \wedge q \wedge r) \prec_Y (p \wedge \neg q \wedge r)))$ |
| $\Box((p \wedge \neg q \wedge r) \rightarrow ((p \wedge \neg q \wedge r) \prec_Y (p \wedge q \wedge r)))$ |
| $\Box((p \wedge q \wedge r) \rightarrow ((p \wedge q \wedge r) \approx_X (p \wedge \neg q)))$ |
| $\Box((p \wedge \neg q \wedge r) \rightarrow ((p \wedge \neg q \wedge r) \approx_X (p \wedge q \wedge r)))$ |
| $\Box((\neg p \wedge q \wedge r) \rightarrow ((\neg p \wedge q \wedge r) \prec_Y (\neg p \wedge \neg q \wedge r)))$ |
| $\Box((\neg p \wedge \neg q \wedge r) \rightarrow ((\neg p \wedge \neg q \wedge r) \prec_Y (\neg p \wedge q \wedge r)))$ |
| $\Box((\neg p \wedge q \wedge r) \rightarrow ((\neg p \wedge q \wedge r) \approx_X (\neg p \wedge \neg q \wedge r)))$ |
| $\Box((\neg p \wedge \neg q \wedge r) \rightarrow ((\neg p \wedge \neg q \wedge r) \approx_X (\neg p \wedge q \wedge r)))$ |
| $\Box((p \wedge \neg q \wedge r) \rightarrow ((p \wedge \neg q \wedge r) \prec_X (\neg p \wedge \neg q \wedge r)))$ |
| $\Box((\neg p \wedge \neg q \wedge r) \rightarrow ((\neg p \wedge \neg q \wedge r) \prec_X (p \wedge \neg q \wedge r)))$ |
| $\Box((p \wedge \neg q \wedge r) \rightarrow ((p \wedge \neg q \wedge r) \approx_Y (\neg p \wedge \neg q \wedge r)))$ |
| $\Box((\neg p \wedge \neg q \wedge r) \rightarrow ((\neg p \wedge \neg q \wedge r) \approx_Y (p \wedge \neg q \wedge r)))$ |

Table 1: Conjuncts of the formula $\gamma$, implying the existence of a two dimensional grid of worlds.

The reflexivity of $\mathcal{M}$ allows us to conclude that the sequences of $X$ and $Y$ coordinates, $x_0, x_1, \ldots$ and $y_0, y_1, \ldots$, are strictly increasing, while the combination of discernibility and weak independence allows us to conclude that we arrive at the same world by going "north" and then "east," as we do by going "east" and then "north."

In order to specify the formulas $\pi$, $\chi$, and $\delta$, we associate to each tile $d \in D$ a sentence letter $p_d$. The worlds of the grid themselves satisfy none of these sentence letters, so one conjunct of $\pi$ is

$$\Box(r \rightarrow \bigwedge_{d \in D} \neg p_d).$$

The other conjunct of $\pi$ marks each world of the grid with a unique sentence letter: $\oplus_{d \in D}(r \prec_X p_d)$, where $\oplus$ signifies exclusive disjunction. The formula $\delta$ is $\Diamond((p \wedge q \wedge r) \prec_X p_{d_0})$. Finally, the formula $\chi$ is the conjunction of the following sentences:

$$\textstyle\bigvee_{(d,e) \in H}(p \wedge q \wedge r) \rightarrow ((p \wedge q \wedge r) \prec_X p_d) \wedge ((\neg p \wedge q \wedge r) \prec_X p_e);$$

$$\textstyle\bigvee_{(d,e) \in H}(\neg p \wedge q \wedge r) \rightarrow ((\neg p \wedge q \wedge r) \prec_X p_d) \wedge ((p \wedge q \wedge r) \prec_X p_e);$$

$$\bigvee_{(d,e)\in H}(p \wedge \neg q \wedge r) \rightarrow ((p \wedge \neg q \wedge r) \prec_X p_d) \wedge ((\neg p \wedge \neg q \wedge r) \prec_X p_e);$$

$$\bigvee_{(d,e)\in H}(\neg p \wedge \neg q \wedge r) \rightarrow ((\neg p \wedge \neg q \wedge r) \prec_X p_d) \wedge ((p \wedge \neg q \wedge r) \prec_X p_e);$$

$$\bigvee_{(d,e)\in V}(p \wedge q \wedge r) \rightarrow ((p \wedge q \wedge r) \prec_X p_d) \wedge ((\neg p \wedge \neg q \wedge r) \prec_X p_e);$$

$$\bigvee_{(d,e)\in V}(\neg p \wedge q \wedge r) \rightarrow ((\neg p \wedge q \wedge r) \prec_X p_d) \wedge ((\neg p \wedge \neg q \wedge r) \prec_X p_e);$$

$$\bigvee_{(d,e)\in V}(p \wedge \neg q \wedge r) \rightarrow ((p \wedge \neg q \wedge r) \prec_X p_d) \wedge ((p \wedge q \wedge r) \prec_X p_e);$$

$$\bigvee_{(d,e)\in V}(\neg p \wedge \neg q \wedge r) \rightarrow ((\neg p \wedge \neg q \wedge r) \prec_X p_d) \wedge ((\neg p \wedge q \wedge r) \prec_X p_e).$$

It can now be verified that $\varphi_{\mathsf{D}}$ is satisfied in a reflexive, discernible and weakly independent model if and only if there is a tiling according to $\mathsf{D}$.

