Model-Based Revealed Preference

Peter Caradonna and Christopher P. Chambers

April 4, 2024

Abstract

We consider the problem of rationalizing a choice data set by a preference satisfying an arbitrary collection of *invariance* axioms. Examples of such axioms include quasilinearity, homotheticity, independence-type axioms for mixture spaces, constant relative/absolute risk and ambiguity aversion axioms, stationarity for dated rewards or consumption streams, and many others. We provide a complete characterization of invariant rationalizability via a novel approach which relies on tools from the theoretical computer science literature on automated theorem proving. We also establish a generalization of the Dushnik-Miller theorem, which we use to give a complete description of the out-of-sample predictions generated by the data under any such model.

1 Introduction

Nearly all economic models are built on a foundation of economic actors, who seek to maximize their individual well-being. Any such model must therefore specify, in a stylized way, the manner in which these actors evaluate various forms of trade-offs and decisions. If these underlying assumptions are inconsistent with the empirical regularities observed in human behavior, these errors bleed into every other aspect of the model, potentially leading to unrealistic or outright incorrect predictions (e.g. Mehra and Prescott 1985).

This motivates a basic need to be able to obtain, in a systematic fashion, the testable implications of various models of preference and decision. Traditionally, revealed preference theory studied the testable implications of rational behavior generally (Samuelson 1938; Richter 1966; Afriat 1967). Since Afriat (1967), a number of papers have utilized the same basic approach to characterize the testable implications of particular models in the special case of price-consumption data (e.g., Varian 1983; Diewert 2012; Echenique and Saito 2015; Chambers et al. 2016). Methodologically, however, these results often rely on model-specific considerations to obtain characterizations, and hence only apply narrowly to single sets of assumption.

In contrast, this paper provides a characterization the testable implications of a wide variety of models, across a host of different domains, through a single, general approach. We consider the class of preferences which satisfy what we term *invariance* axioms. More formally, suppose X is a set of alternatives. A preference relation $\succsim$ is invariant under a transformation $\omega : X \to X$ if:

$$x \succsim y \quad \Longleftrightarrow \quad \omega(x) \succsim \omega(y),$$

for every pair of alternatives x and y. In this setting, we provide complete answers to the two following natural questions:

(I) When is a given revealed preference data set consistent with the maximization of some preference relation that is invariant under every function in some fixed set of transformations?

(II) What comparisons *not* observed in the data are nonetheless agreed upon by *every* invariant rationalizing preference?[1]

Consider the following example.

Example 1. Let $Z = \{x, x', y, y'\}$ be a set of prizes, and suppose X consists of all infinite horizon consumption streams taking values in Z. Suppose we observe a revealed preference relation $\succsim^R$ based on a subject's choices, and are interested in whether this data is consistent with the maximization of any *stationary* preference (à la Koopmans 1960), i.e. which satisfies:

$$(\sigma_1, \sigma_2, \ldots) \succsim (\sigma'_1, \sigma'_2, \ldots) \quad \Longleftrightarrow \quad (z, \sigma_1, \ldots) \succsim (z, \sigma'_1, \ldots)$$

for every pair of consumption streams $\sigma, \sigma' \in X$ and every prize z. Suppose we only observe that:

$$(x', \sigma_1, \ldots) \succ^R (x, \sigma'_1, \ldots)$$
$$(x, \sigma_1, \ldots) \succ^R (x', \sigma'_1, \ldots),$$

and

$$(y', \sigma'_1, \ldots) \succ^R (y, \sigma_1, \ldots)$$
$$(y, \sigma'_1, \ldots) \succ^R (y', \sigma_1, \ldots),$$

for some σ, σ'. Clearly our revealed preference relation contains no cycles, and hence by classical results (e.g., Richter 1966) is consistent with the maximization of some preference relation. However, no such preference can be stationary.

To see this, note that from the first two relations, any such preference $\succeq^*$ must rank $\sigma \succ^* \sigma'$, as if it instead ranked $\sigma' \succeq^* \sigma$, then:

$$(x', \sigma_1, \ldots) \succ^* (x, \sigma'_1, \ldots) \succeq^* (x, \sigma_1, \ldots) \succ^* (x', \sigma'_1, \ldots) \succeq^* (x', \sigma_1, \ldots),$$

violating transitivity. In this cycle, the first and third relations follow from $\succeq^*$ being consistent with $\succsim^R$, and the second and fourth are consequences of stationarity, given that $\sigma \succeq^* \sigma'$. However, an analogous argument applied to the latter two observations implies any such relation must also satisfy $\sigma' \succ^* \sigma$, an impossibility. Thus despite containing no *explicit* contradictions of stationarity or rationality, $\succsim^R$ is inconsistent with every stationary preference. ■

[1] Conditional upon the set of such preferences being non-empty.

This example highlights that even sets of comparisons that are very sparse can lead to strong (sometimes even impossible-to-fulfill) restrictions on the possible comparisons a consistent, invariant preference can make. When we seek to add a comparison between a pair of unranked alternatives, this addition also generates a multitude of knock-on effects: the comparisons between the images of the pair under each transformation in our family. These extra additions can form transitivity violations or cycles, even when the initial added comparison itself does not.

In order to account for the potential infinity of knock-on effects, we are forced to consider *sets* of simultaneous restrictions on the possible comparisons a rationalizing preference can make. In turn, these sets of restrictions can be combined to deduce further restrictions that do not emerge 'directly' from the data.

We introduce a simple, binary operation that we term the 'collapse,' that converts a suitable pair of restriction sets into a new one. This operation may roughly be viewed as a set-valued analogue of the operation mapping a pair of compatible relations $x \succeq y$ and $y \succeq z$ to their implication $x \succeq z$ under transitivity. We show that, no matter the complexity of the environment or structure of the family of transformations, a simple no cycle condition, phrased in terms of our collapse operation, fully characterizes rationalizability by an invariant preference. We also prove that a related generalization of the transitive closure, again in terms of our collapse operation, completely characterizes the set of out-of-sample predictions generated by the data under a given model.

Our methodology relies intimately on a connection with formal logic. We first recast the problem of finding a consistent, invariant preference as one of testing the satisfiability of a set of clauses. We show that the sets of restrictions determined by the data may be expressed as a particular type of clause in this logical framework. Using this, we establish that a 'cycle' our sense may be used to construct as a formal proof of unsatisfiability in the accompanying logical system. To prove the converse, we use a result due to Robinson (1965), which establishes that if a given system of clauses is unsatisfiable, there exists a proof of this fact with a specific combinatorial structure. We then show that any proof of unsatisfiability of such form can always be used to construct a 'cycle.'

The paper proceeds as follows. In Section 2 we formally state our problem and provide numerous economic examples covered by our results. Section 3 considers a special case of our general result—the case in which all the transformations defining our invariance axiom commute. In this special case, we show that our general no-cycle condition reduces to a more familiar form. We additionally show that, when our initial revealed preference data arises from some underlying price-consumption data set, that we are able to immediately recover familiar characterizations from the literature as special cases. In Section 4, we introduce our collapse operation, and provide our general characterization of invariant rationalizability. Finally, Section 5 concludes with a novel generalization of the Dushnik-Miller theorem (Dushnik and Miller 1941) for invariant preferences, which we use to characterize the set of out-of-sample, or counterfactual, predictions generated by a given set of data and model.

1.1 Related Literature

The revealed preference literature is too large to adequately survey here, see Chambers and Echenique (2016) for an overview.[2] Classically, Richter (1966) was the first to characterize rationalizability for the abstract choice model. We obtain Richter's original theorem as a special case of our main results (see Section 3). A similarly classic reference in this vein is Duggan (1999) who retains an abstract framework but imposes additional restrictions on the interpretation of 'rationality.'

Other authors have studied the problem of rationalizing choice data via preferences with various general structures. Nishimura et al. (2017) study this problem for continuous and monotone preferences on various spaces. Demuynck (2009) investigates a general class of 'closure operators' on spaces of binary relations that generalize the transitive closure, and obtains a general extension result for algebraic structures satisfying certain properties.[3] While general, applying these tools requires non-trivial effort to establish their conditions are satisfied. In contrast, our results focuses on a smaller mathematical class of algebraic properties, invariance and monotonicity axioms, but are able to derive results that are immediately applicable.

Other authors have considered invariant preferences in various contexts. Ok and Riella (2014, 2021) consider various extension results for invariant preorders on groups. In contrast, we consider both a more general class of primitive relations and more general notion of invariance.[4] Recently Freer and Martinelli (2022), building off the tools of Demuynck (2009), consider the problem of invariant rationalization by incomplete or non-transitive binary relation.[5] Dubra et al. (2004) show that every 'incomplete' expected utility (EU) preference may be completed in such a way as to preserve the EU axioms.

Dushnik and Miller (1941) show that every partial order is equal to the intersection of its linear order extensions. Several authors in economics have taken interest in such unanimity, or Pareto, representation of incomplete preferences. Abstract approaches include Donaldson and Weymark (1998); Bossert (1999); Weymark (2000) and Alcantud (2009). In concrete economic environments, similar representations can be found in, for example, the theory of expected utility preferences (Dubra et al. 2004; Gorno 2017), Krepsian style preferences over menus (Nehring and Puppe 1999), or rankings of accomplishments (Chambers and Miller 2018).

[2]See also Echenique (2020) for a summary of some recent work in this space.

[3]See Ward (1942) for a general theory of closures.

[4]Mathematically, our notion of invariance corresponds to invariance of a preference under an arbitrary semi-group action on the consumption space. For definitions, see Fuchs (2011).

[5]They also establish an invariant rationalizability result in the special case the collection of transformations, under composition, forms a linearly ordered group.

We are not the first paper to exploit the connection between revealed preference and formal logic. Chambers et al. (2014) study the general form of empirical content for theories in first-order logic and Chambers et al. (2017) study some properties of empirical content. In comparison, our results rely only on the simpler framework of propositional logic. Gonczarowski et al. (2019) show that similar connections with propositional logic obtain in a variety of economic contexts.

Robinson (1965) showed that a certain algorithmic operation on logical clauses called resolution was sound and refutation-complete. This reduced the problem of proving a set of clauses to be inconsistent to a discrete search problem. A number of extensions and refinements giving various 'normal forms' for proofs were established in the early artificial intelligence literature to attempt to further reduce the complexity of this search space (see, e.g., Schöning 2008 for an overview).

Finally, our work presupposes no notion of topology, but many works in economics consider topological aspects of the extension problem. Aumann (1962, 1964); Peleg (1970); Levin (1983) are classical references, but the theory has developed much since then (e.g., Ok 2002; Nishimura et al. 2017).

2 The Model

Let X denote set of alternatives. A **preference relation** $\succsim$ is a complete and transitive binary relation on X. Given a preference relation, we use $\succ$ and $\sim$ to denote its asymmetric and symmetric components, respectively.

Let $\mathcal{M}$ denote a set of transformations, each mapping $X \to X$. We say that a preference relation is **$\mathcal{M}$-invariant** if, for all $x, y \in X$ and all $\omega \in \mathcal{M}$:

$$x \succsim y \quad \iff \quad \omega(x) \succsim \omega(y).$$

Note that if $\omega, \omega' \in \mathcal{M}$, then any $\mathcal{M}$-invariant preference relation also satisfies:

$$x \succsim y \quad \iff \quad (\omega \circ \omega')(x) \succsim (\omega \circ \omega')(y)$$

and

$$x \succsim y \quad \iff \quad (\omega' \circ \omega)(x) \succsim (\omega' \circ \omega)(y),$$

thus it is without loss of generality to assume that $\mathcal{M}$ is closed under composition. Going forward, we will assume both (i) the identity function $\mathrm{id} \in \mathcal{M}$, and (ii) $\mathcal{M}$ is closed under composition.[6]

We consider a pair of primitive **revealed preference** relations, denoted $\langle \succsim^R, \succ^R \rangle$, where $\succ^R$ is a sub-relation of $\succsim^R$.[7] These relations could arise through observed choice behavior, e.g. by defining:

[6] Formally, we assume, without loss of generality, that $(\mathcal{M}, \circ)$ forms a semigroup with identity, or a *monoid*.

[7] We do not, however, assume that $\succ^R$ is necessarily the asymmetric component of $\succsim^R$.

- $x \succsim^R y$ if x and y belong to a common choice set, from which it was observed x was chosen.
- $x \succ^R y$ if $x \succsim^R y$ and, in addition, y was not chosen.

However, we explicitly allow for them also encoding other salient properties such as *monotonicity restrictions*, by setting $x \succ^R y$ if x dominates y in a particular partial order of interest. We will assume, without loss of generality, that $\succsim^R$ is reflexive.

We interpret $\langle \succsim^R, \succ^R \rangle$ as **data**, and seek to understand when it is consistent with the behavior of an economic actor who chooses to maximize some $\mathcal{M}$-invariant preference relation $\succsim$. Formally, an $\mathcal{M}$-invariant preference relation $\succsim$ **rationalizes** the data $\langle \succsim^R, \succ^R \rangle$ if both (i) $\succsim^R \subseteq \succeq$, and (ii) $\succ^R \subseteq \succ$. The primary result of our paper will be to provide a complete characterization of those data sets $\langle \succsim^R, \succ^R \rangle$ that are rationalizability by an $\mathcal{M}$-invariant preference, for any choice of X or $\mathcal{M}$. Equivalently, we characterize which data sets may *not* be rationalized by any $\mathcal{M}$-invariant preference (for fixed choice of $\mathcal{M}$), which therefore provides a complete description of the empirical content of such models.

The transitive closure of our pair $\langle \succsim^R, \succ^R \rangle$ is the pair of relations $\langle \succsim^R_{\mathsf{T}}, \succ^R_{\mathsf{T}} \rangle$, where $x \succsim^R_{\mathsf{T}} y$ if and only if there exists some finite sequence $x_0, \ldots, x_N \in X$ such that:

$$x = x_0 \succsim^R x_1 \succsim^R \cdots \succsim^R x_N = y.$$

Similarly, $x \succ^R_{\mathsf{T}} y$ if $x \succsim^R_{\mathsf{T}} y$ and some relation in the sequence belongs to $\succ^R$. A pair of relations $\langle \succsim^R, \succ^R \rangle$ is said to be **acyclic** if there do not exist $x_0, \ldots, x_N \in X$ such that:

$$x_0 \succsim^R x_1 \succsim^R \cdots \succsim^R x_N \succ^R x_0.$$

We refer to such a sequence as a **cycle**. Note that if the pair $\langle \succsim^R, \succ^R \rangle$ is acyclic, then $\succ^R$ is necessarily the asymmetric component of $\succsim^R$.

