From Local Utility to Neural Networks

Shaowei Ke* Chen Zhao†‡

January 9, 2023

Abstract

By formalizing a preference-based notion of local linearity in the spirit of Machina (1982), we introduce two utility representations. Both are equivalent to continuous finite piecewise linear functions. In the first, it is as if the decision maker has an optimistic self and a cautious self playing a zero-sum game. In the second, the decision maker evaluates an alternative through a neural network. The representations are easy to apply and estimate, can be used for local utility analyses and analyzing choices under ambiguity, and nest the constant loss aversion model and models with hierarchical subjective product attributes as special cases.

1 Introduction

Linear utility functions are widely used in economics. For example, expected utility functions are linear in probability, and empirical research often uses utility functions that are linear in products' attributes. It is perhaps not surprising that these linear utility functions may not describe people's choice behavior well. Therefore, more descriptive nonlinear utility functions have been proposed, such as the constant loss aversion model of Kahneman and Tversky (1979), in which a product's utility depends on whether each of their attributes is in the gain region or the loss region.

One of the most prominent nonlinear utility functions was introduced in a seminal paper by Machina (1982), who, rather than assuming linearity, assumes that the utility function is differentiable. Differentiability ensures the existence of a local expected utility function

*Department of Economics, University of Michigan. Email: shaoweik@umich.edu.

†Faculty of Business and Economics, University of Hong Kong. Email: czhao@hku.hk.

‡We are grateful to David Ahn, Tilman Börgers, In-Koo Cho, Andrew Ellis, Mira Frick, Itzhak Gilboa, Faruk Gul, Ryota Iijima, Annie Liang, Yusufcan Masatlioglu, Paulo Natenzon, Pietro Ortoleva, Wolfgang Pesendorfer, and seminar participants at numerous seminars and conferences for helpful comments.

in an infinitesimal neighborhood of any risky prospect. In other words, the decision maker's utility function exhibits a form of local linearity. It has been argued that the main insights of several important results of expected utility theory continue to hold under differentiable utility functions, and these functions can generate behavior consistent with empirical findings that linear utility functions cannot.[1]

Despite Machina's (1982) ingenious and insightful analysis of nonlinear utility functions, several important issues remain unaddressed. First, it is unclear what axioms a decision maker's preference should satisfy to have a differentiable utility representation. The notion of local linearity is a natural relaxation of global linearity, but ideally we may want local linearity to be defined in terms of the decision maker's choice behavior, rather than in terms of the utility representation directly. Second, differentiable utility functions in general do not offer a simple interpretation of people's choice behavior, and therefore, it is not easy to identify special cases that match behavioral phenomena. Third, empirically, it is not obvious how one should estimate a general differentiable utility function.

In this paper, we take an alternative approach to study local linearity that addresses the above issues. Our theory is based on notions of local linearity imposed directly on choice behavior. Similar to Machina (1982), local utility analyses can be performed under our theory. Moreover, our theory has several advantages. First, compared to differentiable utility functions, the representations of the decision maker's preference that we characterize have a simple and natural axiomatic foundation, and are easier to interpret and apply. Second, our theory applies to many choice settings, and hence can accommodate a wider range of empirical findings. Third, our theory provides a useful framework for us to better understand a variety of choice axioms and decision models. Last, there are well developed and widely used methods to estimate some of our representations.

Specifically, the linearity of utility functions is characterized, for example, by the independence axiom from expected utility theory. Using a variation of the Allais paradox, we argue that although independence does not hold globally, as shown in the original Allais paradox, it does seem reasonable to assume that some notion of independence holds locally. Directly assuming that independence holds locally everywhere, however, will simply bring us back to linear utility functions. Therefore, we first weaken independence, then require that it hold locally everywhere. Roughly speaking, fixing any alternative x, we require that in some neighborhood of x, for any mixtures *with* x, the independence property holds.

We introduce two kinds of such weakening—weak local independence and weak local bi-independence—derived from relaxation of two equivalent axioms that characterize linear

[1] Although the analysis is done in the space of risky prospects, it is clear that the same insights also hold, for example, in the space of product attributes.

utility functions: independence and bi-independence, respectively.

Together with the standard weak order and continuity axioms, we first show that weak local independence implies that the decision maker's preference must exhibit *piecewise independence*, which means that the set of alternatives can be divided into several regions, and independence holds in each region. More importantly, under these axioms, generically, the preference does not have any differentiable utility representations. Thus, our notion of local independence/linearity that is defined based on choice behavior leads to a different class of functions from those of Machina (1982). Nonetheless, we show that the powerful local utility analysis introduced by Machina (1982) continues to work under our axioms.

It might seem that a preference that exhibits piecewise independence can be represented by a *continuous finite piecewise linear* (CFPL) function, but this is not true in general. Rather, weak local bi-independence, which is stronger than weak local independence, together with the same standard axioms, characterizes CFPL functions, which is an extremely important class of functions in many academic fields. Hence, the local linearity property of the utility function we derive from our behavioral definition of local independence is that the utility function is locally exactly linear almost everywhere.

We introduce two useful representations of the decision maker's preference that are equivalent to CFPL functions. The first is called the *cautiously optimistic linear utility* (COLU) representation, whose equivalence to CFPL functions follows from a result by Tarela and Martínez (1999) and Ovchinnikov (2002). A preference has a COLU representation if there are affine functions $U_1, \ldots, U_n$ and $I_1, \ldots, I_m \subseteq \{1, \ldots, n\}$ such that the decision maker's preference is represented by $\max_{1 \leqslant j \leqslant m} \min_{i \in I_j} U_i$. To interpret the COLU representation, it is as if the decision maker has two conflicting selves playing a sequential zero-sum game. The first takes an action (i.e., chooses some set I_j) to maximize the utility of the alternative, and the second takes an action (i.e., chooses some i from the set I_j chosen by the first self) to minimize the utility of the alternative. One may interpret the first self as the decision maker's optimistic self and the second as her cautious self when it comes to evaluating an alternative.[2]

A recent paper by Arora, Basu, Mianjy, and Mukherjee (2018) from the computer science literature shows that neural-network functions with rectified linear units are identical to CFPL functions. Following this observation, our second representation, which is equivalent to CFPL functions, features one of the most powerful ingredients of machine learning: neural networks. We call this representation the *neural-network utility* (NU) representation. In an

[2]The COLU representation and its interpretation are similar to the dual-self expected utility representation and its interpretation in a recent paper by Chandrasekher, Frick, Iijima, and Le Yaouanq (2022), but the motivations, choice domain, characterizations, and classes of functions characterized are different.

NU representation, the decision maker takes an alternative and outputs the utility of the alternative through a feedforward neural network. A feedforward neural network may have multiple hidden layers, and each hidden layer may have multiple neurons. Each neuron first aggregates its child neurons' values in an affine fashion. If the outcome of aggregation is above the normalized threshold—zero—this neuron is activated and the aggregation value may be passed to neurons in the next hidden layer. Otherwise, this neuron remains inactive and has zero value. A neuron in the first hidden layer affinely aggregates all components of the alternative, and the values of neurons in the last hidden layer are aggregated into the utility of the alternative.

To interpret this representation, think of X as a space of products characterized by their attributes. Under the NU representation, first, the decision maker considers multiple ways to evaluate a product, captured by the first-hidden-layer neurons. It is as if the decision maker uses the raw attributes of the product to form multiple advanced subjective attributes. Some of these subjective attributes may be active and some may be inactive. The decision maker may continue to consider multiple ways (captured by the second-hidden-layer neurons) to aggregate those active subjective attributes, which enables her to form more advanced subjective attributes. This process continues until she reaches a final evaluation of the product.

Similar to the COLU representation, the NU representation is easy to apply. Another advantage of the NU representation is that the empirical methods for estimating neural-network models are well developed and widely used in practice. We have seen substantial evidence of the practical success of neural-network models when paired with a large amount of data. Therefore, it is quite likely that the NU model will significantly outperform traditional economic models in prediction and help us identify behavioral phenomena that would be difficult to identify using traditional methods.

As we introduce the representations, we also present several useful applications of our theory. First, we show that the COLU representation provides a useful framework that helps us understand classic axioms from the ambiguity literature in a new way. Second, we show that the constant loss aversion model is a special case of the COLU representation. Writing the constant loss aversion model as a COLU representation gives the model a new interpretation, and suggests natural ways to generalize it. Third, we show how to construct neurons in a simple and intuitive way in the NU representation to capture behavioral effects, such as the decision maker's bias toward risk-free alternatives, which captures the certainty effect. Last, we present an NU representation in which the decision maker uses raw attributes of the product to form hierarchically more advanced/complex subjective attributes, which is an intuitive model of how people evaluate multi-attribute products.

1.1 Related Literature

A growing literature combines economic theory with machine learning. Fudenberg and Liang (2019) use a decision tree algorithm to study the initial play of games. By studying games the algorithm predicts well—but other economic models do not—they identify a new parameter that, if introduced to the best model, improves the model's performance. Cho and Libgober (2021) study a problem in which an agent uses historical data and algorithms to provide action recommendations to a sequence of players in order to maximize their average long-run payoffs. Caplin, Martin, and Marx (2022) and Ke, Wu, and Zhao (2022) analyze the questions of how to model machine learning and how to model people learning from complex machine learning algorithms, respectively.

Our main representations are equivalent to CFPL functions. The closest paper to ours is by Ellis and Masatlioglu (2021), who characterize a categorical thinking model. Fixing any reference point, they assume that bi-independence is preserved for any two cells of an exogenously given partition of the choice domain and allow the preference to be discontinuous across cells. We focus on continuous preferences and identify endogenously a finite number of cells that preserve bi-independence pairwisely from the preference.

In the Anscombe–Aumann choice domain, Siniscalchi (2006) has also characterized CFPL functions that satisfy the C-independence axiom of Gilboa and Schmeidler (1989). Chandrasekher et al. (2022) introduce the dual-self expected utility representation by dropping uncertainty aversion from Gilboa and Schmeidler's model. Both the dual-self expected utility representation and the maxmin expected utility representation of Gilboa and Schmeidler become special cases of CFPL functions that satisfy C-independence if the number of priors is finite.

Our paper is related to non-expected utility theory. As summarized by Karni, Maccheroni, and Marinacci (2015), there are three approaches to relax the expected utility model: the axiomatic approach, the descriptive approach, and the local utility analysis. Our theory, if we focus on the probability simplex as the choice domain, falls into the intersection of all three. Machina (1982) introduces the notion of local linearity and local utility and studies smooth utility functions. Due to differentiability, the utility function can be approximated by affine functions locally everywhere. However, as discussed before, first, local linearity is not defined on choice behavior directly. Second, the smooth utility function does not have a simple interpretation and thus is not easy to apply. Third, it is not obvious how to estimate a smooth utility function from data. Our approach addresses these issues. We provide two notions of local linearity defined on preferences. In sharp contrast to Machina's smooth utility functions, our notions of local linearity lead to utility representations that are generically nondifferentiable. Our COLU representation and the NU representation both have natural interpretations, and we offer simple application examples. The NU representation can be estimated using standard machine learning techniques.

Within the framework of expected utility theory and maintaining independence, Hara, Ok, and Riella (2019) characterize several new representations by relaxing completeness, transitivity, and continuity. The main representation, coalitional expected utility representation, only imposes reflexivity and independence. There is a set of sets (called coalitions) of expected utility functions such that a lottery x is preferred to another lottery y if and only if every coalition has a utility function that ranks x above y.

The rest of the paper is organized as follows. In Section 2, we motivate and introduce our behavioral definition of local linearity. Section 3 presents results based on weak local independence. Section 4 provides the characterization and uniqueness results for the CFPL representation of the decision maker's preference. Sections 5 and 6 introduce the COLU and NU representations, and discuss applications of these representations. Section 7 concludes.

2 Locally Linear Preferences

Consider a convex and compact choice domain $X \subseteq \mathbb{R}^N$ with nonempty interior. Each choice alternative $x = (x_1, \ldots, x_N) \in X$ is an N-tuple. For example, when X is a space of products described by their attributes, x_i is the value of attribute i. When X is the probability simplex in $\mathbb{R}^N$, $x \in X$ is called a lottery over N prizes, with x_i indicating the probability of prize i. We use x, y, z to denote generic choice alternatives. For any $\lambda \in [0, 1]$, we use λxy as shorthand for the convex combination $\lambda x + (1-\lambda)y$. The decision maker has a preference $\succsim$ on X. Its asymmetric and symmetric parts are denoted by $\succ$ and $\sim$, respectively.

We know from expected utility theory that $\succsim$ on X satisfies the following axioms if and only if it has a linear utility representation—that is, there exists an affine function U such that $x \succsim y \iff U(x) \geqslant U(y)$.

Axiom 1 *(Weak Order) $\succsim$ is complete and transitive.*

Axiom 2 *(Continuity) For any $x \in X$, $\{y \in X : y \succsim x\}$ and $\{y \in X : x \succsim y\}$ are closed.*

Axiom 3 *(Independence) For any $x, y, z \in X$ and $\lambda \in (0, 1)$, $x \succsim y \Leftrightarrow \lambda xz \succsim \lambda yz$.*

The idea of independence is simple—if x is better than y, mixing x and y with any weight should also be better than mixing x with z with the same weight, and vice versa. This idea can be expressed in an equivalent way that will be crucial in our paper.

Axiom 4 *(Bi-independence) For any $x, y, z, z' \in X$ with $z \sim z'$ and $\lambda \in (0, 1)$, $x \succsim y \Leftrightarrow \lambda xz \succsim \lambda yz'$.*

If we require $z = z'$, bi-independence implies independence. Conversely, by applying independence twice, we can obtain bi-independence.

Among these axioms, usually (bi-)independence is the one that is violated. For example, a well-known violation of (bi-)independence comes from the Allais paradox in the choice domain with objective uncertainty—i.e., when X is the probability simplex in $\mathbb{R}^N$ and each $x \in X$ is a lottery over n prizes. Confronting the following two pairs of lotteries, most decision makers choose the left-hand lottery from the first pair and the right-hand lottery from the second:

First pair		Second pair	
100%: \$1M	3%: \$0 87%: \$1M 10%: \$1.5M	87%: \$0 13%: \$1M	90%: \$0 10%: \$1.5M

To see why such behavior is a violation of (bi-)independence, let δ_r be the degenerate lottery that pays \$$r$ for sure. Let $x = \delta_{1\mathrm{M}}$, $y = \frac{10}{13}\delta_{1.5\mathrm{M}} + \frac{3}{13}\delta_0$, $z = \delta_{1\mathrm{M}}$, and $z' = \delta_0$. The first pair of lotteries becomes $0.13xz$ and $0.13yz$. The second pair becomes $0.13xz'$ and $0.13yz'$. (Bi-)independence requires that $0.13xz \succsim 0.13yz$ if and only if $0.13xz' \succsim 0.13xz'$. Therefore, the Allais paradox violates (bi-)independence, and thus is inconsistent with linear utility functions.

Many nonlinear utility functions have been proposed to accommodate empirical evidence inconsistent with linearity. One of the most prominent is found in Machina (1982). It is assumed that the utility function is differentiable, which implies the existence of a unique local expected utility function in an infinitesimal neighborhood of any lottery. In other words, the decision maker's utility function exhibits a form of local linearity. It is unclear, however, what axioms a decision maker's preference should satisfy to have such a utility representation. Below, we introduce a definition of locally linear choice behavior, without assuming some form of local linearity on the utility representation directly.