It follows at once that the set $F$ of formulas satisfiable in the class $C$ of reflexive, discernible and weakly independent models is co-r.e.-hard. Moreover, $C$ is first-order definable via the natural translation of $\mathcal{L}(\mathbb{P}, \mathbb{S})$ into first-order formulas. Hence, by the completeness theorem for first-order logic, $F$ is co-r.e. (see Blackburn et al., 2001, §6.5). □

## Appendix: Proof of Theorem (60)

Both sets of formulas are **NP**-hard since the satisfiability problem for non-modal sentential logic is ptime-reducible to each. A straightforward application of the mosaic method yields the conclusion that each is in **NP**. □

## Appendix: Proof of Theorem (61)

We derive the compactness theorem for $\mathcal{L}(\mathbb{P}, \mathbb{S})$, where the signature $(\mathbb{P}, \mathbb{S})$ is countable, as a corollary to the compactness theorem for first-order logic. Our argument follows a standard strategy which proceeds via translating a modal language into a (fragment of) first-order logic. The translation essentially codifies the definition of satisfaction over some class of relational frames. In our case, this direct strategy requires modification since our "frames" are not (first-order) relational structures, in particular, the selection function has a "type 1" argument, the proposition from which a salient confirming representative is chosen. Moreover, the utility functions, whose range is the set of real numbers, present another obstacle to smooth "first-orderization" in the context of our compactness argument. To overcome these difficulties, the first step in our compactness proof is to "compile" a structure $\mathcal{M} = (\mathbb{W}, s, u, t)$ into a relational structure $\mathcal{F}_{\mathcal{M}}$ and to translate each sentence $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ to a first-order formula $\varphi^\dagger(x)$ with one free variable such that for all $w \in \mathbb{W}$,

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

Given $\mathcal{M} = (\mathbb{W}, s, u, t)$ of signature $(\mathbb{P}, \mathbb{S})$, we define $\mathcal{F}_\mathcal{M}$. The signature of $\mathcal{F}_\mathcal{M}$ consists of a unary relation symbol $Q_p$, for each $p \in \mathbb{P}$; a binary relation symbol $\leq_X$, for each $X \in \mathbb{S}$; and a binary relation symbol $R_\varphi$, for each sentence $\varphi$ of $\mathcal{L}(\mathbb{P}, \mathbb{S})$.

(84) The interpretation of each relation symbol in the signature of $\mathcal{F}_\mathcal{M}$ is defined as follows (we suppress the superscript $\mathcal{F}_\mathcal{M}$ on each relation symbol):

(a) $Q_p = t(p)$;

(b) for all $w, w' \in \mathbb{W}$, $w \leq_X w'$ iff $u_X(w) \leq u_x(w')$;

(c) for all $w, w' \in \mathbb{W}$, $R_\varphi(w, w')$ iff $w' = s(w, \varphi[\mathcal{M}])$.

Note that (84)c implies that $R_\varphi$ is the empty relation if and only if $\varphi[\mathcal{M}] = \emptyset$.

(85) We now define, by recursion, for each sentence $\varphi$ of $\mathcal{L}(\mathbb{P}, \mathbb{S})$, its translation $\varphi^\dagger(x)$, a formula with one free variable in the first-order language of $\mathcal{F}_\mathcal{M}$.

(a) $p^\dagger = Q_p(x)$;

(b) $(\varphi \wedge \psi)^\dagger = \varphi^\dagger(x) \wedge \psi^\dagger(x)$;

(c) $(\neg\varphi)^\dagger = \neg\varphi^\dagger(x)$;

(d) $(\varphi \preceq_X \psi)^\dagger = (\exists y)(\exists z)(\varphi^\dagger(y) \wedge \psi^\dagger(z) \wedge R_\varphi(x, y) \wedge R_\psi(x, z) \wedge y \leq_X z)$.

On the basis of (84) and (85), it is now easy to verify (83).

Next, we describe a first-order theory $D$ in the signature of $\mathcal{F}_\mathcal{M}$ such that

(86) for every $\mathcal{M}$, $\mathcal{F}_\mathcal{M} \models D$

and

(87) for every countable first-order structure $A$, if $A \models D$, then for some $\mathcal{M}$, $A = \mathcal{F}_\mathcal{M}$.

We proceed to describe $D$.

(88) $D$ consists of the following first-order sentences.