2.1 Examples of Invariant Preferences

Before we formally state our results, in this section we provide a number of examples of various types of invariance axioms for models of preference. All of these correspond to various special cases of our notion of $\mathcal{M}$-invariance, for particular choices of X and $\mathcal{M}$.

2.1.1 Quasilinearity

Let $X = \mathbb{R}_+ \times Z$. A preference is said to be *quasilinear* if:

$$(t, z) \succsim (t', z') \quad \Longleftrightarrow \quad (t + \alpha, z) \succsim (t' + \alpha, z')$$

for all $(t, z), (t', z') \in X$ and $\alpha \geq 0$. When (t, z) is interpreted as the 'dated reward,' corresponding to the delivery of a prize z to an agent at t units of time into the future, quasilinearity is also referred to as *stationarity* (see Fishburn and Rubinstein 1982). See also Caradonna (2023).

2.1.2 (Generalized) Homotheticity

Let X be a cone in a real vector space. A preference is *homothetic* if:

$$x \succsim y \iff tx \succsim ty$$

for all scalars $t > 0$. Similarly, Cobb-Douglas preferences are the unique, continuous and monotone preferences on $\mathbb{R}^N_+$ satisfying the related but more general form of invariance:

$$(x_1, \ldots, x_N) \succsim (y_1, \ldots, y_N) \iff (t_1 x_1, \ldots, t_N x_N) \succsim (t_1 y_1, \ldots, t_N y_N),$$

for all $x, y \in \mathbb{R}^N_+$ and all $(t_1, \ldots, t_N) \in \mathbb{R}^N_{++}$ (see Trockel 1989).

2.1.3 Mixture Invariance

Suppose that $X = \Delta(Z)$, the set of all Borel probability measures on some metrizable space Z. A preference satisfies the *independence* axiom of von Neumman and Morgenstern (Von Neumann and Morgenstern 1947) if:

$$\mu \succsim \nu \iff \alpha\mu + (1-\alpha)\eta \succsim \alpha\nu + (1-\alpha)\eta$$

for all $\alpha \in (0, 1]$ and $\eta \in X$.[8] If instead X denotes the Anscombe-Aumann domain $\mathcal{F}$ of simple, measurable maps from some measurable space S into $\Delta(Z)$, the *independence* axiom takes the form:

$$f \succsim g \iff \alpha f + (1-\alpha)h \succsim \alpha g + (1-\alpha)h$$

for $\alpha \in (0, 1]$ and some act $h \in X$. Similarly, common weakenings of independence such as *certainty independence* (Gilboa and Schmeidler 1989), *weak certainty independence* (Maccheroni et al. 2006), *worst independence* (Chateauneuf and Faro 2009), *risk independence* (Cerreia-Vioglio et al. 2011) and so forth are all of this form.

2.1.4 Stationarity

Let $X = Z^{\mathbb{N}}$ denote the set of all infinite horizon consumption streams taking values in some set of prizes Z. A preference on X is said to be *stationary* in the sense of Koopmans (1960) if:

$$(x_1, x_2, \ldots) \succsim (x'_1, x'_2, \ldots) \iff (z, x_1, x_2, \ldots) \succsim (z, x'_1, x'_2, \ldots)$$

for all $z \in Z$. See also Epstein (1983).

[8] More generally, this form of invariance may be defined for any mixture space. See, e.g., Herstein and Milnor (1953).

2.1.5 Convolution Invariance

Suppose X consists of all lotteries on $\mathbb{R}$ with bounded support. Mu et al. (2021) consider continuous weak orders on X that are monotone with respect to first-order stochastic dominance, and invariant under convolutions:

$$\mu \succsim \nu \quad \Longleftrightarrow \quad \mu * \eta \succsim \nu * \eta,$$

for all $\eta \in X$.[9] A preference on X is said to exhibit *constant absolute risk aversion* (e.g., Safra and Segal 1998) if:

$$\mu \succsim \nu \quad \Longleftrightarrow \quad \mu * \delta_\alpha \succsim \nu * \delta_\alpha$$

for all $\alpha \in \mathbb{R}$, where δ_α denotes the Dirac measure centered at α.[10]

2.1.6 Product & Dilution Invariance

Let X consist of all finite Blackwell experiments on some fixed, finite set of states of the world Θ. Thus an element of X is a tuple $\left(S, \{\mu_\theta\}_{\theta\in\Theta}\right)$, where S is a finite set of signals, and each μ_θ is a probability measure on S. Pomatto et al. (2023) consider 'costliness' orderings over X that are invariant under two varieties of transformations: the formation *products*, and so-called *dilutions*. In this context, products of Blackwell experiments formalize the idea of running two simultaneous and independent experiments. The α-dilution of an experiment, denoted $\alpha \cdot \left(S, \{\mu_\theta\}_{\theta\in\Theta}\right)$, is the experiment $\left(S \cup \{o\}, \{\mu'_\theta\}_{\theta\in\Theta}\right)$, where o is a completely uninformative signal, and (i) $\mu'_\theta(A) = \alpha\mu_\theta(A)$ for all $A \subseteq S$, and $\mu'_\theta(\{o\}) = 1 - \alpha$. This corresponds to invariance under:

$$\left(S, \{\mu_\theta\}_{\theta\in\Theta}\right) \succsim \left(S', \{\nu_\theta\}_{\theta\in\Theta}\right) \iff \left(S \times T, \{\mu_\theta \otimes \eta_\theta\}_{\theta\in\Theta}\right) \succsim \left(S' \times T, \{\nu_\theta \otimes \eta_\theta\}_{\theta\in\Theta}\right)$$

for all $(T, \{\eta_\theta\}_{\theta\in\Theta}) \in X$, and:

$$\left(S, \{\mu_\theta\}_{\theta\in\Theta}\right) \succsim \left(S', \{\nu_\theta\}_{\theta\in\Theta}\right) \iff \alpha \cdot \left(S, \{\mu_\theta\}_{\theta\in\Theta}\right) \succsim \alpha \cdot \left(S', \{\nu_\theta\}_{\theta\in\Theta}\right)$$

for all $\alpha \in (0, 1]$.

3 Characterizing Rationalizability: The Case of Commuting Transforms

In this section, we consider the case in which each pair of transformations in $\mathcal{M}$ commute, i.e.:

$$(\omega \circ \omega')(x) = (\omega' \circ \omega)(x)$$

[9] The term 'additive' in the paper's title refers to this property when the preference is equivalently regarded as being defined over (bounded) random variables.

[10] Similarly, *constant relative risk aversion* is also a special case of $\mathcal{M}$-invariance, where $\mathcal{M}$ consists of the transformations which multiplicatively scale the support of a lottery.

for all $x \in X$ and $\omega, \omega' \in \mathcal{M}$. Every example in Section 2.1.1, 2.1.2, and 2.1.5 is of this form, as are often families of transformations which depend only on a single real scalar, such as mixing under various weights with a fixed act or lottery (e.g. worst-independence, Section 2.1.3). In such cases, we refer to $\mathcal{M}$ as a **commutative family**.

Let $R \subseteq X \times X$ be an arbitrary binary relation. We define the **$\mathcal{M}$-closure** of R, denoted $R_{\mathcal{M}}$ via:

$$x \; R_{\mathcal{M}} \; y \quad \Longleftrightarrow \quad \omega(x) \; R \; \omega(y)$$

for some $\omega \in \mathcal{M}$. Since we have assumed that the identity function id $\in \mathcal{M}$, by setting $\omega = \text{id}$ we obtain that $R \subseteq R_{\mathcal{M}}$.

Consider now the data $\langle \succsim^R, \succ^R \rangle$. Our first main result says that, when $\mathcal{M}$ is a commutative family, the data $\langle \succsim^R, \succ^R \rangle$ are rationalizable by an $\mathcal{M}$-invariant preference relation if and only if its $\mathcal{M}$-closure is acyclic.

Theorem 1. *Let X be a set, and $\mathcal{M}$ an arbitrary family of commuting transformations. Then $\langle \succsim^R, \succ^R \rangle$ is rationalilzable by an $\mathcal{M}$-invariant preference relation if and only if $\left\langle \succsim^R_{\mathcal{M}}, \succ^R_{\mathcal{M}} \right\rangle$ is acyclic.*

Note that when $\mathcal{M} = \{\text{id}\}$, $\mathcal{M}$ is clearly a commutative family, and that *every* preference $\succsim$ is trivially $\mathcal{M}$-invariant. Thus Theorem 1 strictly subsumes the classical characterization of Richter (1966).

3.1 Price-Consumption Data

In this section, we consider the special case in which our relations $\langle \succsim^R, \succ^R \rangle$ are generated by some price-consumption data set $\{(p_1, x_1), \ldots, (p_K, x_K)\}$. Here, we assume $\langle \succsim^R, \succ^R \rangle$ are the revealed preference relations associated with this data set, via:

$$x \succsim^R y \quad \Longleftrightarrow \quad x \,=\, x^k \text{ for some } k, \text{ and } p_k \cdot x \geq p_k \cdot y$$

(respectively $\succ^R$ and $>$). We show that for various common choices of $\mathcal{M}$, the acyclicity of $\langle \succsim^R_{\mathcal{M}}, \succ^R_{\mathcal{M}} \rangle$ straightforwardly reduces to the standard, model-specific revealed preference axioms.

3.1.1 Quasilinearity

Suppose that $X = Y \times \mathbb{R}_+$, and $\mathcal{M}$ consists of all transformations of the form $(y, t) \mapsto (y, t + \alpha)$ for $\alpha \geq 0$. Let $\langle \succsim^R, \succ^R \rangle$ be an arbitrary data set. Then the $\mathcal{M}$-closure $\langle \succsim^R_{\mathcal{M}}, \succ^R_{\mathcal{M}} \rangle$ is defined by:

$$(y, t) \succsim^R_{\mathcal{M}} (y', t') \quad \Longleftrightarrow \quad (y, t + \alpha) \succsim^R (y', t' + \alpha)$$

for some $\alpha \geq 0$, with a similar definition for $\succ^R_{\mathcal{M}}$. Suppose now that $Y = \mathbb{R}^{L-1}_+$, and $\langle \succsim^R, \succ^R \rangle$ is the revealed preference relation arising from some price-consumption data set; without loss of generality, we normalize the each to $p_k = (\tilde{p}_k, 1)$. Then a $\langle \succsim^R_{\mathcal{M}}, \succ^R_{\mathcal{M}} \rangle$ cycle is equivalent to the existence of $(y_{k_0}, t_0), \ldots, (y_{k_N}, t_N) \in X$, and $\alpha_0, \ldots, \alpha_N \geq 0$ such that:

$$\begin{aligned}
p_{k_0} \cdot (y_{k_0}, t_{k_0}) &\geq p_{k_0} \cdot (y_{k_1}, t_1 + \alpha_0) = p_{k_0} \cdot (y_{k_1}, t_{k_1} + \alpha_0 - \alpha_1) \\
p_{k_1} \cdot (y_{k_1}, t_{k_1}) &\geq p_{k_1} \cdot (y_{k_2}, t_2 + \alpha_1) = p_{k_1} \cdot (y_{k_2}, t_{k_2} + \alpha_1 - \alpha_2) \\
&\vdots \\
p_{k_N} \cdot (y_{k_N}, t_{k_N}) &> p_{k_N} \cdot (y_{k_0}, t_0 + \alpha_N) = p_{k_N} \cdot (y_{k_0}, t_{k_0} + \alpha_N - \alpha_0)
\end{aligned} \tag{1}$$

where $t_{k_i} = t_i + \alpha_i$ for all $i = 1, \ldots N$.[11] Summing over (1):

$$\sum_{i=0}^{N} \tilde{p}_{k_i} \cdot (y_{k_{i+1}} - y_{k_i}) < 0,$$

which precisely corresponds precisely to a negative cycle à la Brown and Calsamiglia (2007).[12]

3.1.2 Homotheticity

Let X be a cone in a real vector space, and let $\mathcal{M}$ consist of all transformations of the form $x \mapsto \alpha x$, for $\alpha > 0$. The particular case of $X = \mathbb{R}^n_+$ is treated in Chambers and Echenique (2016), Theorem 4.2, but we reproduce the ideas here.

Here, the $\mathcal{M}$-closure of the data set $\langle \succsim^R, \succ^R \rangle$ is given by:

$$x \succsim^R_{\mathcal{M}} y \quad \Longleftrightarrow \quad \alpha x \succsim^R \alpha y$$

for some $\alpha > 0$, with a similar definition for $\succ^R_{\mathcal{M}}$. In Chambers and Echenique (2016), $\langle \succsim^R_{\mathcal{M}}, \succ^R_{\mathcal{M}} \rangle$ is referred to as $\langle \succeq^H, \succ^H \rangle$. The $\mathcal{M}$-closure is acyclic if and only if there do not exist $x_0, \ldots, x_N \in X$ and $\alpha_0, \ldots, \alpha_N > 0$ such that:

$$\begin{aligned}
\alpha_0 x_0 &\succsim^R \alpha_0 x_1 \\
\alpha_1 x_1 &\succsim^R \alpha_1 x_2 \\
&\vdots \\
\alpha_N x_N &\succ^R \alpha_N x_0.
\end{aligned}$$

Suppose again that $\langle \succsim^R, \succ^R \rangle$ is the revealed preference relation arising from some set of price-consumption observations; without loss of generality, we normalize each price so $p_k \cdot x_k = 1$. Then (2) is equivalent to the existence of

[11] In other words, $x \succsim^R_{\mathcal{M}} y$ if and only if there is some fixed translation along the numeraire axis that brings x equal to some chosen x_k, and which leaves y within the budget defined by p_k and x_k.

[12] Here, the i indices are understood to satisfy $N + 1 \equiv 0$.

$x_{k_0}, \ldots, x_K \in X$ and $\alpha_0, \ldots, \alpha_K > 0$ such that:

$$\begin{aligned}
p_{k_0} \cdot x_{k_0} &\geq p_{k_0} \cdot (\alpha_0 x_1) = p_{k_0} \cdot \left(\frac{\alpha_0 x_{k_1}}{\alpha_1}\right) \\
p_{k_1} \cdot x_{k_1} &\geq p_{k_0} \cdot (\alpha_1 x_2) = p_{k_0} \cdot \left(\frac{\alpha_1 x_{k_2}}{\alpha_2}\right) \\
&\vdots \\
p_{k_N} \cdot x_{k_N} &> p_{k_N} \cdot (\alpha_N x_0) = p_{k_N} \cdot \left(\frac{\alpha_N x_{k_0}}{\alpha_0}\right)
\end{aligned} \tag{2}$$

where $\alpha_i x_i = x_{k_i}$ for all $i = 1, \ldots, N$. Taking products of (2) leads to the cancellations of all α_i/α_{i+1} terms, resulting in:

$$\prod_{i=0}^{N} p_{k_i} x_{k_{i+1}} < 1,$$

which is precisely a violation of the homothetic axiom of revealed preference of Varian (1983). As mentioned previously, in the case of general $\langle \succsim^R, \succ^R \rangle$, not necessarily arising from price-consumption observations, Demuynck (2009) obtains a similar characterization, in the special case of monotone and homothetic preferences, via a different approach.