Since linearity is characterized by (bi-)independence, a natural idea for formalizing local linearity is to assume that (bi-)independence holds locally, i.e., when the choice alternatives are close to each other. First, let us show that part of this idea offers a natural solution to evidence against full linearity. Take the Allais paradox as an example. If the right-hand lottery $0.13yz$ in the first pair becomes almost degenerate, decision makers may not be much biased toward $\delta_{1\mathrm{M}}$, and hence the certainty effect may not be strong enough to trigger violations of (bi-)independence.[3] To see this, suppose we now have $0.013xz$ and $0.013y^*z$ in

[3]This idea is different from that of Harless (1992), who turns the risk-free lottery in the Allais paradox into a slightly risky one. The four lotteries in Harless's experiment are still far apart.

the first pair and $0.013xz^*$ and $0.013y^*z^*$ in the second, in which $y^* = \frac{10}{13}\delta_{1.5M} + \frac{3}{13}\delta_{0.5M}$ and $z^* = \delta_{0.5\mathrm{M}}$:

First pair		Second pair	
100%: \$1M	0.3%: \$0.5M	98.7%: \$0.5M	99%: \$0.5M
	98.7%: \$1M	1.3%: \$1M	1%: \$1.5M
	1%: \$1.5M		

Can we assume that (bi-)independence holds locally around any choice alternative then? The answer is negative, because in that case (bi-)independence will hold globally. Below, we introduce two novel ways to mildly weaken (bi-)independence, and require that the weaker version of (bi-)independence hold locally everywhere. A subset of X is said to be a *neighborhood* of an alternative x if it is an open convex set that contains x.

Axiom 5 *(Weak Local Independence) Any $z \in X$ has a neighborhood L_z such that for any $x, y \in L_z$ and $\lambda \in (0,1)$, $x \succsim y \Leftrightarrow \lambda xz \succsim \lambda yz$.*

For a subset L of X, we say that L *preserves independence* if $x, y, z, \lambda xz, \lambda yz \in L$ with $\lambda \in (0,1)$ implies that $x \succsim y \Leftrightarrow \lambda xz \succsim \lambda yz$. In other words, any three lotteries in this subset satisfy the property required by independence. Weak local independence does not imply that L_z preserves independence. It only requires that preferences be preserved when the alternatives in L_z are mixed with the *given* z. In fact, if instead we require that any $z \in X$ have a neighborhood L_z that preserves independence, independence will hold globally.

To see how weak local independence differs from preserving independence locally, consider the following example. Let $X = [0,1]$. Suppose $\succsim$ can be represented by

$$U(x) = \begin{cases} -x + 0.01, & \text{if } x < 0.01, \\ x - 0.01, & \text{if } x \geqslant 0.01. \end{cases} \tag{1}$$

It can be verified that no neighborhoods of $x = 0.01$ preserve independence, but weak local independence holds.

Weak local independence only informs us of the decision maker's local choice behavior. It does not impose any structure on the decision maker's preference when the choice alternatives are far apart. A local and weakened version of bi-independence, by contrast, imposes structures on the choice behavior in this situation.

Axiom 6 *(Weak Local Bi-independence) Any $z, z' \in X$ with $z \sim z'$ have neighborhoods L_z and $L_{z'}$, respectively, such that for any $x \in L_z$, $y \in L_{z'}$, and $\lambda \in (0,1)$, $x \succsim y \Leftrightarrow \lambda xz \succsim \lambda yz'$.*

For subsets L_1, L_2 of X, we say that L_1 and L_2 *preserve bi-independence* if $x, z, \lambda xz \in L_i$, $y, z', \lambda yz' \in L_{3-i}$ with $i \in \{1,2\}$, $\lambda \in (0,1)$, and $z \sim z'$ implies that $x \succsim y \Leftrightarrow \lambda xz \succsim \lambda yz'$.

Weak local bi-independence does not require that L_z and $L_{z'}$ preserve bi-independence. It only requires that preferences be preserved when the alternatives x and y are mixed respectively with the *given* z and z'. Clearly, weak local bi-independence weakens bi-independence in a fashion similar to how weak local independence weakens independence.

By letting $z = z'$ in weak local bi-independence, we obtain weak local independence. Thus, weak local bi-independence is stronger than weak local independence. In fact, weak local bi-independence is strictly stronger. Figure 1 is an example that satisfies weak local independence but does not satisfy weak local bi-independence.

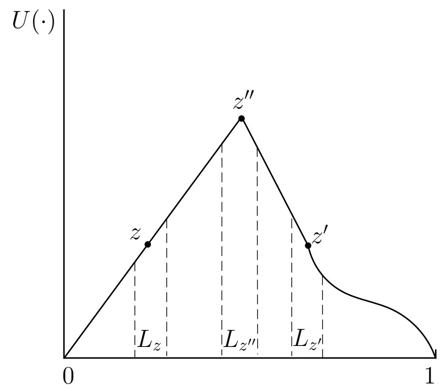

FIGURE 1: Let $X = [0,1]$. The decision maker's utility function $U : X \to \mathbb{R}$ is shown in the figure. In this example, every alternative $\tilde{z} \in [0,1]$ has a neighborhood such that for any x, y in that neighborhood and $\lambda \in (0,1)$, $x \succsim y \Leftrightarrow \lambda x\tilde{z} \succsim \lambda y\tilde{z}$. In particular, the neighborhood L_z satisfies this requirement trivially since U is monotone within L_z. However, z and z' do not satisfy the requirement of weak local bi-independence, since any neighborhood of z' includes a nonlinear segment. Take any alternative x from that nonlinear segment. We can find y in L_z such that $x \succ y$ but $\lambda yz \succ \lambda xz'$.

3 Weak Local Independence

First, we introduce results that mainly use weak local independence. A set $Y \subseteq X$ is *regular closed* if $Y = \mathrm{cl}(\mathrm{int}(Y))$. We say that $\succsim$ exhibits *piecewise independence* if there exists a collection of regular closed subsets whose union is X such that each of those subsets preserves independence. A linear utility function $U : X \to \mathbb{R}$ is said to be a *local utility function* of $\succsim$ if it represents $\succsim$ on a nonempty regular closed subset $Y \subseteq X$ and $U(X) \in \{\{0\}, [0,1]\}$. In other words, we normalize each local utility function so that it is either a constant utility function that assigns 0 to any alternative or has range $[0,1]$.

Theorem 1 *If $\succsim$ satisfies weak order, continuity, and weak local independence, then $\succsim$ exhibits piecewise independence. Furthermore, $\succsim$ has at most a finite number of local utility functions.*

Assuming weak order and continuity, weak local independence implies that one can decompose the choice domain into regular closed regions such that within each region, $\succsim$ has a linear utility representation. Note that the fact that $\succsim$ has a linear utility representation within each of these regions does not imply that $\succsim$ can be represented by a CFPL function (to be formally defined in the next section). For example, the preference in Figure 1 exhibits piecewise linear independence, since it is monotone within $[0, z'']$ and within $[z'', 1]$, but it cannot be represented by a CFPL function.

One important implication of Theorem 1 is that generically, weak local independence leads to nondifferentiability of the utility representation.

Proposition 1 *Suppose $\succsim$ satisfies weak order, continuity, and weak local independence. If each local utility function of $\succsim$ is non-constant and $\succsim$ has a differentiable utility representation, then there exists a linear utility function U such that any local utility function of $\succsim$ is either U or $1 - U$.*

Thus, assuming weak order, continuity, and weak local independence, if $\succsim$ has non-constant local utility functions, a necessary condition for $\succsim$ to have a differentiable utility representation is that it can have at most two local utility functions that represent opposite preferences. This means that for any regular closed regions X, Y that preserve independence, each level set in X must be parallel to each level set in Y. To see the intuition, suppose that there exists two regions in which the level sets are not parallel across regions. If this is the case, we can find some alternative at which the level set has a kink. Then any representation of $\succsim$ with non-constant local utility functions cannot be differentiable at that alternative.

Next, we present an example of $\succsim$ such that (i) $\succsim$ has a differentiable utility representation and (ii) there exists a non-constant U such that U and $1 - U$ are the only local utility functions of $\succsim$. Let $X = [0, 1]$ and suppose $\succsim$ can be represented by $V(x) = (x - 0.01)^2$.[4] Then the (normalized) local utility functions are $U_1(x) = x$ and $U_2(x) = 1 - x$, and the corresponding regular closed regions are $[0, 0.01]$ and $[0.01, 1]$, respectively. This example shows that it is possible for $\succsim$ to exhibit piecewise independence and have a differentiable representation even if it does not satisfy independence. However, situations in which a differentiable representation may exist—namely, when $\succsim$ has a constant local utility function or when $\succsim$ only has two opposite local utility functions—is nongeneric. Therefore, under the axioms in Proposition 1, generically, $\succsim$ will not have a differentiable utility representation.

Although $\succsim$ may not have a differentiable utility representation, it turns out that we can still perform the local utility analysis in the spirit of Machina (1982). Consider a preorder

[4]Note that this $\succsim$ can also be represented by the utility function defined in (1). Thus, the fact that a binary relation has a nondifferentiable representation does not imply that it does not have a differentiable representation.

(a reflexive and transitive binary relation) $\trianglerighteq$ defined on X. We say that a utility function U *respects* $\trianglerighteq$ if for any $x, y \in X$, $x \trianglerighteq y$ implies $U(x) \geqslant U(y)$. For example, in expected utility theory, an expected utility function with a strictly increasing Bernoulli index respects first-order stochastic dominance, which is a preorder. We say that $\succsim$ *respects* $\trianglerighteq$ if it has a utility representation that respects $\trianglerighteq$. We say that $\trianglerighteq$ satisfies *betweenness* if $x \trianglerighteq y$ implies that $x \trianglerighteq \lambda xy \trianglerighteq y$ for any $\lambda \in [0, 1]$.

Proposition 2 *Suppose $\succsim$ satisfies weak order, continuity, and weak local independence. For any preorder $\trianglerighteq$ that satisfies betweenness, if each local utility function of $\succsim$ respects $\trianglerighteq$, then $\succsim$ respects $\trianglerighteq$.*

Suppose X is the set of lotteries over a (finite) set of monetary prizes. Then first-order stochastic dominance and second-order stochastic dominance can be defined in the usual way as partial orders on X. Since both of these partial orders satisfy betweenness, we conclude that if each local utility function of $\succsim$ respects first-order (second-order) stochastic dominance, then $\succsim$ will respect first-order (second-order) stochastic dominance. Therefore, the insights from the main results of Machina (1982) also apply in our theory.

4 Weak Local Bi-independence

Now, we introduce the results that make use of weak local bi-independence. Since weak local bi-independence is stronger than weak local independence, our results in the previous section continue to hold here.

As discussed above, Figure 1 shows that weak local independence is too weak to ensure that the preference can be represented by a CFPL function, and that weak local bi-independence is strictly stronger than weak local independence. It turns out that weak local bi-independence exactly characterizes CFPL functions.

Definition 1 *A function $f : X \to \mathbb{R}$ is CFPL if f is continuous and there exist a finite collection of regular closed subsets whose union is X such that f is affine on each of those subsets.*[5]

If a preference can be represented by a CFPL function, we say that the preference has a CFPL representation.

Theorem 2 *The preference $\succsim$ has a CFPL representation if and only if $\succsim$ satisfies weak order, continuity, and weak local bi-independence.*

[5] In this definition, X may be replaced by a closed subset of X.

Assuming weak order and continuity, (bi-)independence characterizes linear functions on X, whereas weak local bi-independence characterizes CFPL functions on X. CFPL functions have been extremely useful in many academic fields, and the behavioral characterization of CFPL functions is more challenging than it might appear (see Siniscalchi (2006) for a related example).

According to this theorem, our stronger behavioral definition of local linearity, weak local bi-independence, leads to the following local linearity property of the utility function: The utility representation is locally exactly linear almost everywhere. The measure of the set of nondifferentiable points/alternatives is zero, and every differentiable point/alternative has a neighborhood such that the utility representation is affine on that neighborhood.

The uniqueness of CFPL representations is analogous to that of linear representations.

Proposition 3 *Suppose the preference $\succsim$ has a CFPL representation W. Then, $\succsim$ can be represented by another CFPL function $\tilde{W}$ if and only if there exists a strictly increasing CFPL function $f : W(X) \to \mathbb{R}$ such that $\tilde{W} = f \circ W$.*

Therefore, CFPL representations are unique up to positive CFPL transformations, just like linear representations are unique up to positive (continuous) linear transformations.

4.1 Sketch of the Proof of Theorem 2

We explain why the axioms imply that the preference has a CFPL representation. The proof consists of three parts. First, we identify the interior of all regions over which the preference can be represented by an affine function. By weak local independence (implied by weak local bi-independence), every alternative z has a neighborhood L_z such that for any $x, y \in L_z$ and $\lambda \in (0,1)$, $x \succsim y \Leftrightarrow \lambda xz \succsim \lambda yz$. Pick any $z' \in L_z$. We can find a neighborhood $L_{z'} \subseteq L_z$ that has the same property with respect to z'. Inductively, we can find N alternatives inside L_z and prove that the (regular closed) polytope formed by these N alternatives and z preserves independence. Figure 2 illustrates the construction when $N = 2$.

It may appear that the fact that there is a polytope containing z that preserves independence for every $z \in X$ is sufficient for us to construct a CFPL representation, but if this is true, we only need weak local independence rather than weak local bi-independence. We have seen a pathological example in Figure 1 that only satisfies the former but not the latter, and cannot be represented by a CFPL function. In the actual proof, we use weak local bi-independence instead to construct polytopes in a similar fashion. This allows us to show that every pair of such polytopes preserves bi-independence, which is the key to ensure that each of these polytopes is indeed part of a region over which the preference's CFPL

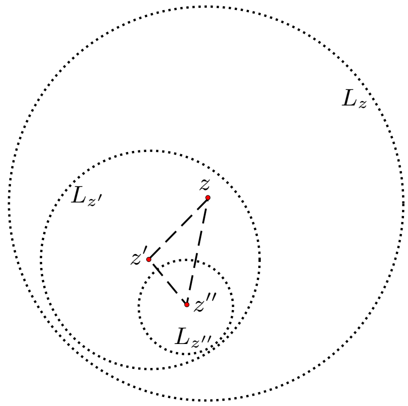

FIGURE 2: Let $N = 2$. Without loss of generality, assume that $L_{z''} \subseteq L_{z'} \subseteq L_z$. It can be shown that the triangle $\overline{z''z'z}$ preserves independence.

representation is affine (called a region for simplicity). Let the union of the interiors of these polytopes be X_o, which is the union of the interiors of all regions.

Next, we identify each region via Zorn's lemma. First, we consider the set of all functions that map X_o into subsets of X that preserve independence individually and bi-independence pairwisely. By Zorn's lemma, we are able to find a maximal element among these functions. It will assign each $z \in X_o$ a maximal region that satisfies the required properties. The image of this maximal function identifies all regions.

The number of regions must be finite. Intuitively, if we do not have finitely many regions, we can select an alternative in each region and then find an accumulation point of these alternatives such that *any* neighborhood of that accumulation point intersects with infinitely many regions. This turns out to violate weak local independence, which means that weak local bi-independence is also violated.

Then, we construct a CFPL representation of the preference. The key step in this part is to show that if a collection of subsets of X preserve independence individually and bi-independence pairwisely, we can construct a CFPL representation on the union of these subsets. The proof of this step is similar to Chapter 2.4 of Schmidt (1998).

5 Cautiously Optimistic Linear Utility

Tarela and Martínez (1999) and Ovchinnikov (2002) show that a CFPL function has a lattice polynomial representation that maximizes over a collection of minimums of sets of affine functions. Building on their results, we define the COLU representation of the decision maker's preference as follows.

Definition 2 *We say that $\succsim$ has a COLU representation if there are affine functions $U_1, \ldots, U_M$*

and index sets $I_1, \ldots, I_m \subseteq \{1, \ldots, M\}$ *such that*

$$x \succsim y \iff \max_{1 \leqslant j \leqslant m} \min_{i \in I_j} U_i(x) \geqslant \max_{1 \leqslant j \leqslant m} \min_{i \in I_j} U_i(y).$$

With a COLU representation, it is as if the decision maker has two selves who are playing a sequential zero-sum game. The first self takes an action (i.e., chooses a set of utility functions indicated by an index set I_j for further evaluations). The first self's goal is to maximize the utility of an alternative eventually. Then, the second self takes an action by choosing some i from the set I_j chosen by the first self. The second self's goal is to minimize the expected utility of the alternative.[6] The second self evaluates an alternative with caution. Given what the second self will do, the first self tries to evaluate the alternative more optimistically.