(a) $(\forall x)(\forall y)(\forall z)(x \leq_X y \to (y \leq_X z \to x \leq_X z))$, for each $X \in \mathbb{S}$;

(b) $(\forall y)(y \leq_X y)$, for each $X \in \mathbb{S}$;

(c) $(\forall y)(\forall z)(y \leq_X z \vee z \leq_X y)$, for each $X \in \mathbb{S}$;

(d) $(\forall x)(\forall y)(R_\varphi(x, y) \to \varphi^\dagger(y))$, for each $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$;

(e) $(\exists x)\varphi^\dagger(x) \to (\forall x)(\exists y)(\forall z)(R_\varphi(x, z) \leftrightarrow y = z)$, for each $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$;

(f) $(\forall x)(\varphi(x) \leftrightarrow \psi(x)) \to (\forall x)(\forall y)(R_\varphi(x, y) \leftrightarrow R_\psi(x, y))$, for each $\varphi, \psi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$.

It is easy to verify (86) by direct inspection of the clauses of (88). In order to verify (87), we argue as follows. Let $A$ be a countable relational structure satisfying $D$. We define a structure $\mathcal{M} = (\mathbb{W}, s, u, t)$. First, let $\mathbb{W} = |A|$ and let $t(p) = Q_p^A$, for each $p \in \mathbb{P}$. Next, by (88)a-c, for each $X \in \mathbb{S}$, $\leq_X^A$ is a countable linear pre-order. It follows from the universality of the rational numbers among countable linear orders that a utility function $u_X$ may be chosen so that for all $w, w' \in \mathbb{W}, w \leq_X w'$ if and only if $u_X(w) \leq u_X(w')$. For each $\varphi \in \mathcal{L}(\mathbb{P}, \mathbb{S})$ and $w, w' \in \mathbb{W}$, let $s(w, \varphi[\mathcal{M}]) = w'$ if and only if $R_\varphi(w, w')$. Finally, let $s(w, P)$ be an arbitrarily chosen element of $P$ for any proposition $P \subseteq \mathbb{W}$ which is not expressed by a sentence. It is easy to see that the structure $\mathcal{M}$ satisfies (87).

We now derive (61) from the compactness theorem for first-order logic. Let $T$ be a set of sentences of $\mathcal{L}(\mathbb{P}, \mathbb{S})$ and suppose that every finite subset $T' \subseteq T$ is satisfiable. It follows at once from (83) and (86) that for every finite $T' \subseteq T$, $\{\varphi^\dagger(c) \mid \varphi \in T'\} \cup D$ is satisfiable (here c is a constant symbol). Therefore, by the Compactness and Löwenheim-Skolem Theorems for first-order logic, there is a countable structure $A$ such that $A \models \{\varphi^\dagger(c) \mid \varphi \in T\} \cup D$. Hence, by (83) and (87), $T$ is satisfiable. $\square$

## References


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.

P. Gärdenfors. *Knowledge in Flux: Modeling the dynamics of Epistemic States.* MIT Press, Cambridge MA, 1988.

D. M. Grether and C. R. Plott. Economic theory of choice and the preference reversal phenomenon. *American Economic Review*, 69:623–628, 1979.

J. Haidt. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108:814–834, 2001.

Paul E. Johnson. *Social Choice: Theory and Research.* Sage Publications, New York NY, 1998.

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.

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.

I. Levi. *Hard Choices: Decision making under unresolved conflict.* Cambridge University Press, Cambridge, 1986.

D. Lewis. *Counterfactuals.* Harvard University Press, 1973.

H. Lewis and C. Papadimitriou. *Elements of the Theory of Computation.* Prentice-Hall, 1981.

S. Lichtenstein and P. Slovic, editors. *The Construction of Preference.* Cambridge University Press, New York NY, 2006.

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

D. M. Messick. Social interdependence and decision making. In G. Wright, editor, *Behavioral Decision Making*, pages 87–109. Plenum, New York NY, 1985.

J. W. Payne and C. Puto. Adding asymmetrically dominated alternatives: violations of regularity and the similarity hypothesis. *The Journal of Consumer Research*, 9: 90–5, 1982.

P. Pettit. *Rules, Reasons, and Norms*. Oxford University Press, Oxford UK, 2002.

L. J. Savage. *The Foundations of Statistics*. Wiley, 1954.

A. Sen. Choice functions and revealed preference. *Review of Economic Studies*, 38: 307–317, 1971.

E. Shafir, I. Simonson, and A. Tversky. Reason-based choice. *Cognition*, 49:11–36, 1993.

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

Katya Tentori, D. Osherson, L. Hasher, and C. May. Wisdom and aging: Irrational preferences in college students but not older adults. *Cognition*, 81(3):B87–B96, 2001.

A. Tversky, S. Sattath, and P. Slovic. Contingent weighting in judgment and choice. *Psychological Review*, 95:371–384, 1988.

J. van Benthem and F. Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logic*, 17:157–182, 2007.

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.

G. H. von Wright. *The Logic of Preference*. Edinburgh University Press, Edinburgh UK, 1963.