3.1.3 Translation-Invariance

Let S be some finite set of states of the world, and let $X = \mathbb{R}^S$ denote the space of portfolios of Arrow securities. Let $\mathcal{M}$ denote the collection of transformations of the form $x \mapsto x + \vec{\alpha}$, where $\vec{\alpha} := (\alpha, \ldots, \alpha)$, for each $\alpha \in \mathbb{R}$. We refer to an $\mathcal{M}$-invariant preference as *translation invariant*. By Theorem 1, the data $\langle \succsim^R, \succ^R \rangle$ are rationalizably by a translation-invariant preference if and only if there does not exist $x_0, \ldots, x_N \in X$ and $\alpha_0, \ldots, \alpha_N \in \mathbb{R}$ such that:

$$\begin{aligned}
p_{k_0} \cdot x_{k_0} &\geq p_{k_0} \cdot \big(x_1 + \vec{\alpha}_0\big) = p_{k_0} \cdot \big(x_{k_1} + \vec{\alpha}_0 - \vec{\alpha}_1\big) \\
p_{k_1} \cdot x_{k_1} &\geq p_{k_1} \cdot \big(x_2 + \vec{\alpha}_1)\big) = p_{k_1} \cdot \big(x_{k_2} + \vec{\alpha}_1 - \vec{\alpha}_2\big) \\
&\vdots \\
p_{k_N} \cdot x_{k_N} &\geq p_{k_N} \cdot \big(x_0 + \vec{\alpha}_N\big) = p_{k_N} \cdot \big(x_{k_0} + \vec{\alpha}_N - \vec{\alpha}_0\big).
\end{aligned} \tag{3}$$

Summing over (3) we obtain:

$$\sum_{i=0}^{N} p_{k_i} \cdot (x_{k_{i+1}} - x_{k_i}) - \sum_{i=0}^{N} (\alpha_i - \alpha_{i+1}) \|p_{k_i}\|_1 < 0,$$

or, normalizing each p_{k_i} by $\|p_{k_i}\|_1$ without loss of generality:

$$\sum_{i=0}^{N} \frac{p_{k_i}}{\|p_{k_i}\|_1} \cdot (x_{k_{i+1}} - x_{k_i}) < 0,$$

precisely the same condition obtained in Chambers et al. (2016).[13]

3.2 Out of Sample Predictions

Suppose now we fix some $\langle \succsim^R, \succ^R \rangle$ such that $\langle \succsim^R_{\mathcal{M}}, \succ^R_{\mathcal{M}} \rangle$ is acyclic, and hence for which the set of $\mathcal{M}$-invariant rationalizating preferences is non-empty.[14] A natural question is to determine the set of *out-of-sample predictions* imposed by $\mathcal{M}$-invariance. More formally, we seek the set of relations which do *not* belong to $\langle \succsim^R, \succ^R \rangle$ but which nonetheless belong to every $\mathcal{M}$-invariant rationalizing preference.

In the classical case, when $\mathcal{M} = \{\text{id}\}$, the only out of sample predictions that can be generated are given by the transitive closure: if we observe that x is preferred to y, and y is preferred to z, then we can conclude x is preferable to z, even if this not directly observed. The Dushnik-Miller theorem (Dushnik and Miller 1941) establishes that deductions drawn by (possibly repeating) this form of inference are, in fact, the *only* relations unanimously agreed upon by the set of rationalizations.

In light of Theorem 1, a natural conjecture would be that the $\mathcal{M}$-closure of $\succsim^R$ (at least when it is transitive) is the intersection of all $\succsim^R$'s $\mathcal{M}$-invariant extensions, and hence characterizes all possible out-of-sample predictions generated by the model. Perhaps surprisingly, this turns out to be false, as the following example illustrates.

Example 2. Suppose X consists of a discrete-time model of dated rewards with two possible prizes, i.e. $X = \{a, b\} \times \mathbb{N}_0$, and define $\omega(i, n) = (i, n + 1)$ for $i \in \{a, b\}$.[15] Let $\mathcal{M} = \{\text{id}, \omega, \omega^2, \ldots\}$.[16] As such, here $\mathcal{M}$-invariance corresponds to *stationarity* of the preference, in the sense of Fishburn and Rubinstein (1982). Finally, let $\succsim^R$ contain only relations of the form:

$$(a, n) \succ^R (b, n + 1)$$

and, for all $n \geq 1$,

$$(a, n) \succ^R (b, n - 1).$$

Suppose an agent has preferences rationalizing $\succsim^R$. Relative to receiving b at a given date, such an agent always prefers to either receive a a day earlier, or a day later.

What can be said about the agent's preferences between the prizes today, i.e. between $(a, 0)$ and $(b, 0)$? First, note that $\succsim^R$ is vacuously transitive. Thus if we are unwilling to assume stationarity, nothing could be said: there exist rationalizing preferences preferring $(a, 0)$ to $(b, 0)$ and vice-versa.

[13] Economically, this normalization may be regarded as treating *bonds* as a numeraire commodity.

[14] As a consequence, $\succ^R$ must be the asymmetric component of $\succsim^R$, hence we will speak of the single data relation $\succsim^R$.

[15] We use the notation $\mathbb{N}_0$ to denote the set of non-negative integers.

[16] We write ω^k to denote the k-fold composite $\omega \circ \cdots \omega$.

In fact, $\succsim^R$ is stationary partial order, i.e. is $\mathcal{M}$-invariant. However, in spite of $(a, 0)$ and $(b, 0)$ being $\succsim^R$-incomparable, every *stationary* rationalization actually strictly prefers $(a, 0)$ to $(b, 0)$. To see this, suppose for sake of contradiction such a preference $\succeq^*$ ranked $(b, 0) \succeq^* (a, 0)$. Then

$$(b, 0) \succeq^* (a, 0) \succ^* (b, 1) \succeq^* (a, 1) \succ^* (b, 0),$$

a contradiction.[17] ■

By considering only $\mathcal{M}$-invariant rationalizations, we obtain a smaller rationalizing set, relative to the classical case, which in turn yields a larger intersection. Thus while classically, the out-of-sample implications of rationality are exceedingly limited, when one considers richer models of agents behavior, we obtain commensurately richer, more nuanced, out-of-sample predictions. We return to this in Section 5, where we provide a characterization and a generalization of the Dushnik-Miller theorem for the $\mathcal{M}$-invariant setting.

4 The General Case

4.1 Overview

In the preceding section, Theorem 1 showed that when the transformations in $\mathcal{M}$ commute with each other, a simple generalization of Richter (1966)'s acyclicity condition characterizes rationalizability by an $\mathcal{M}$-invariant preference. However, when $\mathcal{M}$ is not a commutative family, this conclusion fails. We return to the example from the introduction.

Example 3. Let $Z = \{x, x'y, y'\}$ be a set of prizes, and suppose X consists of all infinite horizon consumption streams taking values in Z, i.e. $X = Z^{\mathbb{N}}$. Let $\mathcal{M}$ consist of all finite compositions of the transformations $\{\omega_i\}_{i \in Z}$ which append the prize i to the start of a consumption stream. Here, $\mathcal{M}$-invariance corresponds to the stationarity of a preference in the sense of Koopmans (1960).

Let $\sigma, \sigma' \in X$ be arbitrary consumption streams, and suppose $\succsim^R$ is given by:

$$\begin{aligned} (x', \sigma_1, \sigma_2, \ldots) &\succ^R (x, \sigma'_1, \sigma'_2, \ldots) \\ (x, \sigma_1, \sigma_2, \ldots) &\succ^R (x', \sigma'_1, \sigma'_2, \ldots), \end{aligned} \tag{4}$$

and

$$\begin{aligned} (y', \sigma'_1, \sigma'_2, \ldots) &\succ^R (y, \sigma_1, \sigma_2, \ldots) \\ (y, \sigma'_1, \sigma'_2, \ldots) &\succ^R (y', \sigma_1, \sigma_2, \ldots). \end{aligned} \tag{5}$$

This relation is vacuously transitive, as is its $\mathcal{M}$-closure. However, it cannot be rationalized by any stationary preference. To see this, note that from the first

[17] The third relation follows from $\mathcal{M}$-invariance, and the second and fourth from $\succeq^*$ rationalizing $\succsim^R$.

two relations, any stationary extension $\succeq^*$ must rank $\sigma \succ^* \sigma'$, as if it instead ranked $\sigma' \succeq^* \sigma$, then:

$$\omega_{x'}(\sigma) \succ^* \omega_x(\sigma') \succeq^* \omega_x(\sigma) \succ^* \omega_{x'}(\sigma') \succeq^* \omega_{x'}(\sigma),$$

violating transitivity. However, by an identical argument, the latter two relations similarly imply that any stationary extension must rank $\sigma' \succ^* \sigma$. Thus when $\mathcal{M}$ is not commutative, acyclicity of the $\mathcal{M}$-closure no longer guarantees rationalizability. ■

The essence of Example 3 is that extending an incomplete relation while preserving $\mathcal{M}$-invariance is far more delicate than extending an acyclic binary relation. Suppose $x, y \in X$ are $\succsim^R$-unrelated, and we wish to manually construct an $\mathcal{M}$-invariant extension $\succeq^*$. To do so, we must posit how $\succeq^*$ ranks x and y. However, any assignment of ranking between x and y in general has infinitely many knock-on effects: it determines the ranking between $\omega(x)$ and $\omega(y)$ for every $\omega \in \mathcal{M}$. Depending on the specifics of the transforms in $\mathcal{M}$, we may inadvertently end up creating cycles by involving multiple different 'translates' of our added relation.[18]

This creates a set of constraints on how any relation can extend $\succsim^R$. Crucially, when $\mathcal{M}$ is a commutative family, Theorem 1 shows that if these 'indirect constraints' rule out any extension ranking x over y or ranking y over x, then $\langle \succsim^R_{\mathcal{M}}, \succ^R_{\mathcal{M}} \rangle$ must contain a cycle. However, when the transforms in $\mathcal{M}$ do not commute, this need no longer be the case, as Example 3 shows.

In this section, we introduce a strengthening of the transitive closure. Unlike the transitive closure, the natural domain of our strengthened notion of closure is not on the data $\langle \succsim^R, \succ^R \rangle$ itself, but rather the collections of 'indirect constraints' jointly imposed by the data and the structure of $\mathcal{M}$. We show that our generalized notion of closure characterizes both (i) the existence of $\mathcal{M}$-invariant rationalizations, for any choice of X, $\mathcal{M}$, and $\langle \succsim^R, \succ^R \rangle$, and (ii) the out-of-sample predictions generated by $\mathcal{M}$-invariance, just as the transitive closure does in the classical setting of $\mathcal{M} = \{\text{id}\}$.

[18] In Example 2, a cycle was formed by adding $(b, 0) \succeq^* (a, 0)$ which relied on both that relation, and its 'translate' $(b, 1) \succeq^* (a, 1)$. Similarly, in Example 3, adding $\sigma' \succeq^* \sigma$ lead to a cycle involving $\omega_x(\sigma') \succeq \omega_x(\sigma)$ and $\omega_w(\sigma') \succeq^* \omega_w(\sigma)$.

4.2 Forcing Collections and Constraint Sets

Let $\omega_0, \ldots, \omega_N \in \mathcal{M}$, and $x_0, y_0, \ldots, x_N, y_N \in X$ be a sequence of $\succsim^R$-unrelated pairs (i.e. x_i and y_i are $\succsim^R$-unrelated). We term a collection of relations:

$$\begin{aligned} \omega_0(x_0) &\succsim^R_{\mathsf{T}} \omega_1(y_1) \\ \omega_1(x_1) &\succsim^R_{\mathsf{T}} \omega_2(y_2) \\ &\vdots \\ \omega_{N-1}(x_{N-1}) &\succsim^R_{\mathsf{T}} \omega_N(y_N) \\ \omega_N(x_N) &\succsim^R_{\mathsf{T}} \omega_0(y_0), \end{aligned} \tag{$*$}$$

a **forcing collection.** If any of the relations $\succsim^R_{\mathsf{T}}$ also belongs to $\succ^R_{\mathsf{T}}$, we say $(*)$ forms a **strong forcing collection.**[19]

Any (strong) forcing collection implies joint restrictions on the possible comparisons any $\mathcal{M}$-invariant rationalization may make between the x_i and y_i. To formalize this, for each $(x, y) \in X \times X$, we introduce a pair of boolean variables (i.e. taking values in $\{\top, \bot\}$),

$$[\mathtt{x} \succeq \mathtt{y}] \quad \text{and} \quad [\mathtt{x} \succ \mathtt{y}].$$

Let $\mathcal{V}$ denote the set of all such variables. A set of boolean variables $C \subseteq \mathcal{V}$ forms a **constraint set** for the forcing collection $(*)$ if:

(i) For all $0 \le i \le N$, exactly one of $[\mathtt{y}_i \succeq \mathtt{x}_i]$ or $[\mathtt{y}_i \succ \mathtt{x}_i]$ belongs to C, and every variable in C corresponds to a unique pair $\{x_i, y_i\}$.

(ii) If $(*)$ is not a strong forcing collection, then C contains at least one variable of the form $[\mathtt{y}_i \succ \mathtt{x}_i]$.