Corollary 1 *The preference* $\succsim$ *has a COLU representation if and only if* $\succsim$ *satisfies weak order, continuity, and weak local bi-independence.*

This result directly follows from our Theorem 2 and Tarela and Martínez (1999) and Ovchinnikov (2002).

A remarkable feature of the COLU representation is its finiteness. If we allow for infinitely many affine functions under the maximum and minimum operators in the representation, the COLU representation may not be truly piecewise linear. For example, one can have a collection of affine functions whose minimum ends up being a parabola. Indeed, similar representations in the literature, such as the dual-self expected utility representation of Chandrasekher et al. (2022) and its special case, the maxmin expected utility representation of Gilboa and Schmeidler (1989), do not have such finiteness. In order to obtain such finiteness for a representation, usually some rather demanding axioms are needed. In our case, we do not have any axioms that directly assume finiteness. Rather, finiteness is naturally implied by weak local independence, as shown by Theorem 1.

In what sense is the COLU representation unique? A COLU representation may have many affine functions under the maximum and minimum operators. Some of the affine functions may be redundant. We might hope that by removing the redundant ones, we can obtain a minimal COLU representation that is unique. However, this is not true. Similar observations have been made in recent papers with similar representations, such as Hara et al. (2019) and Chandrasekher et al. (2022). In those papers, the uniqueness of the representation is

[6] Similar to the dual-self expected utility representation of Chandrasekher et al. (2022), the format of the set of actions is not important. We can write the COLU representation equivalently as $\max_{a_1 \in A_1} \min_{a_2 \in A_2} U(x, a_1, a_2)$, in which A_1, A_2 are arbitrary action sets and there is only one utility function whose value depends on the actions taken by the two selves and the alternative that is being evaluated.

obtained via half-space closures. In our case, due to the finiteness of the COLU representation, we can have a much simpler and interpretable *canonical* COLU representation that is unique in some sense.

We already know that under weak order, continuity, and weak local bi-independence, the preference has a CFPL representation that is unique up to a positive CFPL transformation. Moreover, according to Tarela and Martínez (1999) and Ovchinnikov (2002), every CFPL function can be rewritten as a COLU representation. In general, however, a CFPL function may be equal to multiple distinct COLU representations. Therefore, below we discuss, fixing a particular CFPL representation of the preference, in what sense the COLU representation of the CFPL representation is unique. In particular, we follow the construction of Tarela and Martínez (1999) and Ovchinnikov (2002) to illustrate how distinct COLU representations that are equal to the same CFPL function can be transformed into the same unique canonical COLU representation.

Suppose a COLU representation is equal to a CFPL function V. According to the solution to the maxmin problem of the COLU representation, V's domain is divided into several maximal (in the sense of set inclusion) regular closed subsets, $X_1^*, \ldots, X_j^*$, such that $X = \bigcup_{i=1}^{j} X_i^*$ and V is affine on each X_i^*. Let $X_1, \ldots, X_k$ denote the connected components of $X_1^*, \ldots, X_j^*$. Let U_i denote the affine function that is identical to V on X_i, $i = 1, \ldots, k$. Let $\mathbb{U} = \{U_1, \ldots, U_k\}$. For each X_i, define a set of affine functions $\mathbb{U}_i = \{U \in \mathbb{U} : U_i(x) \leqslant U(x) \text{ for any } x \in X_i\}$. Let $\mathcal{U} = \{\mathbb{U}_1, \ldots, \mathbb{U}_k\}$. For any COLU representation V, we call $\max_{\mathbb{U}_i \in \mathcal{U}} \min_{U \in \mathbb{U}_i} U(x)$ its canonical COLU representation.

The construction of $\mathcal{U}$ can be understood as follows. First, we remove redundant affine functions from the original COLU representation. Each nonredundant affine function solves the maxmin problem of the original COLU representation on some regular closed maximally connected subsets of X, which must be a polytope. For each such polytope X_i such that the affine function U_i solves the maxmin problem on X_i, we construct a subset $\mathbb{U}_i$ of $\mathbb{U}$, which consists of nonredundant affine functions from the original COLU representation that dominate U_i. Putting these subsets $\mathbb{U}_i$ together, we obtain $\mathcal{U}$. The next result follows from Tarela and Martínez (1999) and Ovchinnikov (2002).

Corollary 2 *Every COLU representation is equal to its canonical COLU representation.*

Thus, two distinct COLU representations of the preference that are equal to some CFPL function V must have the same canonical COLU representation because the canonical COLU representation only depends on V.

5.1 COLU and Ambiguity

The COLU representation's functional form is similar to that of a model of ambiguity: the dual-self expected utility representation (Chandrasekher et al. (2022)). We put the COLU representation in the context of ambiguity in this subsection. Our choice domain not only nests the space of product attributes and the probability simplex as special cases, but also allows us to study subjective uncertainty. Let $X = [\underline{u}, \overline{u}]^N \subseteq \mathbb{R}^N$, in which N indicates the number of states. Each $x = (x_1, \ldots, x_N) \in X$ describes an *act* that assigns utility $x_i \in [\underline{u}, \overline{u}]$ to state i. Let $\mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^N$.

Under this interpretation of X, an affine function $f(x) = \mu \cdot x + \alpha$ becomes the sum of a constant and the inner product between a fixed finite signed measure on the state space $\mu \in \mathbb{R}^N$ and any act x. Below we list several standard axioms in the ambiguity literature, first introduced by Gilboa and Schmeidler (1989).

Axiom 7 *(C-Independence) For any $x, y \in X, u \in [\underline{u}, \overline{u}]$ and $\lambda \in (0,1)$, $x \succsim y \Leftrightarrow \lambda x(u\mathbf{1}) \succsim \lambda y(u\mathbf{1})$.*

Axiom 8 *(Monotonicity) For any $x, y \in X$, $x \geq y$ implies $x \succsim y$.*

Axiom 9 *(Uncertainty Aversion) For any $x, y \in X$ and $\lambda \in [0,1]$, $x \sim y$ implies $\lambda xy \succsim x$.*

The next result provides a new perspective on the implications of these axioms, through the lens of the COLU representation. To simplify the exposition, we impose a nondegeneracy assumption. We say that a preference is *nondegenerate* if there exist $u, v \in \mathbb{R}$ such that $u\mathbf{1} \succ v\mathbf{1}$.[7]

Proposition 4 *Suppose a nondegenerate $\succsim$ satisfies weak order, continuity, and weak local bi-independence. The following statements are true:*

1. *The preference satisfies C-independence, monotonicity, and uncertainty aversion if and only if it has a finite maxmin expected utility representation $\min_{i \in I} \mu_i \cdot x$, in which $\mu_i \in \mathbb{R}^N_+$ and $\mu_i \cdot \mathbf{1} = 1$ for any i.*

2. *The preference satisfies C-independence and monotonicity if and only if it has a COLU representation $\max_{1 \leqslant j \leqslant m} \min_{i \in I_j} \mu_i \cdot x$, in which $\mu_i \in \mathbb{R}^N_+$ and $\mu_i \cdot \mathbf{1} = 1$ for any i.*

3. *The preference satisfies monotonicity if and only if it has a COLU representation $\max_{1 \leqslant j \leqslant m} \min_{i \in I_j} \mu_i \cdot x + \alpha_i$, in which $\mu_i \in \mathbb{R}^N_+$ and $\alpha_i \in \mathbb{R}$ for any i.*

[7] Without nondegeneracy, the first two statements in Proposition 4 need to allow for the case in which all finite signed measures are zero measures.

4. *The preference satisfies C-independence if and only if it has a COLU representation* $\max_{1\leqslant j\leqslant m}\min_{i\in I_j}\mu_i\cdot x$*, in which* $\mu_i\in\mathbb{R}^N$ *and* $\mu_i\cdot\mathbf{1}$ *is identical for any* i*.*

The first result in Proposition 4 shows that if weak local bi-independence is assumed in addition to standard axioms in Gilboa and Schmeidler (1989), maxmin expected utility with a finite set of priors is obtained. Siniscalchi (2006) characterizes the same class of preferences by requiring, roughly speaking, that the mixture of nearby acts do not provide an effective hedge. Siniscalchi's axiom can also be regarded as a form of local linearity similar in spirit to weak local independence. Compared with weak local (bi-)independence, the statement of Siniscalchi's axiom is more involved and harder to interpret, and the bite of the axiom relies on C-independence.

The second result shows that removing uncertainty aversion from the axiomatic system yields the finite version of the dual-self expected utility representation by Chandrasekher et al. (2022). Chandrasekher (2019) introduces a partially finite version of the dual-self expected utility representation that has a finite number of index sets for the optimistic self to choose from, but each index set may contain an infinite number of measures. Thus, his axioms do not seem to lead to a CFPL representation.

The last two results show that in our framework, (i) monotonicity's role is to ensure that all finite signed measures in the COLU representation are nonnegative measures, and (ii) C-independence ensures that those finite signed measures have the same total mass, and the constant terms of the affine functions in the COLU representation can be assumed away. It can be inferred from Chandrasekher et al. (2022) that the finite signed measures in the COLU representation are the Clark differentials of the functional that aggregates utilities across states. In particular, Ghirardato, Maccheroni, and Marinacci (2004) point out that the combination of monotonicity and C-independence ensures that the Clark differentials must be a set of probability measures; that is, all finite signed measures are nonnegative and have the same total mass. When only one of the two axioms is assumed, however, the Clark differentials may not exist. In this situation, weak local bi-independence comes in handy, since it ensures the existence of a CFPL representation for which the Clark differentials must exist. Therefore, in our framework, the effects of C-independence and monotonicity can be cleanly separated.[8]

In summary, weak local bi-independence provides a clean framework to relate ambiguity

[8]Chateauneuf and Faro (2009) propose a maxmin representation with a confidence function over the probability measures. The confidence function scales the probability measures, essentially making them nonnegative measures with potentially different total masses, which is similar to our third result. However, to achieve this, Chateauneuf and Faro posit a set of weakenings of C-independence to obtain a homothetic aggregator. By contrast, with weak local bi-independence, the COLU representation is not necessarily homothetic.

representations with each other. On the one hand, weak local bi-independence is weak enough to allow for a plethora of choice behaviors. On the other hand, it is also powerful enough to generate nice technical properties such as the existence of Clark differentials and Lipschitz continuity.

5.2 COLU and Constant Loss Aversion

One of the most important ideas in behavioral economics is that people's choice behavior is affected by their reference points, and they treat gains and losses differently (see Tversky and Kahneman (1991) and Kahneman and Tversky (1979)). We first show how their model of reference dependence and loss aversion may be viewed as a special case of the COLU representation, then show how the COLU representation leads to natural generalizations of that model.

The well-known constant loss aversion model of Tversky and Kahneman (1991) can be easily rewritten as a special case of the COLU representation. To see this, consider a simple example in which X is a space of products described by the values of their two attributes. Let $v : \mathbb{R} \to \mathbb{R}$ be defined as follows:

$$v(x_i) = \begin{cases} x_i, & \text{if } x_i \geqslant 0, \\ 2x_i, & \text{if } x_i < 0. \end{cases}$$

The constant loss aversion model assumes that the utility of a product (x_1, x_2) is equal to $v(x_1 - r_1) + v(x_2 - r_2)$, with (r_1, r_2) being an exogenously given reference point. Let $(0, 0)$ be the reference point for simplicity. We can write the constant loss aversion model as the following equivalent COLU representation, $\min\{x_1 + x_2, 2x_1 + x_2, x_1 + 2x_2, 2x_1 + 2x_2\}$. This COLU representation provides a new interpretation of the constant loss aversion model: The decision maker has multiple ways in mind to evaluate a product, and she adopts the most cautious way.[9]

Viewing the constant loss aversion model in this way, it is clear that the decision maker's caution does not need to be so extreme. For example, the decision maker may dislike losses, but she may also like specialization.[10] In other words, she may appreciate a product relatively more if it excels in one attribute, rather than being mediocre in both attributes. This idea can be captured by the following modification of the previous COLU representation: $\max\{\min\{2x_1 + x_2, 2x_1 + 2x_2\}, \min\{x_1 + 2x_2, 2x_1 + 2x_2\}\}$. In this example, when both x_1 and x_2 are positive, the decision maker uses $\max\{2x_1 + x_2, x_1 + 2x_2\}$ to evaluate the product,

[9] Similar observations have also been made by Cerreia-Vioglio, Dillenberger, and Ortoleva (2022).

[10] There is abundant evidence for product specialization; one obvious reason is that consumers like products with salient good attributes. See, among others, Arndt and Kierzkowski (2001).

which reflects a preference for specialized products. Otherwise, the decision maker uses $2x_1 + 2x_2$ to evaluate the product, which, relative to $\max\{2x_1 + x_2, x_1 + 2x_2\}$, implies loss aversion.[11]

6 Neural-network Utility

A recent paper by Arora et al. (2018) from the computer science literature shows that feed-forward neural-network functions with rectified linear units are identical to CFPL functions. Hence, our Theorem 2 also implies that a preference has a COLU representation if and only if it has the following representation—which we will call the neural-network utility (NU) representation—that features one of the most powerful ingredients of machine learning, neural networks.

Given any vector-valued function τ, we use $\tau^{(j)}$ to denote the j-th component of τ.

Definition 3 *A function $U : X \to \mathbb{R}$ is an NU if there exist*

(i) $h, w_0, \ldots, w_{h+1} \in \mathbb{N}$ with $w_0 = n$ and $w_{h+1} = 1$, and

(ii) affine functions $\tau_i : \mathbb{R}^{w_{i-1}} \to \mathbb{R}^{w_i}$, $i = 1, \ldots, h+1$, such that for any $x \in X$,

$$U(x) = \tau_{h+1} \circ \theta \circ \tau_h \circ \cdots \circ \theta \circ \tau_2 \circ \theta \circ \tau_1(x), \tag{2}$$

in which θ is an entry-wise operation such that for any $w \in \mathbb{N}$ and $b \in \mathbb{R}^w$, we have $\theta(b) = (\max\{b_i, 0\}, \ldots, \max\{b_w, 0\})$.

Each function $\theta \circ \tau_i$ is called the i-th *hidden layer*, and $(\theta \circ \tau_i)^{(j)} = \max\{\tau_i^{(j)}(\cdot), 0\}$ is called a *neuron*.[12] Thus, the i-th hidden layer has w_i neurons, and equation (2) characterizes a network of neurons with h hidden layers. Figure 3 provides an example of an NU function.

Mathematically, to evaluate an alternative x, each neuron in the NU function first aggregates its child neurons' values in an affine fashion. If the outcome of aggregation is above the normalized threshold, zero, this neuron is activated and its value becomes the outcome of aggregation. Otherwise, this neuron remains inactive and has a value zero. It is without loss of generality to normalize all thresholds to 0, since we can add an arbitrary constant to each affine function $\tau_i^{(j)}$ and modify the threshold accordingly. The neurons in the first

[11] For example, when x_1 is in the gain region but x_2 is in the loss region, $\max\{2x_1+x_2, x_1+2x_2\} = 2x_1+x_2$. Therefore, $2x_1 + 2x_2$ puts more weight on the loss, x_2, in this case.

[12] The entry-wise operation θ, called the *activation function*, may take other functional forms in general. However, the form we assume in Definition 3, also known as the rectified linear unit, is considered to be the most popular activation function and to have strong biological motivations (see Hahnloser, Sarpeshkar, Mahowald, Douglas, and Seung (2000); and LeCun, Bengio, and Hinton (2015), among others).

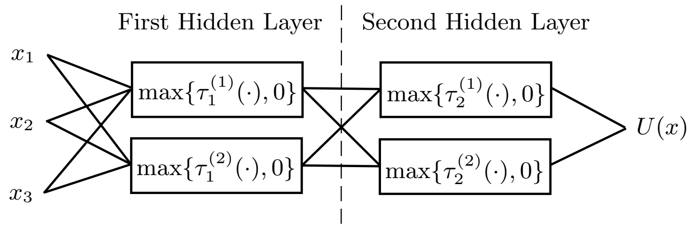

FIGURE 3: Consider an alternative (x_1, x_2, x_3). This NU function has two hidden layers, and each layer has two neurons. Each affine $\tau_1^{(j)}$ is from the choice domain (a subset of $\mathbb{R}^3$) to $\mathbb{R}$, and each affine $\tau_2^{(j)}$ is from $\mathbb{R}^2$ to $\mathbb{R}$. Neurons in the first layer are called *child neurons* of neurons in the second layer. Neurons in the second layer are called *parent neurons* of neurons in the first layer.

hidden layer aggregate the input of the NU function, x_i's, directly, and the values of neurons in the last (h-th) hidden layer are aggregated into the utility of x.

To understand the economic interpretation of the NU representation, think of X as a space of products characterized by their attributes. An affine function on X describes a way to evaluate the product. A decision maker whose preference has an NU representation entertains multiple such ways to evaluate a product. This is captured by the affine functions of the first-hidden-layer neurons. What the first hidden layer achieves is that the decision maker uses the primitive attributes to form multiple more advanced subjective attributes. For example, she might combine several primitive attributes of a car to form a subjective safety attribute. Some of these subjective attributes may be active and some may not be. The decision maker may continue to consider multiple ways (captured by the affine functions of the second-hidden-layer neurons) to aggregate those active subjective attributes. This allows her to form subjective attributes that are even more advanced. This process continues until she aggregates the values of last-hidden-layer neurons to obtain the evaluation of the product.

If X is the probability simplex, the decision maker considers multiple expected utility functions to evaluate the uncertainty of a lottery, which corresponds to the affine functions of the first hidden layer. For instance, she may have one neuron that activates when the expected value of prizes is high, and another that activates whenever the downside risk is high. With multiple risk attitudes, the decision maker wants to aggregate them, and she may not have a unique view about how to do so. This is captured by the second hidden layer. The aggregation continues until the decision maker reaches a final evaluation of the lottery.

The next result follows from our Theorem 2 and Arora et al. (2018).[13]

[13] Arora et al. (2018) adopt a slightly different definition of CFPL functions. In particular, each linear region is assumed to be a polyhedron, i.e., the intersection of finitely many closed half-spaces. This definition is equivalent to our definition since every CFPL function, according to our definition, can be written in the COLU form.

Corollary 3 *The preference $\succsim$ has an NU representation if and only if $\succsim$ satisfies weak order, continuity, and weak local bi-independence.*

The NU representation has two advantages. First, we have efficient empirical methods to estimate a neural-network model, and the same applies to the NU model. This is not true for the differentiable utility function suggested by Machina (1982). In fact, given the substantial evidence of the practical success of neural-network models when paired with a large amount of data, we can expect that the NU model will offer us a superior method to model people's choice behavior as we obtain more and more choice data.

Second, it is convenient to use the NU representation to construct special cases that captures well-known behavioral effects, which is again something not easy to do with a differentiable utility function. We show this in the next subsections.

6.1 NU and the Certainty Effect

We can construct neurons that capture behavioral effects in an NU representation. From the Allais paradox, we know that decision makers are often biased toward certainty. Consider the following example in which X is the set of lotteries over three prizes. Figure 4 presents an NU representation in which the first neuron captures standard expected utility evaluation, while the other three neurons capture the bias toward certainty for the three prizes, respectively.[14] A certainty-effect neuron for a prize is activated if and only if the lottery yields that prize with high probability.

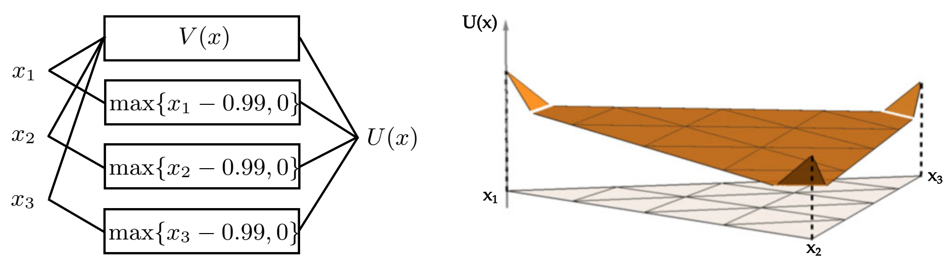

FIGURE 4: In the first neuron, V is affine, which does not seem to satisfy the requirement of a neuron, because it does not compare an affine function with zero—but this is for simplicity and without loss of generality, since $V(x) = \max\{V(x), 0\} - \max\{-V(x), 0\}$. If the probability of the ith prize is larger than 0.99 for some $i \in \{1, 2, 3\}$, a neuron that captures the bias toward certainty will be activated. Finally, $U(x)$ is equal to some weighted sum of all neurons' values.

[14] This function appears in Chapter 2.4.4.2 of Schmidt (1998), although its connection to neural-network models is not explored. We thank David Dillenberger for pointing this out.

6.2 NU and Subjective Attributes

Consider products described by the values of their attributes. A decision maker may have her own perception of what the product's attributes are. For example, suppose X is the set of all electric scooters. Each scooter is described by a speed-related attribute x_1, a steering-related attribute x_2, a brake-related attribute x_3, and a battery-related attribute x_4, in which $x_i \in [-1, 1]$, $i = 1, \ldots, 4$. The following NU representation describes a decision maker who uses x_1 and x_2 to form a subjective attribute about the performance of the scooter and uses x_1, x_2, and x_3 to form a subjective attribute about the safety of the scooter. Then, she combines the performance attribute and the safety attribute, if active, to form a more advanced attribute about the overall riding experience of the scooter. Finally, this attribute is combined with x_4 to form the final evaluation of the scooter.

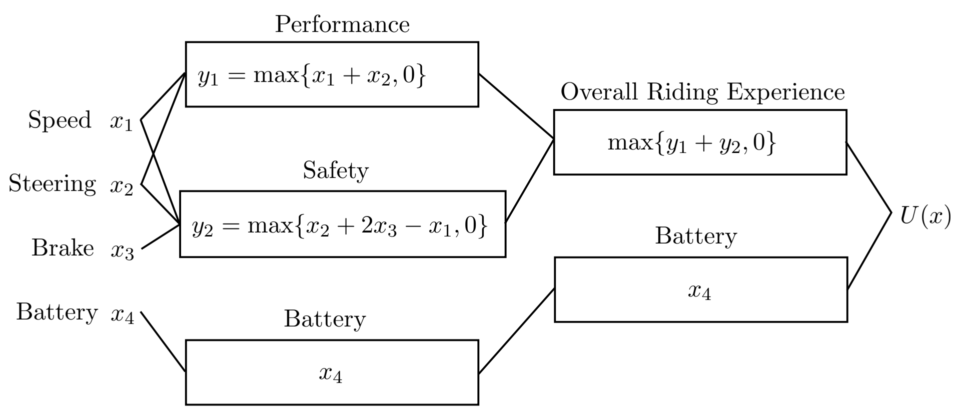

Figure 5: An example in which the decision maker forms hierarchical subjective attributes to evaluate an electric scooter.

7 Concluding Remarks

This paper introduces two notions of local linearity that are defined based on a decision maker's choice behavior, weak local independence and weak local bi-independence. Weak local independence implies that generically, the decision maker's preference does not have any differentiable representations. The stronger notion of local linearity, weak local bi-independence, characterizes CFPL functions. We introduce two new representations of the decision maker's preference that are equivalent to CFPL functions, the COLU representation and the NU representation. Our approach has several advantages over Machina's (1982) approach.

Appendix

Proof of Theorem 1

Proof. We first establish some useful lemmas. In what follows, stating this explicitly, we assume for each lemma that $\succsim$ satisfies weak order and continuity. Since X is separable and connected, by Debreu (1954) it must have a continuous utility representation $V : X \to \mathbb{R}$.

For any $L \subseteq \mathcal{X}$, let $\text{int}(L)$, $\text{cl}(L)$, ∂L, $\text{aff}(L)$, $\dim(L)$ denote the interior, closure, boundary, affine hull, and the dimension of the affine hull of L, respectively, in $\mathbb{R}^N$. For $x \in X$ and $\varepsilon > 0$, let $B_\varepsilon(x)$ denote the open ball centered at x with radius ε. For any finite set of choice alternatives $\{x^1, \ldots, x^m\}$, let $\overline{x^1 \ldots x^m} := \text{co}(\{x^1, \ldots, x^m\})$ be the convex hull of $\{x^1, \ldots, x^m\}$.

For any $L \in X$ and $L \subseteq X$, we write $L \perp z$ if for any $x, y \in L$ and $\lambda \in (0,1)$, $x \succsim y \Leftrightarrow \lambda xz \succsim \lambda yz$. Note that in this definition, λxz and λyz do not have to be in L.

The first two lemmas are straightforward. We omit their proofs.

Lemma 1 *For any $L \subseteq X$ such that int$(L) \neq \emptyset$, the following statements are equivalent:*

(i) L preserves independence.

(ii) cl(L) preserves independence.

(iii) If $x, y, r, s \in L$ and $x - y = \lambda(r - s)$ for some $\lambda > 0$, $x \succsim y \iff r \succsim s$.

(iv) There exists an affine function $U :$ cl$(L) \to \mathbb{R}$ that represents $\succsim$ on cl(L).

Lemma 2 *For any convex $L \subseteq X$ and $x \in L$, if $L \perp x$ and $L \perp y$, then $L \perp \alpha xy$ for any $\alpha \in (0, 1)$.*

Now we present a lemma that is key in our construction of the regular closed pieces.

Lemma 3 *For any convex $L \subseteq X$ such that int$(L) \neq \emptyset$, if L preserves independence and $L \perp x$, then co$(L \cup \{x\})$ preserves independence.*

Proof. We first prove that $\text{co}(\text{int}(L) \cup \{x\}) \backslash \{x\}$ preserves independence. Take any y in the set. There exists $y' \in \text{int}(L)$ and $\alpha \in (0, 1]$ such that $y = \alpha y' x$. Since $y' \in \text{int}(L)$, there exists some $\varepsilon > 0$ such that $B_\varepsilon(y') \subseteq \text{int}(L)$. For any $r, s \in B_{\alpha\varepsilon}(y)$, let $r' = \frac{r-y}{\alpha} + y'$, $s' = \frac{s-y}{\alpha} + y'$. Since $||r - y||, ||s - y|| < \alpha\varepsilon$, we have $r', s' \in B_\varepsilon(y')$ and $\alpha(r' - s') = r - s$. Moreover,

$$\alpha r' x = r - y + \alpha y' + (1 - \alpha)x = r - y + y = r.$$

Similarly, $\alpha s' x = s$. Because $L \perp x$, $r \succsim s \iff r' \succsim s'$. The above arguments show that for any $y \in \text{co}(\text{int}(L) \cup \{x\}) \backslash \{x\}$, there exists $\varepsilon > 0$ such that for any $r, s \in B_\varepsilon(y)$, we can find $r', s' \in \text{int}(L)$ such that $r - s = \alpha(r' - s')$ for some $\alpha > 0$ and $r \succsim s \iff r' \succsim s'$.

By Lemma 1, to show that $\text{co}(\text{int}(L) \cup \{x\}) \backslash \{x\}$ preserves independence, we only need to show that for any $r, s, r^*, s^* \in \text{co}(\text{int}(L) \cup \{x\}) \backslash \{x\}$ such that $r - s = \lambda(r^* - s^*)$ for some $\lambda > 0$, $r \succsim s$ if and only if $r^* \succsim s^*$.

First, focus on r and s. Clearly, $\overline{rs} \subseteq \text{co}(\text{int}(L) \cup \{x\}) \backslash \{x\}$. For any $t \in \overline{rs}$, according to the arguments above, there exists $\varepsilon_t > 0$ such that for any $\tilde{r}_t, \tilde{s}_t \in B_{\varepsilon_t}(t)$, we can find $\tilde{r}'_t, \tilde{s}'_t \in \text{int}(L)$ that satisfy $\tilde{r}_t - \tilde{s}_t = \alpha(\tilde{r}'_t - \tilde{s}'_t)$ for some $\alpha > 0$ and $\tilde{r} \succsim \tilde{s} \iff \tilde{r}' \succsim \tilde{s}'$.

Note that $\{B_{\varepsilon_t}(t) : t \in \overline{rs}\}$ forms an open cover of $\overline{rs}$. Since $\overline{rs}$ is compact, let the Lebesgue number of the open cover be $\rho > 0$ and define

$$t_k := r + \frac{\min\{k\rho, ||s - r||\}}{||s - r||}(s - r)$$

for $k = 0, 1, \ldots, \min\{j \in \mathbb{N} : \rho j \geqslant ||s - r||\}$. Let $m := \min\{j \in \mathbb{N} : \rho j \geqslant ||s - r||\} - 1$. By definition, $t_0 = r$, $t_m = s$, and $||t_k - t_{k+1}|| < \rho$ for any $k \in \{0, \ldots, m\}$. Since ρ is the Lebesgue number of the open cover, for any $k \in \{0, \ldots, m\}$, there exists $t \in \overline{rs}$ such that $t_k, t_{k+1} \in B_{\varepsilon_t}(t)$. Therefore, there exist $r'_k, s'_k \in \text{int}(L)$ such that $t_k - t_{k+1} = \beta_k(r'_k - s'_k)$ for some $\beta_k > 0$ and $t_k \succsim t_{k+1} \iff r'_k \succsim s'_k$. Note that by construction, for any $k \in \{0, \ldots, m\}$, $t_k - t_{k+1} = \lambda_k(r - s)$ for some $\lambda_k > 0$, which implies that for any $k \in \{0, \ldots, m\}$, $r'_k - s'_k = \alpha_k(r'_0 - s'_0)$ for some $\alpha_k > 0$.

Since L preserves independence, by Lemma 1, for any $k \in \{0, \ldots, m\}$, $r'_k \succsim s'_k \iff r'_0 \succsim s'_0$. It follows that $r'_0 \succsim s'_0 \iff t_k \succsim t_{k+1}$. Then, transitivity requires that $r \succsim s \iff r'_0 \succsim s'_0$. Note that $r - s = \frac{\beta_0}{\lambda_0}(r'_0 - s'_0)$.

The same arguments apply to r^* and s^*: There exist some $r^*_0, s^*_0 \in \text{int}(L)$ such that $r^* \succsim s^* \iff r^*_0 \succsim s^*_0$ and $r^* - s^* = \lambda^*(r^*_0 - s^*_0)$ for some $\lambda^* > 0$. Since $r - s = \lambda(r^* - s^*)$, we know that $r'_0 - s'_0 = \alpha^*(r^*_0 - s^*_0)$ for some $\alpha^* > 0$. By Lemma 1, $r'_0 \succsim s'_0 \iff r^*_0 \succsim s^*_0$. Thus, $r \succsim s \iff r^* \succsim s^*$. This completes the proof that $\text{co}(\text{int}(L) \cup \{x\}) \backslash \{x\}$ preserves independence.

It is straightforward to verify that

$$\text{cl}(\text{co}(\text{int}(L) \cup \{x\}) \backslash \{x\}) = \text{cl}(\text{co}(\text{int}(L) \cup \{x\})).$$

Hence, by Lemma 1, $\text{cl}(\text{co}(\text{int}(L) \cup \{x\}))$ preserves independence.