Intuitively, a constraint set C encodes sets of relations that *cannot be simultaneously satisfied* by any $\mathcal{M}$-invariant rationalization. For example, the constraint set:

$$C = \big\{[\mathtt{y}_0 \succ \mathtt{x}_0], [\mathtt{y}_1, \succeq \mathtt{x}_1], \ldots, [\mathtt{y}_N \succeq \mathtt{x}_N]\big\} \tag{6}$$

reflects the observation that in light of $(*)$, adding the relations $y_0 \succ x_0$ and $y_n \succeq x_n$ (for $1 \le i \le N$) would create a cycle. In logical terms, C may be thought of as encoding the clause:

$$\neg[\mathtt{y}_0 \succ \mathtt{x}_0] \vee \neg[\mathtt{y}_1 \succeq \mathtt{x}_1] \vee \cdots \vee \neg[\mathtt{y}_N \succeq \mathtt{x}_N]. \tag{7}$$

Any complete binary relation on X defines an assignment of values $\top$ or $\bot$ to each variable in $\mathcal{V}$ in a natural way. Thus C encodes a logical formula which any

[19]When $N = 0$ we also speak of $\omega_0(x_0) \succsim^R_{\mathsf{T}} \omega_0(y_0)$ (resp. $\omega_0(x_0) \succ^R_{\mathsf{T}} \omega_0(y_0)$) as a valid forcing collection (resp. strong forcing collection), with constraint set $[\mathtt{y} \succ \mathtt{x}]$ (resp. $[\mathtt{y} \succeq \mathtt{x}]$). As always, these may be interpreted as encoding the observation that no $\mathcal{M}$-invariant rationalization $\succeq^*$ can rank $y \succ^* x$ (resp. $y \succeq^* x$).

$\mathcal{M}$-invariant rationalization must evaluate to $\top$. For purposes of exposition, it is simpler to work in terms of the constraint set (6) rather than its equivalent **clausal representation** (7).

Finally, it is straightforward to see that our notion of a constraint set generalizes the transitive closure: suppose $x \succsim^R y$ and $y \succsim^R z$. Then clearly $x \succsim^R_{\mathsf{T}} z$, which defines a forcing collection (with $N = 0$). The associated constraint set is $[\mathtt{z} \succ \mathtt{x}]$, which encodes that no extension $\succeq^*$ can satisfy $z \succ^* x$, or equivalently, that every extension must satisfy $x \succeq^* z$.

4.3 Collapses and the Transitive Closure

The transitive closure of a binary relation $R \subseteq X \times X$ may be viewed as arising from a particular form of (partial) binary operation on pairs of elements in $X \times X$. Given (x, y) and (y, z) belong to R, the pair (x, z) formed by 'cancelling out' the common element y is a first-order implication of transitivity. We denote the set of such implications by R^1. Similarly, the second-order implications of transitivity are deduced by applying this 'cancellation' operation to pairs of elements of $R \cup R^1$, and so forth. The transitive closure then simply is the union of these implications, i.e. $R_{\mathsf{T}} := \cup_n R^n$.

We now introduce an mechanically similar cancellation operation that we term the 'collapse.' Here, however, the collapse operates on pairs of *constraint sets*, rather than pairs of elements in $X \times X$.[20]

Let $\mathcal{C} \subseteq 2^{\mathcal{V}}$ denote the collection of all constraint sets generated by the primitives $\langle \succsim^R, \succ^R \rangle$ and $\mathcal{M}$. Formally, let $C, C' \in \mathcal{C}$ and $D \subseteq 2^{\mathcal{V}}$. We say that D is a **collapse** of C and C' if there exists variables $L \in C$, $L' \in C'$ such that:

(i) D is given by the union of C and C', minus L and L', i.e.:

$$D = \big(C \setminus \{L\}\big) \;\cup\; \big(C' \setminus \{L'\}\big).$$

(ii) Either L and L' are of the form:

$$L = \big[\omega(\mathtt{y}) \succeq \omega(\mathtt{x})\big] \quad \text{and} \quad L' = \big[\omega'(\mathtt{x}) \succeq \omega'(\mathtt{y})\big]$$

or

$$L = \big[\omega(\mathtt{y}) \succeq \omega(\mathtt{x})\big] \quad \text{and} \quad L' = \big[\omega'(\mathtt{x}) \succ \omega'(\mathtt{y})\big]$$

for some $x, y \in X$ and $\omega, \omega' \in \mathcal{M}$.

[20] The need to consider sets of constraints simultaneously is a consequence $\mathcal{M}$-invariance introducing the 'indirect constraints' as discussed above. Informally, the comparison of working with pairs in $X \times X$ versus pairs of constraint sets is somewhat analogous to the situation of solving systems of linear equations. When a system is suitably decomposable, it may be possible to solve equation-by-equation, reducing the machinery needed to that of solving a single linear equation at a time. However, generally such a decomposition is impossible, necessitating an approach that handles the whole system simultaneously.

In comparison, when forming the transitive closure, we cancel the y in (x, y) and (y, z) to form (x, z); here, cancellation is across 'clashing' pairs of relations, modulo the transformation ω and ω'.

Let $\mathcal{C}^0 \equiv \mathcal{C}$, and inductively define:

$$\mathcal{C}^n = \left\{ D \subseteq 2^{\mathcal{V}} : D \text{ is a collapse of } C, C' \in \cup_{i=0}^{n-1} \mathcal{C}^n \right\}.$$

Finally, define:

$$\mathcal{C}^* = \bigcup_{n=0}^{\infty} \mathcal{C}^n.$$

We say that $\langle \succsim^R, \succ^R \rangle$ is $\mathcal{M}$-**acyclic** if and only if $\varnothing \notin \mathcal{C}^*$. For example, in Example 3, we have two (strong) forcing collections, (4) and (5), with constraint sets $[\sigma \succeq \sigma']$ and $[\sigma' \succeq \sigma]$ respectively. The collapse of these two constraint sets is $\varnothing$ and hence $\langle \succsim^R, \succ^R \rangle$ is not $\mathcal{M}$-acyclic.

To motivate this terminology, note that for a binary relation R, a cycle is a collection of pairs $(x_0, x_1), (x_1, x_2), \ldots, (x_N, x_0) \in R$ (with at least one pair in the asymmetric component of R) such that every x_i is a 'cancelled' element between two pairs. In comparison, here $\langle \succsim^R, \succ^R \rangle$ contains a 'cycle' if there is a finite collection of constraint sets and sequence of collapses such that every variable, appearing in any constraint, is collapsed away at some point in the sequence.

The next theorem is the primary result of this paper. It says that $\mathcal{M}$-acyclicity characterizes the existence of an $\mathcal{M}$-invariant rationalizing preference, no matter how large or complex X, $\mathcal{M}$ or $\langle \succsim^R, \succ^R \rangle$ are.

Theorem 2. *The data $\langle \succsim^R, \succ^R \rangle$ are $\mathcal{M}$-acyclic if and only if they are rationalizable by an $\mathcal{M}$-invariant preference.*

To illustrate Theorem 2, consider again the relation $\succsim^R$ from Example 3. There, (4) defines a forcing collection with constraint set $C = \left\{ [\sigma' \succeq \sigma] \right\}$, and (5) a forcing collection with constraint set $C' = \left\{ [\sigma \succeq \sigma'] \right\}$. Thus $\varnothing$ is precisely the collapse of C and C' and hence we may think of the pair C and C' as defining an '$\mathcal{M}$-cycle.'

5 Out-of-Sample Predictions

When $\mathcal{M} = \{\text{id}\}$, every ($\mathcal{M}$-invariant) rationalizing preference $\succeq^*$ ranks $x \succeq^* y$ if and only if $x \succsim^R_{\mathsf{T}} y$. However, as illustrated by Example 2, when $\mathcal{M}$ is richer, so too are the set of counterfactual predictions generated by the class of $\mathcal{M}$-invariant preferences. Moreover, Example 2 shows that the set of such predictions is richer than either the transitive, or $\mathcal{M}$-invariant closure (or, indeed, any iterated application of these operations).

It turns out that the set of out-of-sample predictions generated by the $\mathcal{M}$-invariant rationalizations of an $\mathcal{M}$-acyclic relation $\succsim^R$ are straightforwardly described in terms of collapses.

Theorem 3. *Suppose $\langle \succsim^R, \succ^R \rangle$ is rationalizable by an $\mathcal{M}$-invariant preference. Then $x \succeq^* y$ (resp. $x \succ^* y$), for every such rationalization $\succeq^*$, if and only if:*

$$\big\{[\boldsymbol{y} \succ \boldsymbol{x}]\big\} \in \mathcal{C}^* \quad \textit{(resp. } \big\{[\boldsymbol{y} \succeq \boldsymbol{x}]\big\} \in \mathcal{C}^* \textit{).}$$

It is worth again stressing the connection between Theorem 3 and the classical case involving the transitive closure. Classically, if $x \succsim^R_{\mathsf{T}} y$, then there exist a collection of pairs $(x, x_2), (x_2, x_3), \ldots, (x_N, y) \in \succsim^R$, such that when all feasible 'cancellations' are carried out, the only remaining terms are the unmatched elements x and y. Analogously, Theorem 3 simply says that an identical result holds true for collapses of constraint sets: x is preferred to y by every $\mathcal{M}$-invariant rationalization if and only if there is a collection of constraint sets such that $[\mathtt{x} \succeq \mathtt{y}]$ is the only element not collapsed away.

Appendix

A Proof of Theorem 1

Say that a relation $\succeq \subseteq X \times X$ is **weakly $\mathcal{M}$-invariant** if:

$$x \succeq y \implies \omega(x) \succeq \omega(y)$$

and

$$x \succ y \implies \omega(x) \succ \omega(y),$$

for all $x, y \in X$ and $\omega \in \mathcal{M}$.

Lemma 1. *Suppose that an acyclic relation $\succsim^R$ is weakly $\mathcal{M}$-invariant. Then so is its transitive closure, $\succsim^R_{\mathsf{T}}$.*

Proof. First, let $x, y \in X$ such that $x \succsim^R_{\mathsf{T}} y$. Then there exists $x_1, \ldots, x_K \in X$, $K \geq 2$, such that $x = x_1 \succsim^R \cdots \succsim^R x_K = y$. By $\mathcal{M}$-invariance of $\succsim^R$, for every $\omega \in \mathcal{M}$, we also have that $\omega(x) = \omega(x_1) \succsim^R \cdots \succsim^R \omega(x_k) = \omega(y)$, hence $\omega(x) \succsim^R_{\mathsf{T}} \omega(y)$ as desired.

Now, suppose additionally that it is not the case that $y \succsim^R_{\mathsf{T}} x$. We want to show that it is not the case that $\omega(y) \succsim^R_{\mathsf{T}} \omega(x)$ for any $\omega \in \mathcal{M}$.

Since $x \succsim^R_{\mathsf{T}} y$ but not $y \succsim^R_{\mathsf{T}} x$, it follows that there exist $x_1, \ldots, x_K \in X$, $K \geq 2$ such that $x = x_1 \succsim^R \cdots \succsim^R x_K = y$, where for some $1 \leq i \leq K-1$, we have $x_i \succ x_{i+1}$. Consequently, for any $\omega \in \mathcal{M}$, by $\mathcal{M}$-invariance, we must also have $\omega(x) = \omega(x_1) \succsim^R \cdots \succsim^R \omega(x_K) = \omega(y)$, where $\omega(x_i) \succ \omega(x_{i+1})$. Because $\succsim^R$ is acyclic, there can be such sequence connecting $\omega(y)$ back to $\omega(x)$, hence $\omega(y) \succsim^R_{\mathsf{T}} \omega(x)$ must be false. Since $\omega \in \mathcal{M}$ was arbitrary, the result follows. □

Lemma 2. *Suppose $\mathcal{M}$ is a commutative family. Then every weakly $\mathcal{M}$-invariant preorder has an $\mathcal{M}$-invariant preorder extension.*

Proof. Let $\succsim^R$ be a weakly $\mathcal{M}$-invariant preorder. Define $\succeq$ via $x \succeq y$ if and only if there exists $\omega_1, \ldots, \omega_K \in \mathcal{M}$ such that $(\omega_1 \circ \ldots \omega_K)(x) \succsim^R (\omega_1 \circ \ldots \omega_K)(y)$.

Since the identity function id $\in \mathcal{M}$, we have immediately that $\succsim^R \subseteq \succeq$. Suppose, now, that $x \succ^R y$ and, for purposes of contradiction that additionally $y \succeq x$. Then there exist $\omega_1, \ldots, \omega_K \in \mathcal{M}$ for which $(\omega_1 \circ \ldots \omega_K)(y) \succsim^R (\omega_1 \circ \ldots \omega_K)(x)$. Since $\succsim^R$ is weakly $\mathcal{M}$-invariant, this implies that both: $(\omega_1 \circ \ldots \omega_K)(y) \succsim^R (\omega_1 \circ \ldots \omega_K)(x)$ and $(\omega_1 \circ \ldots \omega_K)(x) \succ^R (\omega_1 \circ \ldots \omega_K)(y)$, a contradiction. Thus $\succsim^R \subseteq \succeq$ as well, and hence $\succeq'$ defines an extension of $\succsim^R$.

We now claim that $\succeq$ is transitive. Suppose that $x \succeq y \succeq z$. As $x \succeq y$, there exist $\omega_1, \ldots, \omega_K \in \mathcal{M}$ for which $(\omega_1 \circ \ldots \circ \omega_K)(x) \succsim^R (\omega_1 \circ \ldots \omega_K)(y)$. Similarly, as $y \succeq z$, there exist $\omega'_1, \ldots, \omega'_L \in \mathcal{M}$ for which $(\omega'_1 \circ \ldots \omega'_L)(y) \succeq (\omega'_1 \circ \ldots \omega'_L)(z)$. By the weak $\mathcal{M}$-invariance of $\succsim^R$ and by commutativity of $\mathcal{M}$, we may conclude that $(\omega_1 \circ \ldots \circ \omega_K \circ \omega'_1 \circ \ldots \omega'_L)(x) \succsim^R (\omega_1 \circ \ldots \circ \omega_K \circ \omega'_1 \circ \ldots \omega'_L)(y)$ and

$(\omega_1 \circ \ldots \circ \omega_K \circ \omega_1' \circ \ldots \omega_L')(y) \succsim^R (\omega_1 \circ \ldots \circ \omega_K \circ \omega_1' \circ \ldots \omega_L')(z)$, so that $(\omega_1 \circ \ldots \circ \omega_K \circ \omega_1' \circ \ldots \omega_L')(x) \succsim^R (\omega_1 \circ \ldots \circ \omega_K \circ \omega_1' \circ \ldots \omega_L')(z)$, and hence $x \succeq z$ by the transitivity of $\succsim^R$.

We claim that $\succeq$ is $\mathcal{M}$-invariant. First we show that it is weakly $\mathcal{M}$-invariant. Suppose that $x \succeq y$ and let $\omega \in \mathcal{M}$. Then there are $\omega_1, \ldots, \omega_K \in M$ for which $(\omega_1 \circ \ldots \omega_K)(x) \succsim^R (\omega_1 \circ \ldots \omega_K)(y)$. Since $\mathcal{M}$ is a commutative family and by the weak $\mathcal{M}$-invariance of $\succsim^R$, we have $(\omega_1 \circ \ldots \omega_K)\big(\omega(x)\big) \succeq (\omega_1 \circ \ldots \omega_K)\big(\omega(y)\big)$, and hence $\omega(x) \succeq \omega(y)$. Suppose additionally that $x \succ y$ and, for purposes of contradiction that for some $\omega \in \mathcal{M}$, $\omega(y) \succeq \omega(x)$. Then there exists $\omega_1', \ldots \omega_L' \in \mathcal{M}$ for which $(\omega_1' \circ \ldots \omega_L')\big(\omega(y)\big) \succsim^R (\omega_1' \circ \ldots \omega_L')\big(\omega(x)\big)$, which by definition implies that $y \succeq x$, a contradiction. This establishes that $\succeq$ is weakly $\mathcal{M}$-invariant.

Finally, to show that $\succeq$ is fully $\mathcal{M}$-invariant, suppose that $\omega(x) \succeq \omega(y)$. Then there exist $\omega_1, \ldots, \omega_K \in \mathcal{M}$ for which $(\omega_1 \circ \ldots \omega_K \circ \omega)(x) \succsim^R (\omega_1 \circ \ldots \omega_K \circ \omega)(y)$, which implies $x \succeq y$. Suppose in addition $\omega(y) \succeq \omega(x)$ is false and, for purposes of contradiction, that $y \succeq x$. Then again there exist $\omega_1, \ldots, \omega_K$ for which $(\omega_1 \circ \ldots \omega_K)(y) \succeq (\omega_1 \circ \ldots \omega_K)(x)$. By weak $\mathcal{M}$-invariance of $\succsim^R$, $(\omega \circ \omega_1 \ldots \omega_K)(y) \succsim^R (\omega \circ \omega_1 \ldots \omega_K)(x)$. By commutativity of $\mathcal{M}$, $(\omega_1 \circ \ldots \omega_K)\big(\omega(y)\big) \succsim^R (\omega_1 \circ \ldots \omega_K)\big(\omega(x)\big)$, so that $\omega(y) \succeq \omega(x)$, a contradiction. The result follows. □

Lemma 3. *Let $\mathcal{M}$ be a commutative family. Let $\succeq$ be a weakly $\mathcal{M}$-invariant preorder, and $w, z \in X$ be $\succeq$-unrelated (and hence distinct) elements of X. Then there is an acyclic, weakly $\mathcal{M}$-invariant extension $\succeq'$ of $\succeq$ that renders w and z comparable.*

Proof. By appeal to commutativity, any finite string of compositions of functions in $\mathcal{M}$ may be expressed (not necessarily uniquely) as:

$$f_1^{n_1} \circ f_2^{n_2} \circ \cdots \circ f_L^{n_L},$$

where $\{f_1, \ldots, f_L\} \subseteq \mathcal{M}$, and $n_1, \ldots, n_L \in \mathbb{N}$.[21] Given such an expression, define $\mathbf{f} : \mathcal{M} \to \mathbb{N}_0$ as the unique function such that $f_l \mapsto n_l$ and $g \mapsto 0$ if and only if $g \notin \{f_1, \ldots, f_L\}$.

Suppose now, for sake of obtaining a contradiction, that no acyclic, weakly $\mathcal{M}$-invariant extension of $\succeq$ exists that compares w and z. Then every weakly $\mathcal{M}$-invariant binary relation that extends $\succeq$ and renders w and z comparable, contains some cycle; in particular, the minimal such extensions obtained either by adding $w \succ' z$ and $\mathbf{f}(w) \succ' \mathbf{f}(z)$ for all $\mathbf{f}$ associated with some finite composition of elements of $\mathcal{M}$, by adding $z \succ' w$ and all $\mathbf{f}(z) \succ' \mathbf{f}(w)$, or by adding $z \sim' w$ and all $\mathbf{f}(z) \sim' \mathbf{f}(w)$, must contain some cycle. Consider first $\succeq' = \succeq \cup \succeq^*$, where $\succeq^*$ contains all relations of the form $w \succ^* z$ and

[21] Note, however, that in general it will be impossible to guarantee a unique representation of this form. For example, suppose $f : X \to X$ is bijective, and $\{f, f^{-1}\} \subseteq \mathcal{M}$.

$\mathbf{f}(w) \succ^* \mathbf{f}(z)$ for all finite compositions of elements of $\mathcal{M}$, $\mathbf{f}$. It follows there exists a cycle composed of relations of two forms:

$$\begin{array}{rl} x \succeq \mathbf{a}^1(w) & \mathbf{a}^1(w) \succ^*_{\mathsf{T}} \mathbf{a}^1(z) \\ \mathbf{a}^1(z) \succeq \mathbf{a}^2(w) & \mathbf{a}^2(w) \succ^*_{\mathsf{T}} \mathbf{a}^2(z) \\ \vdots & \vdots \\ \mathbf{a}^{I-1}(z) \succeq \mathbf{a}^I(w) & \mathbf{a}^I(w) \succ^*_{\mathsf{T}} \mathbf{a}^I(z) \\ \mathbf{a}^I(z) \succeq x, & \end{array} \tag{8}$$

for some $x \in X$, where the left column consists of relations in $\succeq$ and the right sequences solely of relations in $\succeq' \setminus \succeq$. Note that $I \geq 2$, and without loss of generality, each $\mathbf{a}^i$ is distinct.[22]

Analogously, if $\succeq' = \succeq \cup \succeq^*$, where $\succeq^*$ contains all relations of the form $z \succ^* w$ and $\mathbf{f}(z) \succ^* \mathbf{f}(w)$ for finite compositions $\mathbf{f}$, then there exists a cycle of the form:

$$\begin{array}{rl} x' \succeq \mathbf{b}^1(z) & \mathbf{b}^1(z) \succ^*_{\mathsf{T}} \mathbf{b}^1(w) \\ \mathbf{b}^1(w) \succeq \mathbf{b}^2(z) & \mathbf{b}^2(z) \succ^*_{\mathsf{T}} \mathbf{b}^2(w) \\ \vdots & \vdots \\ \mathbf{b}^{J-1}(w) \succeq \mathbf{b}^J(z) & \mathbf{b}^J(z) \succ^*_{\mathsf{T}} \mathbf{b}^J(w) \\ \mathbf{b}^J(w) \succeq x', & \end{array}$$

for some $x' \in X$, where again the left column consists of relations in $\succeq$, the right solely of sequences of relations in $\succeq' \setminus \succeq$, $J \geq 2$, and each $\mathbf{b}^j$ unique.

Finally, suppose $\succeq' = \succeq \cup \succeq^*$, where $\succeq^*$ contains all relations of the form $z \sim^* w$ and $\mathbf{f}(z) \sim^* \mathbf{f}(w)$ for finite compositions $\mathbf{f}$. By hypothesis, there is a cycle of the form:

$$\begin{array}{rl} x'' \succeq \mathbf{c}^1(x_1) & \mathbf{c}^1(x_1) \sim^*_{\mathsf{T}} \mathbf{c}^1(y_1) \\ \mathbf{c}^1(y_1) \succeq \mathbf{c}^2(x_2) & \mathbf{c}^2(x_2) \sim^*_{\mathsf{T}} \mathbf{c}^2(y_2) \\ \vdots & \vdots \\ \mathbf{c}^{K-1}(y_{K-1}) \succeq \mathbf{c}^J(x_K) & \mathbf{c}^K(a_K) \sim^*_{\mathsf{T}} \mathbf{c}^J(y_K) \\ \mathbf{c}^J(y_K) \succeq x'', & \end{array} \tag{$*$}$$

where at least one relation in the left-hand column is strict, $K \geq 2$, each $\mathbf{c}^k$ is unique, and for all $k = 1, \ldots, K$, $\{x_k, y_k\} = \{w, z\}$.

Define:

$$\begin{aligned} \mathbf{p}^i &= \mathbf{a}^{i+1} - \mathbf{a}^i \\ \mathbf{q}^j &= \mathbf{b}^{j+1} - \mathbf{b}^j \\ \mathbf{r}^k &= \mathbf{c}^{k+1} - \mathbf{c}^k, \end{aligned}$$

[22] If $I = 1$, then we have $\mathbf{a}^1(z) \succeq x$ and $x \succeq \mathbf{a}^1(w)$, hence $\mathbf{a}^1(z) \succeq \mathbf{a}^1(w)$. Since $\succeq$ is $\mathcal{M}$-invariant, this would imply w and v are $\succeq$-related, which is false.

where we define indices $I+1, J+1, K+1 \equiv 1$. Note that each $\mathbf{p}^i, \mathbf{q}^j$, and $\mathbf{r}^k$ is not equal to the zero function $\mathbf{0}$, and:

$$\sum_{i=1}^{I} \mathbf{p}^i = \sum_{j=1}^{J} \mathbf{q}^j = \sum_{k=1}^{K} \mathbf{r}^k = \mathbf{0}.$$

Consider the sets:

$$\begin{aligned}
\tilde{A}_{wz} &= \{\mathbf{r}^k \mid y_k = w,\ x_{k+1} = z\} \\
\tilde{A}_{zw} &= \{\mathbf{r}^k \mid y_k = z,\ x_{k+1} = w\} \\
\tilde{A}_{ww} &= \{\mathbf{r}^k \mid y_k = w,\ x_{k+1} = w\} \\
\tilde{A}_{zz} &= \{\mathbf{r}^k \mid y_k = z,\ x_{k+1} = z\}.
\end{aligned}$$

Clearly these sets cover $\{\mathbf{r}^1, \ldots, \mathbf{r}^K\}$. Define:

$$\begin{aligned}
A_{wz} &= \tilde{A}_{wz} \\
A_{zw} &= \tilde{A}_{zw} \setminus \tilde{A}_{wz} \\
A_{ww} &= \tilde{A}_{ww} \setminus \tilde{A}_{zw} \setminus \tilde{A}_{wz} \\
A_{zz} &= \tilde{A}_{zz} \setminus \tilde{A}_{ww} \setminus \tilde{A}_{zw} \setminus \tilde{A}_{wz},
\end{aligned}$$

if these sets are non-empty, and if empty define them as $\{\mathbf{0}\}$. By hypothesis, at least some of the sets must contain non-zero elements. Note that each element of $\{\mathbf{r}^1, \ldots, \mathbf{r}^K\}$ is contained in exactly one set in the collection $\{A_{wz}, A_{zw}, A_{ww}, A_{zz}\}$. Let $\{\mathbf{s}^m_{wz}\}_{m=1}^{|A_{wz}|}$ (resp. $\{\mathbf{s}^m_{zw}\}_{m=1}^{|A_{zw}|}$, $\{\mathbf{s}^m_{ww}\}_{m=1}^{|A_{ww}|}$, and $\{\mathbf{s}^m_{zz}\}_{m=1}^{|A_{zz}|}$) denote enumerations of A_{wz} (resp. A_{zw}, A_{ww}, and A_{zz}).

We now establish a contradiction, by showing that $\succeq$ contains a cycle, contrary to our hypothesis that it is a preorder. Let $\bar{\mathbf{h}}$ denote a sufficiently large map $M \to \mathbb{N}_0$ with finite support.[23] We will consider two cases in turn.

Case 1: $|A_{wz}| + |A_{zw}| > 0$.

To build our cycle, we first define two chains in $\succeq$ which will prove important in our construction.[24] Let us refer to chain one as the sequence:

[23]Sufficiently in the sense only that each vector in the following sequence remain non-negative valued.

[24]The first chain indexes by $|A_{wz}|$ and the second indexes by $|A_{zw}|$; if either of these are zero, these chains are trivial.

$$
\begin{aligned}
\bar{\mathbf{h}}(z) &\succeq (\bar{\mathbf{h}} + \mathbf{p}^1)(w) \\
&\succeq (\bar{\mathbf{h}} + \mathbf{p}^1 + \mathbf{s}^1_{wz})(z) \\
&\qquad\vdots \\
&\succeq \left(\bar{\mathbf{h}} + |A_{wz}| \sum_{i=1}^{I} \mathbf{p}^i + I \sum_{m=1}^{|A_{wz}|} \mathbf{s}^m_{wz}\right)(z) \\
&\qquad\vdots \\
&\succeq \left(\bar{\mathbf{h}} + J\,|A_{wz}| \sum_{i=1}^{I} \mathbf{p}^i + IJ \sum_{m=1}^{|A_{wz}|} \mathbf{s}^m_{wz}\right)(z).
\end{aligned}
$$

The first part of this chain, up to $\left(\bar{\mathbf{h}} + |A_{wz}| \sum_{i=1}^{I} \mathbf{p}^i + I \sum_{m=1}^{|A_{wz}|} \mathbf{s}^m_{wz}\right)(z)$ is constructed as follows. For every $l = 1, \ldots, I|A_{wz}|$, every term of the form $(\bar{\mathbf{h}} + \ldots + \mathbf{p}^l)(w)$ is followed by a term of the form $(\bar{\mathbf{h}} + \ldots + \mathbf{p}^l + \mathbf{s}^l_{wz})(z)$, and for every $l = 0, \ldots, I|A_{wz}| - 1$, every term of the form $(\bar{\mathbf{h}} + \ldots + \mathbf{s}^l_{wz})(z)$ is followed by a term of the form $(\bar{\mathbf{h}} + \ldots + \mathbf{p}^l + \mathbf{s}^{l+1}_{wz})(w)$, where an l index on $\mathbf{p}$ is modulo I and on s_{wz} is modulo $|A_{wz}|$.

The second part of this chain, up through $\left(\bar{\mathbf{h}} + J\,|A_{wz}| \sum_{i=1}^{I} \mathbf{p}^i + IJ \sum_{m=1}^{|A_{wz}|} \mathbf{s}^m_{wz}\right)(z)$, follows by iterating the first $I|A_{wz}|$ steps an additional $|J| - 1$ times.