To complete the proof of this lemma, we only need to show that $\text{co}(L \cup \{x\})$ is a subset of $\text{cl}(\text{co}(\text{int}(L) \cup \{x\}))$. Since L is convex and has nonempty interior, $\text{cl}(\text{int}(L)) = \text{cl}(L)$. To see this, take any $y \in \text{cl}(L)$ and let $\{y_k\}_{k=1}^{\infty}$ be some sequence in L that converges to y. Take any $r \in \text{int}(L)$. Since L is convex, the sequence $\{(1 - \frac{1}{k})y_k r\}_{k=1}^{\infty}$ is a sequence in $\text{int}(L)$ that converges to y as well. Therefore, $y \in \text{cl}(\text{int}(L))$. Because $\text{cl}(\text{int}(L)) = \text{cl}(L)$, for any $y \in L$, let $\{y_k\}_{k=1}^{\infty}$ be some sequence in $\mathrm{int}(L)$ that converges to y. Then, for any $\alpha \in [0,1]$, $\alpha y_k x$ converges to $\alpha y x$, which implies that $\alpha y x \in \mathrm{cl}(\mathrm{co}(\mathrm{int}(L) \cup \{x\}))$. Thus, $\mathrm{co}(L \cup \{x\}) \subseteq \mathrm{cl}(\mathrm{co}(\mathrm{int}(L) \cup \{x\}))$. ■

Now we are ready to prove Theorem 1. Suppose that in addition to weak order and continuity, $\succsim$ also satisfies weak local independence. Thus, for each $x \in X$, there exists $\varepsilon_x > 0$ such that $B_{\varepsilon_x}(x) \perp x$. Hereafter, for any $x \in X$, let $B_x = B_{\varepsilon_x}(x)$. We proceed to construct a polytope that preserves independence for each $x \in X$.

Lemma 4 *Suppose $\succsim$ satisfies weak local independence. Then for any $x \in X$, there exist $x^1, \ldots, x^N \in X$ such that $\overline{xx^1 \ldots x^N}$ preserves independence and has nonempty interior.*

Proof. Let $x^0 := x \in X$. Then, recursively for $i = 1, \ldots, N$, let x^i be an arbitrary point in $(\bigcap_{j<i} B_{x^j}) \backslash \mathrm{aff}(\{x^0, \ldots, x^{i-1}\})$. Since each B_{x^i} is open and $\mathrm{aff}(\{x^0, \ldots, x^{i-1}\})$'s dimension is at most $i-1$, such x^i's always exist. By construction, the dimension of $\Delta := \overline{xx^1 \ldots x^N}$ is equal to N, the dimension of X, and Δ has nonempty interior.

Pick some $\alpha \in (0,1)$ such that for any $j = 0, \ldots, N-1$, $y^j := \alpha x^N x^j \in \bigcap_{i=0}^{N} B_{x^i}$. Clearly, $\Delta' = \overline{y^0 \ldots y^{N-1} x^N}$ also has nonempty interior. In addition, by construction, $\Delta' \perp x^i$ for $i = 0, \ldots, N$ because $\Delta' \subseteq \bigcap_{i=0}^{N} B_{x^i}$. Since $x^N \in \Delta'$, it follows from Lemma 2 that $\Delta' \perp y^i$ for $i = 0, \ldots, N-1$. Applying Lemma 2 again, we know that $\Delta' \perp y$ for any $y \in \Delta'$, which implies that Δ' preserves independence. Then, applying Lemma 3 iteratively, we know that $\mathrm{co}(\Delta' \cup \{x^{N-1}\})$ preserves independence, $\mathrm{co}(\Delta' \cup \{x^{N-1}, x^{N-2}\})$ preserves independence, and so on. Since $\Delta = \mathrm{co}(\Delta' \cup \{x^0, \ldots, x^{N-1}\})$, Δ preserves independence. ■

It follows directly from Lemma 4 that $\succsim$ exhibits piecewise independence. To show the second statement of Theorem 1, the next step is to identify the "largest" pieces that induce the same local utility function using Zorn's Lemma. Let $\mathcal{D}$ be the collection of all possible polytopes constructed using the procedure in Lemma 4. Let $X_o := \bigcup_{\Delta \in \mathcal{D}} \mathrm{int}(\Delta)$. It is clear that

$$X = \bigcup_{\Delta \in \mathcal{D}} \Delta = \bigcup_{\Delta \in \mathcal{D}} \mathrm{cl}(\mathrm{int}(\Delta)) \subseteq \mathrm{cl}\left(\bigcup_{\Delta \in \mathcal{D}} \mathrm{int}(\Delta)\right) = \mathrm{cl}(X_o) \subseteq X.$$

Thus, X_o is an open and dense subset of X. For any $x \in X_o$, pick $\Delta_x \in \mathcal{D}$ such that $x \in \mathrm{int}(\Delta_x)$.

For any nonempty open subset $Y \subseteq X$, we say that Y *induces local utility function* U if any $x \in Y$ has a neighborhood $L \subseteq Y$ such that U represents $\succsim$ on L. Note that when Y is not convex, the fact that Y induces U does not imply U represents $\succsim$ on Y. By definition, if Y_i induces local utility function U_i for $i = 1, 2$ and $Y_1 \cap Y_2 \neq \emptyset$, then $U_1 = U_2$.

Let $\mathcal{O} := \{L \subseteq X : L \text{ is nonempty, connected, and open}\}$. Let $\mathcal{F}$ be the set of all functions $P : X_o \to \mathcal{O}$ such that for any $x \in X_o$, (i) $\mathrm{int}(\Delta_x) \subseteq P(x)$, and (ii) $P(x)$ induces some local utility function. Clearly, $\mathcal{F}$ is nonempty, since for any $x \in X_0$ we can simply let $P(x) = \text{int}(\Delta_x)$.

Define a binary relation $\Subset$ on $\mathcal{F}$ as follows: For any $x, y \in \mathcal{F}$, $P \Subset Q$ if for any $x \in X_o$, $P(x) \subseteq Q(x)$. It is straightforward to verify that $\Subset$ is a partial order on $\mathcal{F}$. Take any totally ordered subset of $\mathcal{F}$, $\{P_i\}_{i \in I}$, in which I is an index set. Let $P^* : X_o \to \mathcal{O}$ be a function such that for any $p \in X_o$, $P^*(x) := \bigcup_{i \in I} P_i(x)$. We show that $P^* \in \mathcal{F}$. First of all, $P^*(x)$ is open since every $P_i(x)$ is open. Second, $P^*(x)$ is connected, since every $P_i(x)$ is connected and contains $\text{int}(\Delta_x)$, which is connected. Furthermore, each $P_i(x)$ induces the same local utility function as $\text{int}(\Delta_x)$ does. Thus, $P^*(x)$ also induces the same local utility function as $\text{int}(\Delta_x)$ does. Hence, P^* is an upper bound of $\{P_i\}_{i \in I}$ in terms of $\Subset$.

Now, we can apply Zorn's lemma and know that $\mathcal{F}$ contains some $\Subset$-maximal element. With a harmless abuse of notation, denote this $\Subset$-maximal element by P^*. Next we show that $\{x \in X_o : P^*(x)\}$ is finite.

For any $x \in X$ and $\varepsilon > 0$, let $C_\varepsilon(x) := \{y \in X : ||x - y||_\infty < \varepsilon\}$; that is, $C_\varepsilon(x)$ is the open hypercube that is centered at x with edge length 2ε. For any $x \in X$ and nonempty $L \subseteq X$, let $\text{cone}_x(L) := \{y \in X : y = x + \alpha(z - x) \text{ for some } \alpha \geqslant 0 \text{ and } z \in L\}$; that is, $\text{cone}_x(L)$ is the smallest cone with vertex x that contains L.

Lemma 5 *Suppose $\succsim$ satisfies weak local independence and P^* is a $\Subset$-maximal element. Then $\{x \in X_o : P^*(x)\}$ is finite.*

Proof. Suppose $\mathcal{P} = \{x \in X_o : P^*(x)\}$ is not finite. Let $\mathcal{P} = \{P_i\}_{i \in I}$. Suppose I is not finite. Pick a countable subset of P_i's and form a sequence with one choice alternative in each. By the compactness of X, we can pick an accumulation point of the sequence, denoted as x. It is clear that any neighborhood of x intersects P_i for an infinite number of i's in I. Fix some $\varepsilon > 0$ such that $C_\varepsilon(x) \subseteq B_x$, in which B_x is where the second vertex is chosen in the procedure for constructing the polytope for x in Lemma 4. Let $J := \{i \in I : C_\varepsilon(x) \cap P_i \neq \emptyset\}$ and $Q_i := C_\varepsilon(x) \cap \text{cl}(P_i)$ for each $i \in J$. It is easy to verify by the maximality of P^* that $Q_j \cap P_i = \emptyset$ for any $i \in I, j \in J$ with $i \neq j$.

First, we show that for any $y \in C_\varepsilon(x)$ with $x \neq y$, there exists $i \in J$ such that $\overline{xy} \subseteq Q_i$. Suppose $x \neq y$. Applying the procedure in Lemma 4, we can construct $\Delta \in \mathcal{D}$ such that $\overline{xy} \subseteq \Delta$. By the denseness of X_o, there exists $z \in X_o$ such that $P^*(z) \cap \text{int}(\Delta) \neq \emptyset$. It follows that $\text{int}(\Delta)$ induces the same local utility function as $P^*(z)$ does. Then the maximality of P^* implies that $\text{int}(\Delta) \subseteq P^*(z)$. It follows that $\overline{xy} \subseteq \Delta \subseteq Q_i$ for some $i \in J$.

Second, we show that $Q_i = C_\varepsilon(x) \cap \text{cone}_x(Q_i)$ for all $i \in J$. Loosely speaking, the intersection of Q_i and the hypercube must be a cone with vertex x. On one hand, by definition, $Q_i \subseteq C_\varepsilon(x)$ and $Q_i \subseteq \text{cone}_x(Q_i)$, which imply $Q_i \subseteq C_\varepsilon(x) \cap \text{cone}_x(Q_i)$. On the other hand, let $y \in C_\varepsilon(x) \cap \mathrm{cone}_x(C_\varepsilon(x) \cap P_i)$ for some $i \in J$. Then there exists $r \in C_\varepsilon(x) \cap P_i$ such that $y \in \overline{xr}$ or $r \in \overline{xy}$. Since $r \in C_\varepsilon(x)$, there exists $j \in J$ such that $\overline{xr} \subseteq Q_j$. By $r \in P_i$, we have $j = i$ and thus $\overline{xr} \subseteq Q_i$. If $y \in \overline{xr}$, then $y \in Q_i$. If $r \in \overline{xy}$, then since $r \in \overline{xy} \cap P_i \neq \emptyset$, we have $\overline{xy} \subseteq Q_i$. In both cases, $y \in Q_i$. Thus, $C_\varepsilon(x) \cap \mathrm{cone}_x(C_\varepsilon(x) \cap P_i) \subseteq Q_i \subseteq \mathrm{cl}(P_i)$. It follows that $C_\varepsilon(x) \cap \mathrm{cone}_x(Q_i) \subseteq \mathrm{cl}(C_\varepsilon(x) \cap \mathrm{cone}_x(C_\varepsilon(x) \cap P_i)) \subseteq \mathrm{cl}(P_i)$. Hence, $C_\varepsilon(x) \cap \mathrm{cone}_x(Q_i) \subseteq C_\varepsilon(x) \cap \mathrm{cl}(P_i) = Q_i$.

Now let $P_i^o := \mathrm{int}(\mathrm{cl}(P_i))$ for each $i \in I$. Note that P_i and P_i^o may not be the same set. The next step is to show that $P_i^o \cap P_j^o = \emptyset$ for any $i \neq j$. Since $P_i \subseteq \mathrm{cl}(P_i)$ and P_i is open, $P_i \subseteq P_i^o$. Suppose $P_i^o \cap P_j^o \neq \emptyset$. Then there exist $r \in X$ and $\delta > 0$ such that $B_\delta(r) \in P_i^o \cap P_j^o \subseteq \mathrm{cl}(P_i) \cap \mathrm{cl}(P_j)$. Since $B_\delta(r) \subseteq \mathrm{cl}(P_i)$ and P_i is open, we can find an open ball $B \subseteq B_\delta(r)$ such that $B \subseteq P_i$. Again, since $B \subseteq \mathrm{cl}(P_j)$ and P_j is open, we can find an open ball $B' \subseteq B$ such that $B' \subseteq P_j$. This is a contradiction, since $P_i \cap P_j = \emptyset$. Note that for any $i \neq j$, since $P_i^o \cap P_j^o = \emptyset$, $P_i^o \cap \mathrm{cl}(P_j) = \emptyset$.

We are now ready to present the main induction argument. Let $\varepsilon_1 := \frac{\varepsilon}{2}$ and $C_1 := \partial C_{\varepsilon_1}(x)$; that is, C_1 is the surface of hypercube $C_{\varepsilon_1}(x)$. Clearly, $C_1 \subseteq C_\varepsilon(x)$. Note that $C_1 \cap P_i^o \neq \emptyset$ for all $i \in J$. To see that, consider any $i \in J$ and $y \in C_\varepsilon(x) \cap P_i \subseteq C_\varepsilon(x) \cap P_i^o$ with $y \neq x$. We have $\mathrm{cone}_x(\{y\}) \cap C_\varepsilon(x) \subseteq Q_i = C_\varepsilon(x) \cap \mathrm{cone}_x(Q_i)$. Moreover, since $y \in C_\varepsilon(x) \cap P_i^o = \mathrm{int}(Q_i)$,

$$\mathrm{cone}_x(\{y\}) \cap C_1 \subseteq \mathrm{int}(C_\varepsilon(x) \cap \mathrm{cone}_x(Q_i)) = \mathrm{int}(Q_i) = C_\varepsilon(x) \cap P_i^o.$$

Since $\mathrm{cone}_x(\{y\}) \cap C_1 \neq \emptyset$, we have $C_1 \cap P_i^o \neq \emptyset$ for any $i \in J$.

When $N = 1$, C_1 contains at most two points, which cannot intersect with an infinite number of disjoint open sets—a contradiction. Hereafter, we assume $N \geqslant 2$.

Let A_1 be a face of C_1 that intersects P_i^o for an infinite number of i's in J. By the compactness of A_1, there exists $x^1 \in A_1$ such that if L is a neighborhood of x^1, then $L \cap A_1$ intersects P_i^o for an infinite number of i's in J. Now pick $\varepsilon' < \varepsilon_1$ such that $C_{\varepsilon'}(x^1) \subseteq B_{x^1}$. Let $J_1 := \{i \in J : (C_{\varepsilon'}(x^1) \cap A_1) \cap P_i^o \neq \emptyset\}$ and $y_i^1 := C_{\varepsilon'}(x^1) \cap \mathrm{cl}(P_i)$ for all $i \in J_1$.

Let $\varepsilon_2 := \frac{\varepsilon'}{2}$ and $C_2 := \partial C_{\varepsilon_2}(x^1) \subseteq C_{\varepsilon'}(x^1)$. We show that $(C_2 \cap A_1) \cap P_i^o \neq \emptyset$ for all $i \in J_1$. Following the same logic as above, for any $y \in C_{\varepsilon'}(x^1) \cap A_1$, there exists $i \in J_1$ such that $\overline{x^1 y} \subseteq y_i^1$. Furthermore, $y_i^1 = C_{\varepsilon'}(x^1) \cap \mathrm{cone}_{x^1}(y_i^1)$. Consider any $i \in J_1$ and $y \in (C_{\varepsilon'}(x^1) \cap A_1) \cap P_i^o$ with $y \neq x^1$. Since $y \in C_{\varepsilon'}(x^1) \cap P_i^o = \mathrm{int}(y_i^1)$,

$$\mathrm{cone}_{x^1}(\{y\}) \cap C_2 \subseteq \mathrm{int}(C_{\varepsilon'}(x^1) \cap \mathrm{cone}_x(y_i^1)) = \mathrm{int}(y_i^1) = C_{\varepsilon'}(x^1) \cap P_i^o.$$

Since $\emptyset \neq \mathrm{cone}_{x^1}(\{y\}) \cap C_2 \subseteq A_1$, we have $(C_2 \cap A_1) \cap P_i^o \neq \emptyset$.