Similarly, we refer to chain two as the sequence of relations:

$$
\begin{aligned}
\bar{\mathbf{h}}(z) &\succeq (\bar{\mathbf{h}} + \mathbf{s}^1_{zw})(w) \\
&\succeq (\bar{\mathbf{h}} + \mathbf{s}^1_{zw} + \mathbf{q}^1)(z) \\
&\qquad\vdots \\
&\succeq \left(\bar{\mathbf{h}} + J \sum_{m=1}^{|A_{zw}|} \mathbf{s}^m_{wz} + |A_{zw}| \sum_{j=1}^{J} \mathbf{q}^j\right)(z) \\
&\qquad\vdots \\
&\succeq \left(\bar{\mathbf{h}} + IJ \sum_{m=1}^{|A_{zw}|} \mathbf{s}^m_{wz} + I\,|A_{zw}| \sum_{j=1}^{J} \mathbf{q}^j\right)(z).
\end{aligned}
$$

Appending these chains together then yields a chain:

$$
\bar{\mathbf{h}}(z) \succeq \cdots \succeq \left(\bar{\mathbf{h}} + I\,|A_{zw}| \sum_{j=1}^{J} \mathbf{q}^j + J\,|A_{wz}| \sum_{i=1}^{I} \mathbf{p}^i + IJ \sum_{m=1}^{|A_{wz}|} \mathbf{s}^m_{wz} + IJ \sum_{m=1}^{|A_{zw}|} \mathbf{s}^m_{wz}\right)(z).
$$

Consider now the following modification to this chain: immediately after the first instance of an $\mathbf{f}(z) \succeq \mathbf{g}(w)$ relation, apply IJ applications of each transformation in A_{ww}. Similarly, after the first $\mathbf{f}(w) \succeq \mathbf{g}(z)$ relation, insert IJ repetitions of each transformation in A_{zz}. The result is a chain:

$$\bar{\mathbf{h}}(z) \succeq \cdots \succeq \left(\bar{\mathbf{h}} + I\,|A_{zw}| \sum_{j=1}^{J} \mathbf{q}^j + J\,|A_{wz}| \sum_{i=1}^{I} \mathbf{p}^i + IJ \sum_{k=1}^{K} \mathbf{r}^k\right)(z).$$

However, since $\sum_i \mathbf{p}^i = \sum_j \mathbf{q}^j = \sum_k \mathbf{r}^k = \mathbf{0}$, this chain is in fact a cycle. Moreover, since every relation in the left-hand column of $(*)$ appears in this cycle, it contains at least one strict relation, contradicting the hypothesis that $\succ$ is a preorder.

Case 2: $|A_{wz}| + |A_{zw}| = 0$.

Here, we follow a similar construction to the preceding case, except here we first consider a single chain of the form:

$$\begin{aligned}
\bar{\mathbf{h}}(z) &\succeq (\bar{\mathbf{h}} + \mathbf{p}^1)(w) \\
&\succeq (\bar{\mathbf{h}} + \mathbf{p}^1 + \mathbf{q}^1)(z) \\
&\qquad\vdots \\
&\succeq \left(\bar{\mathbf{h}} + J \sum_{i=1}^{I} \mathbf{p}^i + I \sum_{j=1}^{J} \mathbf{q}^j\right)(z).
\end{aligned}$$

Consider now the following modification to this chain: immediately after the first instance of an $\mathbf{f}(z) \succeq \mathbf{g}(w)$ relation, insert one application of each transformation in A_{ww}. Similarly, after the first $\mathbf{f}(w) \succeq \mathbf{g}(z)$ relation, insert an application of each transformation in A_{zz}. The result is a chain:

$$\bar{\mathbf{h}}(z) \succeq \cdots \succeq \left(\bar{\mathbf{h}} + I \sum_{j=1}^{J} \mathbf{q}^j + J \sum_{i=1}^{I} \mathbf{p}^i + \sum_{k=1}^{K} \mathbf{r}^k\right)(z).$$

But by analogous logic to the former case, this also defines a cycle, contradicting the assumption that $\succeq$ is a preorder. Since these cases are exhaustive, we conclude such an extension must exist, which completes the proof. □

We now are in a position to prove Theorem 1.

Proof. Suppose $\succsim_{\mathcal{M}}^{R}$ is acyclic. By Lemma 1, the transitive closure of $\succsim_{\mathcal{M}}^{R}$ is a weakly $\mathcal{M}$-invariant pre-order, and hence by Lemma 2 admits an $\mathcal{M}$-invariant preorder extension.

The remainder of the proof follows from a standard transfinite induction argument. Let $\mathcal{P}_{\mathcal{M}}$ denote the set of $\mathcal{M}$-invariant preorders on X, partially ordered by extension. Given $\succeq_1, \succeq_2 \in \mathcal{P}_{\mathcal{M}}$, we write $\succeq_1 \rhd \succeq_2$ whenever $\succeq_1$ extends

$\succeq_2$. Let $\{\succeq_\lambda\}_{\lambda\in\Lambda}$ be an arbitrary $\rhd$-chain of $\mathcal{M}$-invariant preorders. It follows from standard arguments (see, e.g., Richter 1966; Chambers and Echenique 2016) that:

$$\bar{\succeq} = \bigcup_{\lambda\in\Lambda} \succeq_\lambda$$

is a preorder extension of every $\succeq_\lambda$. Similarly, it follows that $\bar{\succeq}$ is $\mathcal{M}$-invariant: if $(x, y) \in \bar{\succeq}$, then there exists some $\lambda \in \Lambda$ such that $x \succeq_\lambda y$, and since $\succeq_\lambda$ is $\mathcal{M}$-invariant, so must be $\bar{\succeq}$ since it extends $\succeq_\lambda$. Hence $\bar{\succeq}$ belongs to $\mathcal{P}_\mathcal{M}$, and by Zorn's Lemma, there exists a maximal $\mathcal{M}$-invariant preorder $\succeq^*$ which extends $\succsim^R_\mathcal{M}$. Suppose, for purposes of obtaining a contradiction, that $\succeq^*$ is not complete. Then there exist $w, z \in X$ that are $\succeq^*$-unrelated. By Lemma 3 there exists an $\mathcal{M}$-invariant preorder extension of $\succeq^*$ that renders w and z comparable, however, this contradicts the $\rhd$-maximality of $\succeq^*$. Thus $\succeq^*$ is complete and hence is an $\mathcal{M}$-invariant rationalizing preference for $\succsim^R_\mathcal{M}$, and hence $\succsim^R$. $\square$

B Proof of Theorem 2

B.1 Preliminaries from Propositional Logic

As in the main text, for all $(x, y) \in X \times X$, define two boolean variables:

$$[\mathtt{x} \succeq \mathtt{y}] \quad \text{and} \quad [\mathtt{x} \succ \mathtt{y}].$$

Let $\mathcal{V}$ denote the set of all such variables. A **model** is a mapping $\mu : \mathcal{V} \to \{\top, \bot\}$ assigning a truth value to every variable in $\mathcal{V}$.[25] We may extend any model from boolean variables to well-formed logical formulae in the obvious manner. For a proof of this fact, and an introduction to propositional logic, the interested reader is referred to Schöning (2008).

Every formula in propositional logic is equivalent to one in conjunctive normal form (CNF).[26] A **literal** is an atomic formula, of the form A or $\neg A$, for some $A \in \mathcal{V}$. A finite formula F in conjunctive normal form can be written as:

$$F = (A_{1,1} \vee \cdots \vee A_{1,n_1}) \wedge \cdots \wedge (A_{K,1} \vee \cdots \vee A_{K,n_K}),$$

where each $A_{i,j}$ is a literal. We view the formula F as being formed by the individual **clauses**:

$$C_i = A_{i,1} \vee \cdots \vee A_{i,n_i}.$$

A formula such as F can be compactly expressed in set notation:

$$\Big\{ \underbrace{\{A_{1,1}, \dots, A_{1,n_1}\}}_{C_1}, \dots, \underbrace{\{A_{K,1}, \dots, A_{K,n_K}\}}_{C_K} \Big\},$$

[25] In this appendix, we will exclusively use the word 'model' in its logical interpretation, rather than its economic meaning in the main text.

[26] See Schöning 2008.

where each $C_i = \{A_{i,1}, \dots, A_{i,n_i}\}$ is a clause. In other words, within a clause, a comma denotes an OR operation (i.e. $\vee$), and a comma between clauses denotes an AND (i.e. $\wedge$). The formula consisting only of the empty clause $\{\varnothing\}$ is a valid formula; by definition it is unsatisfiable.

Let C_1, C_2, and R be clauses. We say that R is a **resolvent** of C_1 and C_2 if there exists some literal L such that $L \in C_1$ and $\neg L \in C_2$, and

$$R = \big(C_1 \setminus \{L\}\big) \,\cup\, \big(C_2 \setminus \{\neg L\}\big).$$

The value of resolution is made clear by the following lemma (for a proof see Schöning 2008 p.32).

Lemma 4. *Let Θ be a set of clauses, and let R be the resolvent of two clauses C_1 and C_2 in Θ. Then Θ and $\Theta \cup \{R\}$ are logically equivalent.*

When speaking of resolvents, we explicitly allow for R to be the empty set. Suppose Θ is a set of clauses. A **derivation** of $\varnothing$ via resolution is a finite sequence of clauses $\{C_1, \dots, C_N\}$ such that:

(i) $C_N = \varnothing$; and

(ii) For all $i = 1, \dots, N$, C_i is either a clause in Θ, or a resolvent some C_j and C_k (the **parents** of C_i), where $j, k < i$.

More generally, if we remove condition (i) we speak of a **partial derivation** (of C_N). A set of clauses Θ is said to be **unsatisfiable** if and only if there is no model which evaluates every formula in Θ to $\top$. Remarkably, by forming a finite number of resolvents, one is always capable of detecting whether any finite set of formulas is unsatisfiable.

Theorem 4 (Robinson 1965)**.** *Let Θ be a finite set of clauses. Then Θ is unsatisfiable if and only if there exists a derivation of $\varnothing$ via resolution.*

Robinson (1965) paper actually proves stronger analogous result, in the more general setting of first-order logic. For a proof of the above result in propositional logic, the interested reader is referred to Schöning (2008), Chapter 1, Section 5. Many refinements of Theorem 4 exist, intended to further reduce the search space for proofs in the context of machine learning. We will have use of the following modification: say a derivation $\{C_1, \dots, C_N\}$ of $\varnothing$ is via **negative resolution** if:

(i) $C_N = \varnothing$; and

(ii') For all $i = 1, \dots, N$, C_i is either a clause in Θ, or a resolvent of some C_j and C_k, where $j, k < i$ and either C_j or C_k contains no positive literals.

The following theorem is proven on p.102 in Schöning (2008).

Theorem 5. *Let Θ be a finite set of formulas. Then Θ is unsatisfiable if and only if there exists a derivation of $\varnothing$ via negative resolution.*

The value of Theorem 5 is that it provides a 'representation theorem' of sorts for proofs of inconsistency. There may be (many) proofs that a given set of clauses is unsatisfiable; Theorem 5 however guarantees that at least one such proof can be carried out wholly via resolution where one parent contains no positive literals. Crucially, every constraint set in $\mathcal{C}$ may be regarded as a disjunction of negative literals. This gives a means of connecting derivations of $\varnothing$ from Φ via negative resolution with collapses of constraint sets.

B.2 $\mathcal{M}$-invariant Rationalization

Let Φ denote the collection of all logical formulas of the following form:

(T.1) **Completeness**: For all $x, y \in X$:

$$[\mathtt{x} \succeq \mathtt{y}] \vee [\mathtt{y} \succeq \mathtt{x}].$$

This is in conjunctive normal form (CNF).

(T.2) **Coherency**: For all $x, y \in X$:

$$[\mathtt{x} \succeq \mathtt{y}] \iff \neg[\mathtt{y} \succ \mathtt{x}].$$

In CNF, this may be regarded as two separate clauses,

$$\neg[\mathtt{x} \succeq \mathtt{y}] \vee \neg[\mathtt{y} \succ \mathtt{x}] \tag{T.2.a}$$

and

$$[\mathtt{x} \succeq \mathtt{y}] \vee [\mathtt{y} \succ \mathtt{x}]. \tag{T.2.b}$$

(T.3) **Transitivity**: For all $x, y, z \in X$:

$$[\mathtt{x} \succeq \mathtt{y}] \wedge [\mathtt{y} \succeq \mathtt{z}] \implies [\mathtt{x} \succeq \mathtt{z}],$$

or, in CNF:

$$\neg[\mathtt{x} \succeq \mathtt{y}] \vee \neg[\mathtt{y} \succeq \mathtt{z}] \vee [\mathtt{x} \succeq \mathtt{z}].$$

(T.4) **Extension**: For all $(x, y) \in \succsim^R$,

$$[\mathtt{x} \succeq \mathtt{y}].$$

Moreover, if $(x, y) \in \succ^R$ then:

$$[\mathtt{x} \succ \mathtt{y}].$$

(T.5) **Invariance**: For all $x, y \in X$ and $\omega \in \mathcal{M}$ such that x, y belong to the domain of ω:

$$[\mathtt{x} \succeq \mathtt{y}] \iff [\omega(\mathtt{x}) \succeq \omega(\mathtt{y})],$$

or

$$\neg[\mathtt{x} \succeq \mathtt{y}] \vee [\omega(\mathtt{x}) \succeq \omega(\mathtt{y})] \tag{T.5.a}$$

and

$$[\mathtt{x} \succeq \mathtt{y}] \vee \neg[\omega(\mathtt{x}) \succeq \omega(\mathtt{y})]. \tag{T.5.b}$$

By construction, the set of models which evaluate to $\top$ for every formula in Φ are in 1-1 correspondence with the $\mathcal{M}$-invariant weak order extensions of $\langle \succsim^R, \succ^R \rangle$.

Let $\mathcal{C}^0$ denote the collection of all clausal forms of the constraint sets generated by $\langle \succsim^R, \succ^R \rangle$. Thus for $C \in \mathcal{C}^0$ we write:

$$C = \neg[\mathtt{y}_0 \succ \mathtt{x}_0] \vee \neg[\mathtt{y}_1 \succeq \mathtt{x}_1] \vee \cdots \vee \neg[\mathtt{y}_N \succeq \mathtt{x}_N],$$

or:

$$C = \big\{\neg[\mathtt{y}_0 \succ \mathtt{x}_0], \neg[\mathtt{y}_1 \succeq \mathtt{x}_1], \cdots, \neg[\mathtt{y}_N \succeq \mathtt{x}_N]\big\}.$$

B.3 Proofs

We proceed in the proof of Theorem 2 via several lemmas.

Lemma 5. *Suppose $\varnothing \in \mathcal{C}^*$. Then there does not exist any $\mathcal{M}$-invariant preference relation extending $\langle \succsim^R, \succ^R \rangle$.*

Proof. Let Θ denote the collection of all clauses of the form (T.1) - (T.5), as well as all clauses in $\mathcal{C}^0$.[27] If $\succeq$ is an $\mathcal{M}$-invariant weak order extension of $\langle \succsim^R, \succ^R \rangle$, by defining a model via (i) $[\mathtt{x} \succeq \mathtt{y}] = \top$ if and only if $x \succeq y$ and (ii) $[\mathtt{x} \succ \mathtt{y}]$ if and only if $x \succ y$, this model must evaluate every clause in Θ to $\top$.[28]

Let $C, C' \in \mathcal{C}^0$ denote the reduced form representations derived from some forcing collections (either strict or weak), and suppose $D \in \mathcal{C}^1$ is the collapse of C and C'. Regarding these as sets of negative literals, there exists (negative) literals $L \in C$ and $L' \in C'$ such that:

$$D = \big(C \setminus \{L\}\big) \ \cup \ \big(C' \setminus \{L'\}\big)$$

and either:

$$L = \neg[\omega(\mathtt{y}) \succeq \omega(\mathtt{x})] \quad \text{and} \quad L' = \neg[\omega'(\mathtt{x}) \succeq \omega'(\mathtt{y})]$$

or

$$L = \neg[\omega(\mathtt{y}) \succeq \omega(\mathtt{x})] \quad \text{and} \quad L' = \neg[\omega'(\mathtt{x}) \succ \omega'(\mathtt{y})]$$

for some $x, y \in X$, $\omega, \omega' \in \mathcal{M}$. Suppose L and L' are of the former type. Then $\neg[\omega(\mathtt{y}) \succeq \omega(\mathtt{x})] \in C$, hence we may form C_1 by resolving C with the (T.5.a) clause $\neg[\mathtt{y} \succeq \mathtt{x}] \vee [\omega(\mathtt{y}) \succeq \omega(\mathtt{x})]$. Then form C_2 by resolving C_1 with the (T.1) clause $[\mathtt{x} \succeq \mathtt{y}] \vee [\mathtt{y} \succeq \mathtt{x}]$, and finally form C_3 by resolving C_2 with the (T.5.a) clause $\neg[\mathtt{x} \succeq \mathtt{y}] \vee [\omega'(\mathtt{x}) \succeq \omega'(\mathtt{y})]$. Thus:

$$C_3 = \big(C \setminus \{L\}\big) \ \cup \ \big\{[\omega'(\mathtt{x}) \succeq \omega'(\mathtt{y})]\big\}.$$

[27] We view Φ as a collection of sets, given that each clause is viewed as a set itself.

[28] The clauses of the form (T.1)-(T.5) are clearly necessary as they define the basic properties of an invariant weak order extension. Clauses in $\mathcal{C}$ must also hold lest $\succeq$ contain a cycle. Every clause in $\mathcal{C}$ can be obtained from (T.1) - (T.5) as primitives, however we do not need this fact in light of these clauses' obvious necessity via standard order-theoretic arguments.

Then C_3 and C' can be resolved to form D. By Lemma 4, Φ and $\Phi \cup \{D\}$ are logically equivalent (i.e. Φ is satisfiable if and only if $\Phi \wedge \{D\}$ is satisfiable).

Proceeding, suppose now instead that L and L' are of the latter type. Again form C_1 via resolving C with the (T.5.a) clause $\neg[\mathtt{y} \succeq \mathtt{x}] \vee [\omega(\mathtt{y}) \succeq \omega(\mathtt{x})]$, and then C_2 by resolving C_1 and the (T.5.b) clause $[\mathtt{y} \succeq \mathtt{x}] \vee \neg[\omega(\mathtt{y}) \succeq \omega(\mathtt{x})]$. Finally, form C_3 by resolving C_2 with the type (T.2.b) $[\omega(\mathtt{y}) \succeq \omega(\mathtt{x})] \vee [\omega(\mathtt{x}) \succ \omega(\mathtt{y})]$. Then D is the resolvent of C_3 and C', and hence by an analogous argument, Φ and $\Phi \cup \{D\}$ are again logically equivalent.

We have thus shown that Θ and $\Theta \cup \mathcal{C}^1$ are logically equivalent. However, nothing in our argument relied on C, C' belonging to $\mathcal{C}^0$, rather than any other $\mathcal{C}^n$, hence we have actually shown that $\Theta \cup \mathcal{C}^n$ and $\Theta \cup \mathcal{C}^{n+1}$ are logically equivalent, and threfore so too are Θ and $\Theta \cup \mathcal{C}^*$. Since any model evaluates the empty clause $\varnothing$ to $\perp$, if $\varnothing \in \mathcal{C}^*$, then $\mathcal{C}^*$ is unsatisfiable, and by the above therefore so too is Θ. Thus no $\mathcal{M}$-invariant weak order extension of $\langle \succsim^R, \succ^R \rangle$ can exist. □

Lemma 6. *Let C be a disjunction of negative literals such that $C \in \Phi$ or C is the resolvent of two elements of Φ, one of which contains no positive literals. Then $C \in \mathcal{C}^1$.*

Proof. Suppose first that $C \in \Phi$. Since C is a disjunction of negative literals, it must be of the form (T.2.a), i.e.:

$$C = \neg[\mathtt{x} \succeq \mathtt{y}] \vee \neg[\mathtt{y} \succ \mathtt{x}].$$

Then C corresponds to a constraint set for the forcing collection:

$$x \succsim^R x$$
$$y \succsim^R y$$

and hence $C \in \mathcal{C}^0 \subseteq \mathcal{C}^1$. Suppose instead then that C is the resolvent of $C', D \in \Phi$, where D is a disjunction of negative literals and hence $D = \neg[\mathtt{x} \succeq \mathtt{y}] \vee \neg[\mathtt{y} \succ \mathtt{x}]$. Since C also contains no positive literals, it must be the case that $C' \in \Phi$ contains exactly one positive literal. Therefore it must be of the form (T.3), (T.4) or (T.5).

Case: $C' = \neg[\mathtt{x} \succeq \mathtt{z}] \vee \neg[\mathtt{z} \succeq \mathtt{y}] \vee [\mathtt{x} \succeq \mathtt{y}]$. Then:

$$x \succsim_c x$$
$$z \succsim_c z$$
$$y \succsim_c y$$

defines a forcing collection which C is a constraint set for, and hence again $C \in \mathcal{C}^0 \subseteq \mathcal{C}^1$.

Case: $C' = [\mathtt{x} \succeq \mathtt{y}]$ or $C' = [\mathtt{y} \succ \mathtt{x}]$. If the former is true, then $x \succsim_c y$ and hence:

$$x \succsim_c y$$

is a forcing collection for $C = \neg[\mathtt{x} \succ \mathtt{y}]$. If instead the latter is true, then $x \succ_c y$ and:

$$x \succ_c y$$

is a forcing collection whose reduced form equals $C = \neg[\mathtt{x} \succeq \mathtt{y}]$. In either case, we again find $C \in \mathcal{C}^0 \subseteq \mathcal{C}^1$.

Case: $C' = [\mathtt{x} \succeq \mathtt{y}] \vee \neg[\omega(\mathtt{x}) \succeq \omega(\mathtt{y})]$. Then:

$$C = \neg[\mathtt{y} \succ \mathtt{x}] \vee \neg[\omega(\mathtt{x}) \succeq \omega(\mathtt{y})].$$

However, the forcing collections:

$$\begin{array}{ll} x \succsim^R x & \omega(x) \succsim^R \omega(x) \\ y \succsim^R y & \omega(y) \succsim^R \omega(y) \end{array}$$

yield constraint sets (written both in set form and as disjunctions of negative literals):

$$\big\{[\mathtt{x} \succeq \mathtt{y}], [\mathtt{y} \succ \mathtt{x}]\big\} \quad \longleftrightarrow \quad \neg[\mathtt{x} \succeq \mathtt{y}] \vee \neg[\mathtt{y} \succ \mathtt{x}]$$

and

$$\big\{[\omega(\mathtt{x}) \succeq \omega(\mathtt{y})], [\omega(\mathtt{y}) \succ \omega(\mathtt{x})]\big\} \quad \longleftrightarrow \quad \neg[\omega(\mathtt{x}) \succeq \omega(\mathtt{y})] \vee \neg[\omega(\mathtt{y}) \succ \omega(\mathtt{x})].$$

Thus letting $L = [\mathtt{x} \succeq \mathtt{y}]$ and $L' = [\omega(\mathtt{y}) \succ \omega(\mathtt{x})]$, C is simply the collapse of these two constraint sets (written as a disjunction of negative literals) and hence belongs to $\mathcal{C}^1$. An analogous argument obtains if instead $C' = \neg[x' \succeq y'] \vee [\mathtt{x} \succeq \mathtt{y}]$ where for some $\omega \in \mathcal{M}$ we have $\omega(x') = x$ and $\omega(y') = y$. $\square$

Lemma 7. *Suppose there does not exist an $\mathcal{M}$-invariant weak order extension of $\langle \succsim^R, \succ^R \rangle$. Then $\varnothing \in \mathcal{C}^*$.*

Proof. By construction, the $\mathcal{M}$-invariant preference relations extending $\langle \succsim^R, \succ^R \rangle$ are in one-to-one correspondence with the models for Φ. Thus if no such extension exists, Φ is unsatisfiable. By Propositional Compactness (see Schöning 2008 Chapter I.4), there exists a finite unsatisfiable subset $\Phi^* \subseteq \Phi$.

By Theorem 5, there exists a derivation of the empty set via negative resolution, i.e. there exists a sequence of clauses $\{C_1, \ldots, C_N\}$ such that (i) $C_N = \varnothing$, (ii) for all $1 \leq n \leq N-1$ the clause C_n either belongs to Φ^* or is the resolvent of two clauses C_i and C_j, with $i, j < n$, one of which contains no positive literals.

Let $\{D_1, \ldots, D_K\}$ denote those clauses in $\{C_1, \ldots, C_N\}$ which contain no positive literals.[29] For each D_k, if D_k is the resolvent of some C_i and D_j, define D_j to be its **negative parent** (if D_k is not a resolvent, then we say D_k has no negative parent). Furthermore, if C_i itself is the resolvent of some $C_{i'}$ and $D_{j'}$, then we say $D_{j'}$ is the **negative grandparent** of D_k (similarly, if $C_i \in \Phi^*$, i.e. C_i is not a resolvent, then we say D_k has no negative grandparent). Define

[29]Without loss of generality we may suppose that C_1 contains no positive literals and hence $D_1 = C_1$.

$\mathcal{NP}(D_k)$, the **negative predecessors** of D_k, as the set consisting of D_k's negative parent and grandparent (if these exist).

Let $\mathcal{D}^0 \subseteq \{D_1, \ldots, D_K\}$ denote the subset of all D_k which belong to $\mathcal{C}^0$.[30] For each $n \geq 1$, define inductively:

$$\mathcal{D}^n = \left\{D_k : \mathcal{NP}(D_k) \subseteq \mathcal{D}^{n-1}\right\} \,\cup\, \mathcal{D}^{n-1}.$$

In other words, $\mathcal{D}^n$ consists of those positive-literal-free clauses D_k all of whose negative predecessors (if these exist) belong to $\mathcal{D}^{n-1}$ or lower. Viewing $\{C_1, \ldots, C_N\}$ as a binary tree (Schöning 2008, Chapter I.5), by Lemma 6 the sets $\{\mathcal{D}^n\}_{n=0}^{\infty}$ cover $\{D_1, \ldots, D_K\}$.[31] We now wish to show that for all $n \geq 1$, $\mathcal{D}^n \subseteq \mathcal{C}^n \subseteq \mathcal{C}^*$. By definition, $\mathcal{D}^0 \subseteq \mathcal{C}^0$. Thus suppose now that for all $n \leq M$, we have $\mathcal{D}^n \subseteq \mathcal{C}^n$, and consider $n = M + 1$. Let $D_k \in \mathcal{D}^{M+1}$. We consider three cases.

Case 1: D_k has negative parent D_j and negative grandparent $D_{j'}$, both of which belong to $\mathcal{D}^M$ and hence $\mathcal{C}^M$ by the inductive hypothesis. Then D_k is the resolvent of D_j and some C_i, and C_i the resolvent of $D_{j'}$ and some $C_{i'}$. Since D_k and D_j contain no positive literals, this means C_i must contain exactly one positive literal. In turn, since $D_{j'}$ contains no positive literals, this implies $C_{i'}$ must contain exactly two positive literals. Since Φ^* contains no clauses with more than two positive literals, and since every resolvent in $\{C_1, \ldots, C_N\}$ has a parent containing no positive literals, no resolvent in $\{C_1, \ldots, C_N\}$ can have more than 1 positive literal. This means that $C_{i'} \in \Phi^*$ and hence is either of the form $C_{i'} = [\mathtt{x} \succeq \mathtt{y}] \vee [\mathtt{y} \succeq \mathtt{x}]$ or $C_{i'} = [\mathtt{x} \succeq \mathtt{y}] \vee [\mathtt{y} \succ \mathtt{x}]$. Suppose first that $C_{i'}$ is of the former form. Then C_i consists of $D_{j'}$ but with one literal reversed and made positive. Since this is C_i's only positive literal, it must be the cancelling literal when it is resolved with D_j, thus D_k is precisely the collapse of $D_{j'}$ and D_j, where the collapse comes from cancelling a pair of reversed weak literals. If, instead, $C_{i'}$ is of the latter form, then once again C_i consists of $D_{j'}$ but now the one literal is reversed, made positive, and made strict if it was weak, or vice-versa. This is then cancelled by resolving with D_j and hence D_k consists of the collapse of $D_{j'}$ and D_j where the collapse occurs between weak and strict opposing negative literals. In either case, we find that D_k is the collapse of two elements of $\mathcal{C}^M$ and hence belongs to $\mathcal{C}^{M+1}$ as desired.

Case 2: D_k has a negative parent D_j but no negative grandparent, i.e. $C_i \in \Phi^*$. Since D_j and D_k contain no positive literals, it must be that C_i contains exactly one positive literal. Thus C_i is either of the form:

(i) $C_i = \neg[\mathtt{x} \succeq \mathtt{z}] \vee \neg[\mathtt{z} \succeq \mathtt{y}] \vee [\mathtt{x} \succeq \mathtt{y}]$

(ii) $C_i = [\mathtt{x} \succeq \mathtt{y}]$ or $C_i = [\mathtt{y} \succ \mathtt{x}]$

[30] Note $\mathcal{D}^0$ is non-empty as it contains at least $D_1 \in \Phi^*$.

[31] Viewing the resolution proof as a finite binary tree, Lemma 6 shows that (i) every leaf that belongs to $\{D_1, \ldots, D_K\}$ belongs to $\mathcal{D}^0$, and (ii) every element of $\{D_1, \ldots, D_K\}$ that two leaves for parents belongs to $\mathcal{D}^1$. The claim then follows by inducting on how many generations of ancestors an element of $\{D_1, \ldots, D_K\}$ has in the tree.