Now let A_2 be a face of C_2 such that $A_1 \cap A_2$ intersects P_i^o for an infinite number of

i's in J_1. By the compactness of $A_1 \cap A_2$, there exists $x^2 \in A_1 \cap A_2$ such that if L is a neighborhood of x^2, then $L \cap A_1 \cap A_2$ intersects P_i^o for an infinite number of i's in J_1. Inductively, for any k, there exists $x^k \in A_1 \cap \cdots \cap A_k$ such that if L is a neighborhood of x^k, then $L \cap A_1 \cap \cdots \cap A_k$ intersects P_i^o for an infinite number of i's in J. Let $k = N-1$. Then $A_1 \cap \cdots \cap A_{N-1}$ is a line segment that intersects P_i^o for $i \in J_{N-1}$ with $|J_{N-1}| = \infty$. Pick x^{N-1} in a way similar to before. Note that there exists $\hat{\varepsilon} > 0$ small enough such that $C_{\hat{\varepsilon}}(x^{N-1}) \cap P_i^o = C_{\hat{\varepsilon}}(x^{N-1}) \cap \mathrm{cone}_{x^{N-1}}(C_{\hat{\varepsilon}}(x^{N-1}) \cap P_i^o)$ for any $i \in J_{N-1}$. Since $A_1 \cap \cdots \cap A_{N-1}$ is a line segment, it follows that $C_{\hat{\varepsilon}}(x^{N-1}) \cap (A_1 \cap \cdots \cap A_{N-1})$ intersects P_i^o for at most two i's, which is a contradiction. ■

Finally, since each $P^*(x)$ induces a single local utility function and any regular closed subset of X must contain some $x \in X_0$, we establish the second statement of Theorem 1, ■

Proof of Proposition 1

Proof. Suppose $\succsim$ satisfies weak order, continuity, and weak local independence, and has a differentiable representation $V : X \to \mathbb{R}$. Let P^* be the mapping identified in the proof of Theorem 4. By Lemma 5, let $\{x \in X_o : P^*(x)\} = \{P_i\}_{i=1}^n$ for some n and U_i be the non-constant local utility function induced by P_i for each i.

When $N = 1$, $X = [a, b]$ for real numbers $a < b$. The claim holds trivially, since the only possible non-constant local utility functions are $U(x) = 0, U(x) = \frac{x-a}{b-a}$, and $U(x) = 1 - \frac{x-a}{b-a}$ due to the normalization. Thus, we assume $N \geqslant 2$ hereafter.

Next, we prove the following lemma.

Lemma 6 *Suppose $N \geqslant 2$. Then for any open ball B in X, if $\dim(Y) \leqslant N-2$, then $B \backslash \mathit{aff}(Y)$ is connected.*

Proof. We will prove that $B \backslash \mathrm{aff}(Y)$ is path-connected. Pick any distinct $x, y \in B \backslash \mathrm{aff}(Y)$. We know that $\dim(Y \cup \{x\}) \leqslant N - 1$. Since X is a convex set with nonempty interior, any open ball in X is a convex set of dimension N. Thus there exists $z \in B \backslash \mathrm{aff}(Y \cup \{x\})$. Note that $\overline{xz} \cap \mathrm{aff}(Y \cup \{x\}) = \{x\}$. To see that, suppose there exists $z' \neq x$ such that $z' \in \overline{xz} \cap \mathrm{aff}(Y \cup \{x\})$. Then since $z \in \mathrm{aff}(\overline{xz'})$, it has to be the case that $z \in \mathrm{aff}(Y \cup \{x\})$, which is a contradiction. Suppose $y \in \mathrm{aff}(Y \cup \{x\})$. Then similarly we have $\overline{yz} \cap \mathrm{aff}(Y \cup \{x\}) = \{y\}$, and $\overline{xz} \cup \overline{yz}$ forms a path from x to y in $B \backslash \mathrm{aff}(Y)$. Suppose $y \notin \mathrm{aff}(Y \cup \{x\})$. Then we directly have $\overline{xy} \cap \mathrm{aff}(Y \cup \{x\}) = \{x\}$. Thus, $\overline{xy} \subseteq B \backslash \mathrm{aff}(Y)$ and we are done. ■

Pick any i and let $P = \bigcup\{P_j : U_j = U_i \text{ or } U_j = 1 - U_i\}$ and $Q = (\bigcup_{i=1}^n P_i) \backslash P$. By way of contradiction, assume that $Q \neq \varnothing$. By construction, P and Q are both open and $\mathrm{cl}(P \cup Q) = \mathrm{cl}(P) \cup \mathrm{cl}(Q) = X$. Since X is connected, we must have $\mathrm{cl}(P) \cap \mathrm{cl}(Q) \neq \varnothing$. Pick any $x^1 \in \mathrm{cl}(P) \cap \mathrm{cl}(Q)$ and let B_1 be the open ball centered at x^1 such that $B_1 \perp x^1$, given by weak local independence. We show that there exists $x^2 \in B_1$ such that $x^2 \neq x^1$ and $x^2 \in \mathrm{cl}(P) \cap \mathrm{cl}(Q)$. Suppose not. Then $B_1 \backslash \{x\} = ((B_1 \backslash \{x\}) \cap \mathrm{cl}(P)) \cup ((B_1 \backslash \{x\}) \cap \mathrm{cl}(Q))$ and $(B_1 \backslash \{x\}) \cap \mathrm{cl}(P) \cap \mathrm{cl}(Q)) = \emptyset$. This is a contradiction, since $B_1 \backslash \{x\}$ is connected when $N \geqslant 2$. Hence, we can pick $x^2 \in B_1$ such that $x^2 \neq x^1$ and $x^2 \in \mathrm{cl}(P) \cap \mathrm{cl}(Q)$.

Inductively, by Lemma 6, for $j = 2, \ldots, N$, we can pick $x^j \in B_{j-1} \backslash \mathrm{aff}(\overline{x^1 \cdots x^{j-1}})$ such that $x^j \in \mathrm{cl}(P) \cap \mathrm{cl}(Q)$ and $B_j \subseteq B_{j-1}$ such that $B_i \perp x^i$. By $x^N \in \mathrm{cl}(P) \cap \mathrm{cl}(Q)$, there exists $r \in P \cap B_N$ and $s \in Q \cap B_N$. Since P and Q are both open, we can assume, without loss of generality, that $r, s \notin \mathrm{aff}(\overline{x^1 \cdots x^N})$. Using the same argument in Lemma 4, $\Delta_1 = \overline{x^1 \cdots x^N r}$ and $\Delta_2 = \overline{x^1 \cdots x^N s}$ both preserve independence. By $r \in P$, either U_i or $1 - U_i$ represents $\succsim$ on Δ_1. We focus on the case in which U_i represents $\succsim$ on Δ_1, since the other case is symmetric. In addition, there exists a non-constant local utility function $U \notin \{U_i, 1 - U_i\}$ that represents $\succsim$ on Δ_2. Pick any z in the relative interior of $\overline{x^1 \cdots x^N}$; it is clear that

$$\dim(\{\tilde{x} \in \Delta_1 : U_i(\tilde{x}) = U_i(z)\} = \dim(\{\tilde{x} \in \Delta_2 : U(\tilde{x}) = U(z)\} = N - 1.$$

In addition, since U is a local utility function distinct from U_i and $1 - U_i$, the gradient of V at z, $\nabla V(z)$, which needs to be orthogonal to both $\{\tilde{x} \in \Delta_1 : U_i(\tilde{x}) = U_i(z)\}$ and $\{\tilde{x} \in \Delta_2 : U(\tilde{x}) = U(z)\}$, must be $\vec{0}$. It then follows from the continuity of V that $V(x) = V(y)$ for any $x, y \in \overline{x^1 \cdots x^N}$. Since both U_i and U are non-constant and represent $\succsim$ on $\overline{x^1 \cdots x^N}$, which is of dimension $N - 1$, it must be the case that $U = U_i$ or $U = 1 - U_i$, which is a contradiction. ■

Proof of Proposition 2

Proof. Let $\trianglerighteq$ be a preorder that satisfies betweenness and suppose each local utility function of $\succsim$ respects $\trianglerighteq$. Pick $x, y \in X$ such that $x \trianglerighteq y$. For any $z \in \overline{xy}$, let B_z be an open ball centered at z such that $B_z \perp z$, given by weak local independence. Thus, $\{B_z : z \in \overline{xy}\}$ forms an open cover of $\overline{xy}$, which is compact, and thus it has a finite subcover $\{B_{z^i}\}_{i=1}^m$. Let the Lebesgue's number of this finite subcover be $\delta > 0$. Pick $x^1, x^2, \ldots, x^k \in \overline{xy}$ such that $x^1 = x$, $x^m = y$ and $||x^j - x^{j+1}|| < \delta$ for $j = 1, \ldots, k-1$. By definition of the Lebesgue's number, for any $j = 1, \ldots, k-1$, there exists $i(j) \in \{1, \ldots, m\}$ such that $\overline{x^j x^{j+1}} \subseteq B_{z_{i(j)}}$. Thus, following the same procedure as in Lemma 4, for each j, we can construct two polytopes $\overline{z^{i(j)} r^1 \ldots r^N}$ and $\overline{z^{i(j)} s^1 \ldots s^N}$ with $r^1 = x^j$ and $s^1 = x^{j+1}$, both of which preserve independence. Since each local utility function of $\succsim$ respects $\trianglerighteq$, $\succsim$ restricted to $\overline{z^{i(j)} x^j}$ and $\succsim$ restricted to $\overline{z^{i(j)} x^{j+1}}$ both respect $\trianglerighteq$. Since $\bigcup_{j=1}^m \left(\overline{z^{i(j)} x^j} \cup \overline{z^{i(j)} x^{j+1}} \right) = \overline{xy}$, it is easy to see that $\succsim$ restricted to $\overline{xy}$ respects $\trianglerighteq$, and thus, $x \succsim y$. ■

Proof of Theorem 2

Proof. We will only show the sufficiency of the axioms. We prove this via a sequence of lemmas. We present the proof of the necessity in the Online Appendix.

In what follows, without stating this explicitly, we assume for each lemma that $\succsim$ satisfies weak order and continuity. Since X is separable and connected, by Debreu (1954) it must have a continuous utility representation $V : X \to \mathbb{R}$.

The first two lemmas are about the (bi-)independence of line segments. Lemma 7 characterizes when a line segment preserves independence. The proof for Lemma 7 is standard in the literature and is thus omitted. Lemma 8 characterizes when two line segments preserve bi-independence, provided that they each preserve independence. We provide the proof in the Online Appendix.

Lemma 7 *For any $x, r \in X$, $\overline{xr}$ preserves independence if and only if $\overline{xr} \perp x$.*

Lemma 8 *For any $x, y, r, s \in X$, if $\overline{xr}$ and $\overline{ys}$ each preserve independence, then the following statements are equivalent:*

(i) $\overline{xr}$ and $\overline{ys}$ preserve bi-independence.

(ii) For any $x', r' \in \overline{xr}$ and $y', s' \in \overline{ys}$ such that $x' \sim y'$ and $r' \sim s'$, $\lambda x' r' \sim \lambda y' s'$ for any $\lambda \in (0, 1)$.

(iii) There exists $\varepsilon > 0$ such that for any $y', s' \in \overline{ys}$ with $||y' - s'|| < \varepsilon$, $\overline{xr}$ and $\overline{y's'}$ preserve bi-independence.

Given x and y with $x \sim y$, we say that neighborhoods L_x and L_y *preserve weak bi-independence with respect to x and y* if for any $r \in L_x, s \in L_y$, and $\lambda \in (0, 1)$, $r \succsim s \Leftrightarrow \lambda xr \succsim \lambda ys$. We omit the phrase "with respect to x and y" from time to time when there is no risk of confusion.

We introduce one last lemma before constructing the linear regions.

Lemma 9 *Suppose $\succsim$ satisfies weak local bi-independence. Then for any x, there exists $\varepsilon > 0$ such that for any $r \in B_\varepsilon(x)$ and any convex regular closed subset L that preserves independence, $\overline{xr}$ and L preserve bi-independence.*

Proof. There is nothing to prove if $L = \emptyset$. Let $L \neq \emptyset$. By way of contradiction, suppose for any n there exist $r^n \in B_{1/n}(x)$ and a nonempty, convex, and regular closed L_n that preserves independence, such that $\overline{xr^n}$ and L_n do not preserve bi-independence. Thus, there exists $y^n, s^n \in L_n$ such that $\overline{xr^n}$ and $\overline{y^n s^n}$ do not preserve bi-independence. By weak local independence (implied by weak local bi-independence), for a sufficiently large n, $B_{1/n}(x) \perp x$, which implies that $\overline{xr^n}$ preserves independence by Lemma 7. Then by Lemma 8, there exist $\hat{x}^n, \hat{r}^n \in \overline{xr^n}$ and $\hat{y}^n, \hat{s}^n \in \overline{y^n s^n} \subseteq L_n$ such that $\hat{x}^n \sim \hat{y}^n$ and $\hat{r}^n \sim \hat{s}^n$, $||\hat{y}^n - \hat{s}^n|| < \frac{1}{n}$, and $\overline{\hat{x}^n \hat{r}^n}$ and $\overline{\hat{x}^n \hat{s}^n}$ do not preserve bi-independence.

It is clear that $\hat{x}^n, \hat{r}^n$ converges to x as n goes to infinity. Since X is compact, the sequence $\{\hat{y}^n\}$ has a subsequence that converges to some y. We assume, without loss of generality, that the subsequence is $\{\hat{y}^n\}$ itself and that $||\hat{y}^n - y||$ is monotonically decreasing in n. By continuity, $V(\hat{y}^n) = V(\hat{x}^n)$ for all n implies that $V(y) = V(x)$.

To derive the desired contradiction, we show that for any $\varepsilon, \delta > 0$, $B_\varepsilon(x)$ and $B_\delta(y)$ cannot preserve weak bi-independence. Fix any $\varepsilon, \delta > 0$. We can, without loss of generality, assume δ is small enough such that $B_\delta(y) \perp y$, because if $B_\varepsilon(x)$ and $B_\delta(y)$ cannot preserve weak bi-independence, $B_\varepsilon(x)$ and $B_{\delta'}(y)$ cannot preserve weak bi-independence for any $\delta' > \delta$.

Clearly, there exists m such that $||y - \hat{y}^m|| < \delta - \frac{1}{m}$. Then it follows that $||y - \hat{s}^m|| < \delta - \frac{1}{m} + \frac{1}{m} = \delta$. Hence $\hat{y}^m, \hat{s}^m \in B_\delta(y)$. There also exists k such that $n \geqslant k$ implies $\hat{x}^n, \hat{r}^n \in B_\varepsilon(x)$. Let $N = \max\{m, k\}$. Then $\hat{x}^N, \hat{r}^N \in B_\varepsilon(x)$, $\hat{y}^N, \hat{s}^N \in B_\delta(y)$. Since L_N is regular closed, $L_N = \mathrm{cl}(\mathrm{int}(L_N))$, which implies that $L_N \cap B_\delta(y)$ has nonempty interior. By Lemma 3, we have that $\mathrm{co}((L_N \cap B_\delta(y)) \cup \{y\})$ preserves independence. Thus, $\overline{\hat{y}^N \hat{s}^N y}$ preserves independence.

By construction, since $\hat{x}^N, \hat{r}^N \in \overline{xr^N}$, and $\overline{xr^N}$ preserves independence, we either have $\hat{x}^N, \hat{r}^N \succsim x$ or $x \succsim \hat{x}^N, \hat{r}^N$. The two cases are symmetric, so we will only consider the first case. Note that $\hat{x}^N \sim \hat{y}^N$ and $\hat{r}^N \sim \hat{s}^N$, and that $\overline{\hat{x}^N \hat{r}^N}$ and $\overline{\hat{y}^N \hat{s}^N}$ each preserve independence but do not preserve bi-independence. We either have $\hat{x}^N \succ \hat{r}^N$ or $\hat{r}^N \succ \hat{x}^N$. Again, the two cases are symmetric so we will only prove the first case. Hence, $\hat{x}^N \succ \hat{r}^N \succsim x$, and $\hat{r}^N \in \overline{\hat{x}^N x}$. It follows that $\hat{y}^N \succ \hat{s}^N \succsim y$. Then by continuity there exists $s \in \overline{\hat{y}^N y}$ such that $s \sim \hat{s}^N$. If $B_\varepsilon(x)$ and $B_\delta(y)$ preserve weak bi-independence, since $x \sim y$ and $\hat{x}^N \sim \hat{y}^N$, we have $\lambda \hat{x}^N x \sim \lambda \hat{y}^N y$ for all $\lambda \in [0, 1]$. Since $\hat{r}^N \in \overline{\hat{x}^N x}$, $s \in \overline{\hat{y}^N y}$, and $\hat{r}^N \sim s$, it follows that $\lambda \hat{x}^N \hat{r}^N \sim \lambda \hat{y}^N s$ for all $\lambda \in [0, 1]$. Note that for all $\lambda \in [0, 1]$, since $\overline{\hat{y}^N \hat{s}^N y}$ preserves independence, $\lambda \hat{y}^N s \sim \lambda \hat{y}^N \hat{s}^N$. It then follows that $\lambda \hat{x}^N \hat{r}^N \sim \lambda \hat{y}^N \hat{s}^N$ for all $\lambda \in [0, 1]$. Thus $\overline{\hat{x}^N \hat{r}^N}$ and $\overline{\hat{y}^N \hat{s}^N}$ preserve bi-independence, which is a contradiction. ■

Now we proceed to construct a polytope for each $x \in X$ using similar procedures as in Lemma 4.