(iii) $C_i = \neg[\mathtt{x} \succeq \mathtt{y}] \vee [\omega(\mathtt{x}) \succeq \omega(\mathtt{y})]$ or $C_i = [\mathtt{x} \succeq \mathtt{y}] \vee \neg[\omega(\mathtt{x}) \succeq \omega(\mathtt{y})]$ for some $\omega \in \mathcal{M}$.

Suppose first C_i is of form (i). Then the cancelling literal must be $[\mathtt{x} \succeq \mathtt{y}]$. However, since:

$$\begin{aligned} x &\succsim_c x \\ z &\succsim_c z \\ y &\succsim_c y \end{aligned}$$

is a forcing collection, we know $\neg[\mathtt{x} \succsim \mathtt{z}] \vee [\mathtt{z} \succsim \mathtt{y}] \vee \neg[\mathtt{y} \succ \mathtt{x}]$ belongs to $\mathcal{C}^0$. Therefore D_k can be formed from collapsing $\neg[\mathtt{x} \succeq \mathtt{z}] \vee [\mathtt{z} \succeq \mathtt{y}] \vee \neg[\mathtt{y} \succ \mathtt{x}] \in \mathcal{C}^0$ with D_j, where the collapse is between a strict and weak edge. Since $D_j \in \mathcal{C}^M$, this means $D_k \in \mathcal{C}^{M+1}$ as desired.

Suppose now that C_i is of type (ii). In the first case,

$$D_k = D_j \setminus \{\neg[\mathtt{x} \succeq \mathtt{y}]\}.$$

Note however that if $[\mathtt{x} \succeq \mathtt{y}] \in \Phi^*$, then $x \succsim_c y$, and thus:

$$x \succsim_c y$$

is a forcing collection for $\neg[\mathtt{y} \succ \mathtt{x}]$ and hence this clause belongs to $\mathcal{C}^0$. Thus D_k may be obtained as the collapse of $\neg[\mathtt{y} \succ \mathtt{x}]$ and D_j and hence belongs to $\mathtt{D}^{M+1}$. On the other hand, if C_i equals $[\mathtt{y} \succ \mathtt{x}]$ then $y \succ_c x$ and hence:

$$y \succ_c x$$

is a strict forcing collection for $\neg[\mathtt{x} \succeq \mathtt{y}]$ and since:

$$D_k = D_j \setminus \{\neg[\mathtt{y} \succ \mathtt{x}]\},$$

D_k is just the collapse of D_j and $\neg[\mathtt{x} \succeq \mathtt{y}]$, and hence once again belongs to $\mathcal{D}^{M+1}$.

Finally, suppose that C_i is of the former type (iii). Then the cancelling literal must be $[\omega(\mathtt{x}) \succeq \omega(\mathtt{y})]$. Thus D_k is equal to D_j but with the literal $\neg[\omega(\mathtt{x}) \succeq \omega(\mathtt{y})] \in D_j$ becoming $\neg[\mathtt{x} \succeq \mathtt{y}] \in D_k$. Now,

$$\begin{aligned} x &\succsim_c x \\ y &\succsim_c y \end{aligned}$$

is a forcing collection hence $\neg[\mathtt{x} \succeq \mathtt{y}] \vee \neg[\mathtt{y} \succ \mathtt{x}]$ belongs to $\mathcal{C}^0$. Then D_k arises as the collapse of $D_j \in \mathcal{D}^M$ and $\neg[\mathtt{x} \succeq \mathtt{y}] \vee \neg[\mathtt{y} \succ \mathtt{x}] \in \mathcal{C}^0 \subseteq \mathcal{D}^M$ along the pair $\neg[\mathtt{y} \succ \mathtt{x}]$ and $\neg[\omega(\mathtt{x}) \succeq \omega(\mathtt{y})]$, and hence belongs to $\mathcal{D}^{M+1}$ as desired. If instead C_i is of the latter type (iii), an analogous argument suffices.

Case 3: D_k has no negative parent. In this case, D_k cannot be a resolvent at all, and hence belongs to Φ^*. The only clauses in Φ^* which contain no positive literals are of the form $\neg[\mathtt{x} \succeq \mathtt{y}] \vee \neg[\mathtt{y} \succ \mathtt{x}]$. If D_k is of this form, then it belongs to $\mathcal{C}^0$ as

$$x \succsim_c x$$
$$y \succsim_c y$$

is a forcing collection for it, and hence it belongs to $\mathcal{C}^*$.

Now, as D_k cannot have a negative grandparent without a negative parent (as our proof of inconsistency is by *negative* resolution), these cases are exhaustive, and we find that for all $1 \leq k \leq K$, the clause $D_k \in \mathcal{C}^*$. Since $D_K = \varnothing$, this implies that $\varnothing \in \mathcal{C}^*$ as desired. □

The proof of Theorem 2 follows from these lemmas.

C Proof of Theorem 3

Proof. Suppose first that $\{[\mathtt{y} \succ \mathtt{x}]\} \in \mathcal{C}^*$. By an identical argument to that in the proof of Theorem 2, $\Phi \cup \{[\mathtt{y} \succ \mathtt{x}]\}$ is unsatisfiable. Thus no model μ for Φ evaluates $\mu([\mathtt{y} \succ \mathtt{x}]) = \top$. Since the set of models for Φ are in 1-1 correspondence with the set of $\mathcal{M}$-invariant rationalizing preferences of $\langle \succsim^R, \succ^R \rangle$ (which is non-empty by hypothesis), we conclude every such rationalizing preference must weakly rank x above y. An identical argument holds for the case of $\{[\mathtt{y} \succeq \mathtt{x}]\} \in \mathcal{C}^*$ case.

Conversely, suppose every $\mathcal{M}$-invariant rationalizing preference $\succeq^*$ ranks $x \succeq^* y$. Then no model for Φ evaluates $[\mathtt{y} \succ \mathtt{x}]$ to $\top$, and hence $\Phi \cup \{[\mathtt{y} \succ \mathtt{x}]\}$ is unsatisfiable. Define Φ' as follows. First, remove from Φ any clause containing the literal $[\mathtt{y} \succ \mathtt{x}]$; then for every remaining clause that contains the negative literal $\neg[\mathtt{y} \succ \mathtt{x}]$, delete this literal from it. By construction, any model μ' for Φ' uniquely extends to a model μ for Φ which evaluates $\mu([\mathtt{y} \succ \mathtt{x}]) = \top$. Since no such models μ exist, Φ' must be unsatisfiable. By Propositional Compactness (see Schöning (2008) Chapter I.4), there exists a finite subset of $\Phi'' \subseteq \Phi'$ that is unsatisfiable; by Theorem 5, there exists a derivation $\{C_1, \ldots, C_N\}$ of $\varnothing$ from Φ'' via negative resolution. Let $\{D_1, \ldots, D_K\} \subset \{C_1, \ldots, C_N\}$ denote the elements of $\{C_1, \ldots, C_N\}$ belonging to Φ''. Note that each D_k either (i) belongs to Φ as well, or (ii) $D_k \cup \{\neg[\mathtt{y} \succ \mathtt{x}]\}$ belongs to Φ. Moreover, since Φ is satisfiable by hypothesis, at least one D_k must be of the latter type. Define:

$$\bar{D}_k = \begin{cases} C_i & \text{if } D_k \in \Phi \\ D_k \cup \{\neg[\mathtt{y} \succ \mathtt{x}]\} & \text{else.} \end{cases}$$

Then resolving the $\{\bar{D}_1, \ldots, \bar{D}_K\}$ in the same order as in the derivation $\{C_1, \ldots C_N\}$ generates a partial derivation $\{\bar{C}_1, \ldots, \bar{C}_N\}$ of $\neg[\mathtt{y} \succ \mathtt{x}]$ from Φ via negative resolution, and hence by an identical argument to Lemma 7 $[\mathtt{y} \succ \mathtt{x}] \in \mathcal{C}^*$. An identical argument again works for the case in which every extension ranks $x \succ^* y$. □

References

Afriat, S. N. (1967): "The construction of utility functions from expenditure data," *International economic review*, 8, 67–77.

Alcantud, J. C. (2009): "Conditional ordering extensions," *Economic Theory*, 39, 495–503.

Aumann, R. J. (1962): "Utility theory without the completeness axiom," *Econometrica*, 30, 445–462.

——— (1964): "Utility theory without the completeness axiom: a correction," *Econometrica*, 32, 210–212.

Bossert, W. (1999): "Intersection quasi-orderings: An alternative proof," *Order*, 16, 221–225.

Brown, D. J. and C. Calsamiglia (2007): "The nonparametric approach to applied welfare analysis," *Economic Theory*, 31, 183–188.

Caradonna, P. P. (2023): "Preference Regression," .

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2011): "Uncertainty averse preferences," *Journal of Economic Theory*, 146, 1275–1330.

Chambers, C. P. and F. Echenique (2016): *Revealed Preference Theory*, vol. 56, Cambridge University Press.

Chambers, C. P., F. Echenique, and K. Saito (2016): "Testing theories of financial decision making," *Proceedings of the National Academy of Sciences*, 113, 4003–4008.

Chambers, C. P., F. Echenique, and E. Shmaya (2014): "The axiomatic structure of empirical content," *American Economic Review*, 104, 2303–2319.

——— (2017): "General revealed preference theory," *Theoretical Economics*, 12, 493–511.

Chambers, C. P. and A. D. Miller (2018): "Benchmarking," *Theoretical Economics*, 13, 485–504.

Chateauneuf, A. and J. H. Faro (2009): "Ambiguity through confidence functions," *Journal of Mathematical Economics*, 45, 535–558.

Demuynck, T. (2009): "A general extension result with applications to convexity, homotheticity and monotonicity," *Mathematical Social Sciences*, 57, 96–109.

Diewert, W. E. (2012): "Afriat's theorem and some extensions to choice under uncertainty," *The Economic Journal*, 122, 305–331.

Donaldson, D. and J. A. Weymark (1998): "A quasiordering is the intersection of orderings," *Journal of Economic Theory*, 78, 382–387.

Dubra, J., F. Maccheroni, and E. A. Ok (2004): "Expected utility theory without the completeness axiom," *Journal of Economic Theory*, 115, 118–133.

Duggan, J. (1999): "A general extension theorem for binary relations," *Journal of Economic Theory*, 86, 1–16.

Dushnik, B. and E. W. Miller (1941): "Partially ordered sets," *American Journal of Mathematics*, 63, 600–610.

Echenique, F. (2020): "New developments in revealed preference theory: decisions under risk, uncertainty, and intertemporal choice," *Annual Review of Economics*, 12, 299–316.

Echenique, F. and K. Saito (2015): "Savage in the Market," *Econometrica*, 83, 1467–1495.

Epstein, L. G. (1983): "Stationary cardinal utility and optimal growth under uncertainty," *Journal of Economic Theory*, 31, 133–152.

Fishburn, P. C. and A. Rubinstein (1982): "Time preference," *International economic review*, 23, 677–694.

Freer, M. and C. Martinelli (2022): "An algebraic approach to revealed preference," *Economic Theory*, 1–26.

Fuchs, L. (2011): *Partially ordered algebraic systems*, vol. 28, Courier Corporation.

Gilboa, I. and D. Schmeidler (1989): "Maxmin expected utility with non-unique prior," *Journal of mathematical economics*, 18, 141–153.

Gonczarowski, Y. A., S. D. Kominers, and R. I. Shorrer (2019): "To Infinity and Beyond: A General Framework for Scaling Economic Theories," *arXiv preprint arXiv:1906.10333*.

Gorno, L. (2017): "A strict expected multi-utility theorem," *Journal of Mathematical Economics*, 71, 92–95.

Herstein, I. N. and J. Milnor (1953): "An axiomatic approach to measurable utility," *Econometrica, Journal of the Econometric Society*, 291–297.

Koopmans, T. C. (1960): "Stationary ordinal utility and impatience," *Econometrica*, 28, 287–309.

Levin, V. L. (1983): "Measurable utility theorems for closed and lexicographic preorders," *Soviet Mathematics Doklady*, 27, 639–643.

Maccheroni, F., M. Marinacci, and A. Rustichini (2006): "Ambiguity aversion, robustness, and the variational representation of preferences," *Econometrica*, 74, 1447–1498.

Mehra, R. and E. C. Prescott (1985): "The equity premium: A puzzle," *Journal of monetary Economics*, 15, 145–161.

Mu, X., L. Pomatto, P. Strack, and O. Tamuz (2021): "Monotone additive statistics," *arXiv preprint arXiv:2102.00618*.

Nehring, K. and C. Puppe (1999): "On the multi-preference approach to evaluating opportunities," *Social Choice and Welfare*, 16, 41–63.

Nishimura, H., E. A. Ok, and J. K.-H. Quah (2017): "A comprehensive approach to revealed preference theory," *American Economic Review*, 107, 1239–1263.

Ok, E. A. (2002): "Utility representation of an incomplete preference relation," *Journal of Economic Theory*, 104, 429–449.

Ok, E. A. and G. Riella (2014): "Topological closure of translation invariant preorders," *Mathematics of Operations Research*, 39, 737–745.

——— (2021): "Fully preorderable groups," *Order*, 38, 127–142.

Peleg, B. (1970): "Utility functions for partially ordered topological spaces," *Econometrica*, 38, 93–96.

Pomatto, L., P. Strack, and O. Tamuz (2023): "The cost of information: The case of constant marginal costs," *American Economic Review*, 113, 1360–1393.

Richter, M. K. (1966): "Revealed preference theory," *Econometrica*, 34, 635–645.

Robinson, J. A. (1965): "A machine-oriented logic based on the resolution principle," *Journal of the ACM (JACM)*, 12, 23–41.

Safra, Z. and U. Segal (1998): "Constant risk aversion," *journal of economic theory*, 83, 19–42.

Samuelson, P. A. (1938): "The empirical implications of utility analysis," *Econometrica, Journal of the Econometric Society*, 344–356.

Schöning, U. (2008): *Logic for computer scientists*, Springer Science & Business Media.

Trockel, W. (1989): "Classification of budget-invariant monotonic preferences," *Economics Letters*, 30, 7–10.

Varian, H. R. (1983): "Non-parametric tests of consumer behaviour," *The review of economic studies*, 50, 99–110.

Von Neumann, J. and O. Morgenstern (1947): "Theory of games and economic behavior, 2nd rev," .

Ward, M. (1942): "The closure operators of a lattice," *Annals of Mathematics*, 191–196.

Weymark, J. A. (2000): "A generalization of Moulin's Pareto extension theorem," *Mathematical Social Sciences*, 39, 235–240.