Lemma 10 *Suppose $\succsim$ satisfies weak local bi-independence. Then for any $x^0 \in X$, there exist $x^1, \ldots, x^N \in X$ such that $\overline{x^0 x^1 \ldots x^N}$ preserves independence and has nonempty interior, and $\overline{x^i x^j}$ and L preserve bi-independence for any i, j and any convex regular closed subset L that preserves independence.*

Proof. By Lemma 9, for each $x \in X$, there exists $\varepsilon_x > 0$ such that $B_{\varepsilon_x}(x) \perp x$, and that for any $r \in B_{\varepsilon_x}(x)$ and any convex regular closed subset L that preserves independence, $\overline{xr}$ and L preserve bi-independence. For any x, let $B_x = B_{\varepsilon_x}(x)$. Then construct the polytope in the same way as in the proof of Lemma 4 and we are done. ■

The following lemma shows that each polytope constructed in Lemma 10 can be a linear region of some CFPL representation of $\succsim$.

Lemma 11 *Suppose $\succsim$ satisfies weak local bi-independence and $\hat{\mathcal{D}}$ is the collection of all polytopes constructed in Lemma 10. Then for any $\Delta, \Delta' \in \hat{\mathcal{D}}$, Δ and Δ' preserve bi-independence.*

Proof. Since Δ and Δ' both preserve independence and have nonempty interior, Lemma 1 implies that there are affine functions $U : \Delta \to \mathbb{R}$ and $U' : \Delta' \to \mathbb{R}$ that represent $\succsim$ on Δ and Δ', respectively. To prove that Δ and Δ' preserve bi-independence, pick any $x, r \in \Delta$, $y, s \in \Delta'$, and $\lambda \in (0, 1)$ such that $\lambda xr \in \Delta$, $\lambda ys \in \Delta'$, and $x \sim y$. We want to show that $r \succsim s \Leftrightarrow \lambda xr \succsim \lambda ys$.

Since $\Delta = \overline{x^0 \ldots x^N}$ and $\Delta' := \overline{y^0 \ldots y^N}$ for some $x^0, \ldots, x^N \in X$ and $y^0, \ldots, y^N \in X$, without loss of generality, let $U(x^0) = \min_i U(x^i)$, $U(x^N) = \max_i U(x^i)$, $U'(y^0) = \min_i U'(y^i)$, and $U'(y^N) = \max_i U'(y^i)$. Clearly, $U(\lambda xr) \in [U(x^0), U(x^N)]$ and $U'(\lambda ys) \in [U'(y^0), U'(y^N)]$. The cases with $x^0 \sim x^N$ or $y^0 \sim y^N$ are straightforward. Therefore, assume that $x^N \succ x^0$ and $y^N \succ y^1$. Without loss of generality, let $U(x^N) = U'(y^N) = 1$ and $U(x^0) = U'(y^0) = 0$. Standard arguments imply that there exist unique $\alpha, \alpha', \alpha'', \beta, \beta', \beta'' \in [0, 1]$ such that $\alpha x^0 x^N \sim x$, $\alpha' x^0 x^N \sim r$, $\alpha'' x^0 x^N \sim \lambda xr$, $\beta y^0 y^N \sim y$, $\beta' y^0 y^N \sim s$, and $\beta'' y^0 y^N \sim \lambda ys$. Then,

$$U(\lambda pr) = \alpha'' = \lambda U(p) + (1 - \lambda)U(r) = \lambda\alpha + (1 - \lambda)\alpha'.$$

Similarly,

$$U'(\lambda qs) = \beta'' = \lambda U'(q) + (1 - \lambda)U'(s) = \lambda\beta + (1 - \lambda)\beta'.$$

By the construction in Lemma 10, since each $\Delta \in \hat{\mathcal{D}}$ is regular closed and convex and preserves independence, $\overline{x^0 x^N}$ and Δ' preserve bi-independence, which implies that $\overline{x^0 x^N}$ and $\overline{y^0 y^N}$ preserve bi-independence. Therefore, since $\alpha x^0 x^N \sim x \sim y \sim \beta y^0 y^N$, we have $\alpha' x^0 x^N \succsim \beta' y^0 y^N \Rightarrow (\lambda\alpha + (1 - \lambda)\alpha') x^0 x^N \succsim (\lambda\beta + (1 - \lambda)\beta') y^0 y^N$ and $\alpha' x^0 x^N \succ \beta' y^0 y^N \Rightarrow (\lambda\alpha + (1 - \lambda)\alpha') x^0 x^N \succ (\lambda\beta + (1 - \lambda)\beta') y^0 y^N$, which establishes the lemma. ■

Similar to the proof of Theorem 1, we identify the "largest" linear regions using Zorn's Lemma. Let $\hat{X}_o := \bigcup_{\Delta \in \hat{\mathcal{D}}} \operatorname{int}(\Delta)$. It is clear that $\hat{X}_o$ is an open and dense subset of X. For any $x \in \hat{X}_o$, pick $\hat{\Delta}_x \in \hat{\mathcal{D}}$ such that $x \in \operatorname{int}(\hat{\Delta}_x)$. Recall that $\mathcal{O} = \{L \subseteq X : L \text{ is nonempty, connected, and open}\}$. Let $\hat{\mathcal{F}}$ be the set of all functions $P : \hat{X}_o \to \mathcal{O}$ such that for any $x, y \in \hat{X}_o$, (i) $\operatorname{int}(\hat{\Delta}_x) \subseteq P(x)$, (ii) $P(x)$ and $P(y)$ preserve bi-independence, and (iii) $P(x)$ and Δ preserve bi-independence for any $\Delta \in \hat{\mathcal{D}}$. Clearly, $\hat{\mathcal{F}}$ is nonempty since it contains $x \mapsto \hat{\Delta}_x$.

Define a binary relation $\hat{\Subset}$ on $\mathcal{F}$ as follows: For any $x, y \in \mathcal{F}$, $P \mathrel{\hat{\Subset}} Q$ if for any $x \in \hat{X}_o$, $P(x) \subseteq Q(x)$. It is straightforward to verify that $\hat{\Subset}$ is a partial order on $\hat{\mathcal{F}}$. Take any totally ordered subset of $\hat{\mathcal{F}}$, $\{P_i\}_{i\in I}$, in which I is an index set. Let $P^* : X_o \to \mathcal{O}$ be a function such that for any $x \in \hat{X}_o$, $P^*(x) := \bigcup_{i\in I} P_i(x)$. It must be true that $P^* \in \mathcal{P}$. First of all, $P^*(x)$ is open since every $P_i(x)$ is open. Second, $P^*(x)$ is connected, since every $P_i(x)$ is connected and contains $\mathrm{int}(\hat{\Delta}_x)$, which is connected. Now we show that $P^*(x)$ and $P^*(y)$ preserve bi-independence for all $x, y \in \hat{X}_o$. To see this, for any $\lambda \in (0, 1)$, if $x', r', \lambda x'r' \in P^*(x)$ and $y', s', \lambda y's' \in P^*(y)$, by $\{P_i\}_{i\in I}$ is totally ordered by $\hat{\Subset}$, there must exist some index $j \in I$ such that $x', r', \lambda x'r' \in P_j(x)$ and $y', s', \lambda y's' \in P_j(y)$. Then, the property that we want $x', r', \lambda x'r', y', s', \lambda y's'$ to satisfy to ensure that $P^*(x)$ and $P^*(y)$ preserve bi-independence follows from the fact that $P_j(x)$ and $P_j(y)$ preserve bi-independence. Similarly, it is easy to show that $P^*(x)$ and Δ preserve bi-independence for any $\Delta \in \hat{\mathcal{D}}$. Hence, P^* is an upper bound of $\{P_i\}_{i\in I}$ in terms of $\hat{\Subset}$.

We can apply now Zorn's lemma and know that $\hat{\mathcal{F}}$ contains some $\hat{\Subset}$-maximal element. Denote this $\hat{\Subset}$-maximal element by $\hat{P}^*$. Clearly, $\bigcup_{x\in\hat{X}_o} \hat{P}^*(p)$ is an open and dense subset of X. The next step is to prove that $\hat{P}^*$ has some nice properties. To do that, we will need the following two lemmas.

For any two subsets L_1, L_2 of X, we write $L_1 \rightleftarrows L_2$ if there exist $x_h, x_l \in L_1$ and $y_h, y_l \in L_2$ such that $x_h \succ x_l$, $y_h \succ y_l$, $x_h \succ y_l$, and $y_h \succ x_l$. For a finite sequence of subsets $L_1, \ldots, L_m$ of X, we write $L_1 \rightleftarrows \cdots \rightleftarrows L_m$ if $L_i \rightleftarrows L_{i+1}$ for any $i \in \{1, \ldots, m-1\}$. Note that $\rightleftarrows$ is not a transitive binary relation; that is, $L_1 \rightleftarrows L_2 \rightleftarrows L_3$ does not imply $L_1 \rightleftarrows L_3$.

The proof of Lemma 12 is similar to Chapter 2.4 of Schmidt (1998), which we present in the Online Appendix.

Lemma 12 *Suppose $L_1, \ldots, L_m \subseteq X$ are nonempty, connected, and open subsets such that L_i and L_j preserve bi-independence for any $i, j \in \{1, \ldots, m\}$. If $L_1 \rightleftarrows \cdots \rightleftarrows L_m$, there exist affine functions $U_i : L_i \to \mathbb{R}$, $i = 1, \ldots, m$, such that the function $U : \bigcup_{i=1}^m L_i \to \mathbb{R}$ that satisfies $p \in L_i \Rightarrow U(p) = U_i(p)$, $i = 1, \ldots, m$, represents $\succsim$ on $\bigcup_{i=1}^m L_i$.*

To show that different linear regions must have empty intersections, we need one more lemma, whose proof can be found in the Online Appendix.

Lemma 13 *Suppose $L_1, L_2, L_3 \subseteq X$ are nonempty, connected, open subsets such that L_i and L_j preserve bi-independence for any $i, j \in \{1, \ldots, m\}$. If $L_1 \cap L_2 \neq \emptyset$, then*

(i) $L_1 \cup L_2$ preserves independence;

(ii) $L_1 \cup L_2$ and L_3 preserve bi-independence.

Now we proceed to show that the maximal linear regions cannot overlap.

Lemma 14 *Suppose $\succsim$ satisfies weak local bi-independence and $\hat{P}^*$ is a $\hat{\Subset}$-maximal element. Then for any $x, y \in \hat{X}_o$, if $\hat{P}^*(x) \cap \hat{P}^*(y) \neq \emptyset$, then $\hat{P}^*(x) = \hat{P}^*(y)$.*

Proof. Suppose $\hat{P}^*(x) \cap \hat{P}^*(y) \neq \emptyset$. By Lemma 13, we have (i)$\hat{P}^*(x) \cup \hat{P}^*(y)$ preserves independence, (ii) $\hat{P}^*(x) \cup \hat{P}^*(y)$ and $\hat{P}^*(r)$ preserve bi-independence for all $r \in \hat{X}_o$, and (iii) $\hat{P}^*(x) \cup \hat{P}^*(y)$ and Δ preserve bi-independence for any $\Delta \in \hat{\mathcal{D}}$. Thus, if $\hat{P}^*(x) \cap \hat{P}^*(y) \neq \emptyset$, then $\hat{P}^*$ is not $\hat{\Subset}$-maximal unless $\hat{P}^*(x) = \hat{P}^*(y)$. To see this, if $\hat{P}^*(x) \neq \hat{P}^*(y)$, we can define a new function $\hat{P} : \hat{X}_o \to \mathcal{O}$ that agrees with $\hat{P}^*$ except at x and y. Let $\hat{P}(x) = \hat{P}(y) = \hat{P}^*(x) \cup \hat{P}^*(y)$. Then, we have $\hat{P} \neq P^*$, $\hat{P} \in \hat{\mathcal{F}}$, and $\hat{P}^* \hat{\Subset} \hat{P}$. ■

Lemma 15 *Suppose $\succsim$ satisfies weak local bi-independence and $\hat{P}^*$ is a $\hat{\Subset}$-maximal element. Then $\{\hat{P}^*(x) : x \in \hat{X}_o\}$ is finite.*

Proof. Suppose $\{\hat{P}^*(x) : x \in \hat{X}_o\} = \{P_i\}_{i \in I}$ for some infinite index set I. Let B_x be where the second vertex is chosen in the procedure for constructing $\Delta \in \hat{\mathcal{D}}$ in Lemma 10. We show that for any $y \in B_x$, there exists $i \in I$ such that $\overline{xy} \subseteq \mathrm{cl}(P_i)$. Without loss of generality, assume $x \neq y$. Applying the procedure in Lemma 10, we can construct $\Delta \in \hat{\mathcal{D}}$ such that $\overline{xy} \subseteq \Delta$. By the denseness of $\hat{X}_o$, there exists $i \in I$ such that $P_i \cap \mathrm{int}(\Delta) \neq \emptyset$. Then by Lemma 14, it must be the case that $\mathrm{int}(\Delta) \subseteq P_i$, which implies $\overline{xy} \subseteq \Delta \subseteq \mathrm{cl}(P_i)$. The rest is similar to the proof of Lemma 5. ■

Now we start to construct the CFPL representation. The idea is to separate the linear regions into groups according to their utility range and perform positive affine transformations group by group.

For any $P, P' \in \mathcal{P} := \{\hat{P}^*(x) : x \in \hat{X}_o\}$, we write $P' \leftrightsquigarrow P$ if $V(P) = V(P')$ or there is a finite sequence of subsets $P_1, \ldots, P_m \in \mathcal{P}$ such that $P_1 = P$, $P_m = P'$, and $P_1 \rightleftarrows \cdots \rightleftarrows P_m$. By definition, $\leftrightsquigarrow$ is reflexive and transitive, and hence an equivalence relation defined on $\mathcal{P}$. Let $\mathcal{Q}_1, \ldots, \mathcal{Q}_K$ be the equivalence classes induced by $\leftrightsquigarrow$. Since $\mathcal{P}$ is finite, it is clear that there are only a finite number of equivalent classes. Let $Q_i^* = \bigcup_{P \in \mathcal{Q}_i} P$ for each $i \in \{1, \ldots, K\}$. It is easy to see that (i) $V(Q_i^*)$ is a (possibly degenerate) interval for any i, (ii) $V(Q_i^*) \cap V(Q_j^*)$ has empty interior if $i \neq j$, and (iii) $V(X_o) = \bigcup_{i=1}^K V(Q_i^*)$.

For each i, let $V_i^h := \sup V(Q_i^*)$ and $V_i^l := \inf V(Q_i^*)$. If $V(X_o)$ is a singleton, then the whole theorem is trivially true. Without loss of generality, let $V(Q_i^*)$ be a nondegenerate interval if and only if $i \in \{1, \ldots, k\}$, and assume $V_i^h \leqslant V_{i+1}^l$ for each $i \in \{1, \ldots, k-1\}$.

Consider V_1 first. By Lemma 12, there exists a CFPL function $U_1 : Q_1^* \to \mathbb{R}$ that represents $\succsim$ on Q_1^*. We can a perform positive affine transformation to U_1 such that $\inf U^1 = V_1^l$ and $\sup U^1 = V_1^h$. Since V also represents $\succsim$ on Q_1^*, there exists a strictly increasing function $f_1 : V(Q_1^*) \to \mathbb{R}$ such that $f_1(V(p)) = U_1(p)$ for any $p \in Q_1^*$. Extend f_1's domain to $V(X)$ by letting $f_1(v) = v$ for any $v \in V(X) \backslash V(Q_1^*)$.

The proof of the next lemma can be found in the Online Appendix.

Lemma 16 *The function f_1 is strictly increasing and continuous.*

Thus, $f_1 \circ V$ is continuous on X and CFPL on Q_1^*. Recursively, for each $2 \leqslant i \leqslant k$, repeat the exercise above to construct continuous and strictly increasing function $f_i : V(X) \to \mathbb{R}$ such that $f_i \circ f_{i-1} \circ \cdots f_1 \circ V$ represents $\succsim$, and is CFPL on Q_i^*. In the end, we have $U = f_k \circ \cdots \circ f_1 \circ V$ represents $\succsim$, and is CFPL on $\bigcup_{i=1}^k Q_i^*$. Since each $V(Q_i^*)$ is a constant for $i > k$, U is CFPL on X_o. By Lemma 1, it is clear that U is affine on $\mathrm{cl}(P)$ for any $P \in \mathcal{P}$, and $\bigcup_{P \in \mathcal{P}} \mathrm{cl}(P) = X$. Since each $P \in \mathcal{P}$ is open, $\mathrm{cl}(P)$ is regular closed and we are done with the sufficiency of the axioms. ■

Proof of Proposition 3

Proof. Suppose W is a CFPL representation of $\succsim$ and for some strictly increasing CFPL function $f : W(X) \to \mathbb{R}$, $\tilde{W} = f \circ W$. Since f is strictly increasing, $\tilde{W}$ must represent $\succsim$. Since f is CFPL, $\tilde{W}$ is also CFPL. Hence, $\tilde{W}$ is a CFPL representation of $\succsim$.

Next, suppose $W, \tilde{W}$ are CFPL representations of $\succsim$. For the regular closed subsets in X such that their union is X and W is affine on each of them, suppose $X_1, \ldots, X_{m_1}$ are the connected components of those subsets. For the regular closed subsets in X such that their union is X and $\tilde{W}$ is affine on each of them, suppose $Y_1, \ldots, Y_{m_2}$ are the connected components of those subsets. Consider the collection of intersections between any X_i and Y_j, denoted by $\{Z_1, \ldots, Z_m\}$. Clearly, m is finite, both W and $\tilde{W}$ must be affine on each Z_k, and the union of all Z_k's is X. Let $W(Z_k) = [W_k^l, W_k^h]$ and $\tilde{W}(Z_k) = [\tilde{W}_k^l, \tilde{W}_k^h]$. Then, $W_1^l, W_1^h, W_2^l, W_2^h, \ldots, W_m^l, W_m^h$ are elements of $W(X)$. Rearrange these numbers in an ascending order and denote them by $W_1 = \min_X W(x) \leqslant W_2 \leqslant \cdots \leqslant W_{2m} = \max_X W(x)$. Similarly, $\tilde{W}_1^l, \tilde{W}_1^h, \tilde{W}_2^l, \tilde{W}_2^h, \ldots, \tilde{W}_m^l, \tilde{W}_m^h$ are elements of $\tilde{W}(X)$. Rearrange them in an ascending order and denote them by $\tilde{W}_1 = \min_X \tilde{W}(x) \leqslant \tilde{W}_2 \leqslant \cdots \leqslant \tilde{W}_{2m} = \max_X \tilde{W}(x)$. We know that each $W^{-1}([W_i, W_{i+1}]) = \tilde{W}^{-1}([\tilde{W}_i, \tilde{W}_{i+1}])$ must be the union of some subsets of of $Z_1, \ldots, Z_m$. Hence, W and $\tilde{W}$ are affine on each connected component of $W^{-1}([W_i, W_{i+1}]) = \tilde{W}^{-1}([\tilde{W}_i, \tilde{W}_{i+1}])$.

Construct a function $f : W(X) \to \mathbb{R}$ as follows. Let $f(W_i) = \tilde{W}_i$. Between W_i and W_{i+1}, make f an affine function. It is easy to see that $\tilde{W} = f \circ W$. By construction, f is CFPL. Finally, since both W and $\tilde{W}$ represent the same preference, f is strictly increasing. ■

Proof of Proposition 4

Proof. The first and second statements are straightforward. We only prove the only-if part of the third and fourth statements. For the former, take a canonical COLU representation of $\succsim$, $V(x) = \max_{1 \leqslant j \leqslant m} \min_{i \in I_j} \mu_i \cdot x + \alpha_i$. Since each $\mu_i \cdot x + \alpha_i$ is equal to V in some regular closed subset of X, applying monotonicity in that subset immediately implies that $\mu_i \in \mathbb{R}^N_+$ for any i.

For the only-if part of the last statement, take a CFPL representation of $\succsim$, V. Let $X_1, \ldots, X_k$ be the connected components of the regular closed subsets in the definition of the representation. Each X_i is a regular closed connected subset of X such that V is affine on X_i. First, V is not constant on any of these X_i's. To see this, suppose V is constant on X_i. By Lemma 3 and C-independence, $\mathrm{co}(X_i \cup \{\underline{u}\mathbf{1}\})$ must preserve independence, and hence $\mathrm{co}(X_i \cup \{\underline{u}\mathbf{1}, \overline{u}\mathbf{1}\})$ must preserve independence. This implies that V is constant on $\{u\mathbf{1} : u \in [\underline{u}, \overline{u}]\}$, which violates the assumption that the preference is nondegenerate.

Next, without loss of generality, suppose X_1 contains an alternative that maximizes V. By Lemma 3 and C-independence, $Y_1 = \mathrm{co}(X_1 \cup \{\underline{u}\mathbf{1}, \overline{u}\mathbf{1}\})$ preserves independence. On the restricted domain Y_1, V is CFPL. We can find a strictly increasing CFPL function to transform V into W such that W is equal to $\mu_1 \cdot x$ on Y_1, in which $\mu_1 \cdot \mathbf{1}$ is normalized to either 1 or -1. Without loss of generality, assume that $\mu_1 \cdot \mathbf{1} = 1$. Note that by Proposition 3, W also represents $\succsim$ and is CFPL. Let $Y_1, \ldots, Y_l$ be the connected components of the regular closed subsets in the definition of a CFPL function for W. Clearly, Y_1 contains an alternative that maximizes W.

Similar to the binary relations $\rightleftarrows$ and $\leftrightsquigarrow$ defined in the proof of Theorem 2, let $Y_i \rightleftarrows Y_j$ if $W(Y_i) \cap W(Y_j)$ has nonempty interior. Let $Y_i \leftrightsquigarrow Y_j$ if there exist $i_1, \ldots, i_m \in \{1, \ldots, l\}$ such that $i_1 = i$, $i_m = j$, and $Y_{i_1} \rightleftarrows \cdots \rightleftarrows Y_{i_m}$. By definition and because W is not constant on any Y_i, $\leftrightsquigarrow$ is reflexive, symmetric, and transitive. Different equivalent classes induced by $\leftrightsquigarrow$ have at most one point in common in terms of their utility ranges.

Take the equivalent class of Y_1 induced by $\leftrightsquigarrow$. On an arbitrary set in that equivalent class, suppose W's local utility function is $\mu \cdot x + \alpha$. We want to prove that $\mu \cdot \mathbf{1} = 1$ and $\alpha = 0$. If Y_1 is the only element of that equivalent class, we are done. Suppose this is not true. Let Y_i be an element of that equivalent class. There must be some Y_j in that class such that $Y_i \rightleftarrows Y_j$. Since $W(Y_i) \cap W(Y_j)$ has nonempty interior, there must exist $x_i \in \mathrm{int}(Y_i)$ and $x_j \in \mathrm{int}(Y_j)$ such that $W(x_i) = W(x_j)$. Fixing two arbitrary distinct constant u's in $[\underline{u}, \overline{u}]$, by C-independence, for a sufficiently small $\lambda \in (0, 1)$ such that $\lambda x_i(u\mathbf{1}) \in \mathrm{int}(Y_i)$ and $\lambda x_j(u\mathbf{1}) \in \mathrm{int}(Y_j)$, it must be true that $W(\lambda x_i(u\mathbf{1})) = W(\lambda x_j(u\mathbf{1}))$. Suppose W's local utility functions on Y_i and Y_j are $\mu_i \cdot x + \alpha_i$ and $\mu_j \cdot x + \alpha_j$ respectively. Then, we have $\lambda W(x_i) + (1 - \lambda)(\mu_i \cdot (u\mathbf{1}) + \alpha_i) = \lambda W(x_j) + (1 - \lambda)(\mu_j \cdot (u\mathbf{1}) + \alpha_j)$ for two distinct u's, which implies that $\mu_i \cdot \mathbf{1} = \mu_j \cdot \mathbf{1}$ and $\alpha_i = \alpha_j$. Following the same argument, we know that for all Y_i's in this equivalent class with W on each Y_i written as $\mu_i \cdot x + \alpha_i$, μ_i's must have the same total mass and α_i's must be identical. Since Y_1 is in this equivalent class, we know that μ_i's in this equivalent class must have the same total mass 1 and α_i's in this equivalent class must be 0.

If $Y_1, \ldots, Y_l$ are all in the same equivalent class, we are done. Suppose this is not true. Then, at least one of $Y_1, \ldots, Y_l$ must contain an alternative that minimizes W and does not belong to the equivalent class of Y_1. Without loss of generality, suppose it is Y_l. Consider $Y_L = \mathrm{co}(Y_l \cup \{\underline{u}\mathbf{1}, \overline{u}\mathbf{1}\})$. First, observe that by Lemma 3 and C-independence, Y_L preserves independence. Also observe that the utility range of Y_L must overlap with that of the equivalent class of Y_1, because they both contain $\{u\mathbf{1} : u \in [\underline{u}, \overline{u}]\}$ and $W(\{u\mathbf{1} : u \in [\underline{u}, \overline{u}]\})$ is not a singleton.

Next, denote the subset of Y_L whose utility range does not overlap with that of the equivalent class of Y_1 by Y_L^*. We can find a strictly increasing CFPL transformation that is equal to the identity function on $W(X)\backslash W(Y_L^*)$ to transform W into W^* such that at any point in Y_L^*, W^*'s local utility function $\mu^* \cdot x + \alpha^*$ satisfies $\mu^* \cdot \mathbf{1} = 1$. We can normalize $\mu^* \cdot \mathbf{1}$ to 1 rather than -1 because of the following reason. Note that Y_L has a subset whose utility range overlaps with that of the equivalent class of Y_1 but does not overlap with that of Y_L^*. Therefore, at any point of that subset, either W's or W^*'s local utility function, $\mu \cdot x + \alpha$, must satisfy $\mu \cdot \mathbf{1} = 1$ and $\alpha = 0$, as we have established previously. Then, since Y_L preserves independence, it cannot be the case that $\mu^* \cdot \mathbf{1} = -1$.

Finally, we prove that α^* must also be zero. If α^* is not zero, W^* must be discontinuous at the boundary between $\mathrm{cl}(Y_L \backslash Y_L^*)$ and Y_L, because on either of these two sets, the total mass of the finite signed measure of the local utility function of W^* is 1, but the constant term of the local utility function is different, which makes W^* discontinuous and hence we reach a contradiction. Then, note that any point in X, whose local utility function of W^* is $\mu \cdot x + \alpha$, must belong to an element of either the equivalent class of Y_1 or the equivalent class of Y_L^*. Then, we know that $\mu \cdot \mathbf{1} = 1$ and $\alpha = 0$, which follows from the argument that we use to show that any two Y_i, Y_j that overlap with each other in terms of the utility range must satisfy $\mu_i \cdot \mathbf{1} = \mu_j \cdot \mathbf{1}$ and $\alpha_i = \alpha_j$. ■

References

Arndt, S. W. and H. Kierzkowski (2001). *Fragmentation: New Production Patterns in the World Economy*. OUP Oxford.

Arora, R., A. Basu, P. Mianjy, and A. Mukherjee (2018). Understanding Deep Neural Networks with Rectified Linear Units. *International Conference on Learning Representations*.

Caplin, A., D. Martin, and P. Marx (2022). Modeling Machine Learning. Working Paper.

Cerreia-Vioglio, S., D. Dillenberger, and P. Ortoleva (2022). Caution and Reference Effects. Working Paper.

Chandrasekher, M. (2019). Source-Dependent Uncertainty Aversion. Working Paper.

Chandrasekher, M., M. Frick, R. Iijima, and Y. Le Yaouanq (2022). Dual-Self Representations of Ambiguity Preferences. *Econometrica 90*(3), 1029–1061.

Chateauneuf, A. and J. Faro (2009). Ambiguity through Confidence Functions. *Journal of Mathematical Economics 45*(9-10), 535–558.

Cho, I.-K. and J. A. Libgober (2021). Algorithm Games and Rational Play with Strategic Inference. Working Paper.

Debreu, G. (1954). *Decision Processes*, Chapter XI Representation of a Preference Ordering by a Numerical Function, pp. 159–165. John Wiley & Sons, Inc.

Ellis, A. and Y. Masatlioglu (2021). Choice with Endogenous Categorization. *Review of Economic Studies* (forthcoming).

Fudenberg, D. and A. Liang (2019). Predicting and Understanding Initial Play. *American Economic Review 109*(12), 4112–4141.

Ghirardato, P., F. Maccheroni, and M. Marinacci (2004). Differentiating Ambiguity and Ambiguity Attitude. *Journal of Economic Theory 118*(2), 133–173.

Gilboa, I. and D. Schmeidler (1989). Maxmin Expected Utility with Non-Unique Prior. *Journal of Mathematical Economics 18*(2), 141–153.

Hahnloser, R., R. Sarpeshkar, M. Mahowald, R. Douglas, and H. Seung (2000). Digital Selection and Analogue Amplification Coexist in a Cortex-Inspired Silicon Circuit. *Nature 405*, 947–951.

Hara, K., E. A. Ok, and G. Riella (2019). Coalitional Expected Multi-Utility Theory. *Econometrica 87*(3), 933–980.

Harless, D. (1992). Predictions about Indifference Curves inside the Unit Triangle: A Test of Variants of Expected Utility Theory. *Journal of Economic Behavior & Organization 18*(3), 391–414.

Kahneman, D. and A. Tversky (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica 47*(2), 263–292.

Karni, E., F. Maccheroni, and M. Marinacci (2015). Chapter 17 - Ambiguity and Nonexpected Utility. Volume 4 of *Handbook of Game Theory with Economic Applications*, pp. 901–947. Elsevier.

Ke, S., B. Wu, and C. Zhao (2022). Learning from a Black Box. Working Paper.

LeCun, Y., Y. Bengio, and G. Hinton (2015). Deep Learning. *Nature 521*, 436–444.

Machina, M. (1982). "Expected utility" Analysis without the Independence Axiom. *Econometrica 50*(2), 277–323.

Ovchinnikov, S. (2002). Max-Min Representation of Piecewise Linear Functions. *Contributions to Algebra and Geometry 43*(1), 297–302.

Schmidt, U. (1998). *Axiomatic Utility Theory Under Risk: Non-Archimedean Representations and Application to Insurance Economics*. Springer-Verlag Berlin Heidelberg.

Siniscalchi, M. (2006). A Behavioral Characterization of Plausible Priors. *Journal of Economic Theory 128*(1), 1–17.

Tarela, J. and M. Martínez (1999). Region Configurations for Realizability of Lattice Piecewise-Linear Models. *Mathematical and Computer Modelling 30*(11–12), 17–27.

Tversky, A. and D. Kahneman (1991). Loss Aversion in Riskless Choice: A Reference-Dependent Model. *Quarterly Journal of Economics 106*(4), 1039–1061.

Online Appendix

Proof of Lemma 8

Lemma 8 *For any $x, y, r, s \in X$, if $\overline{xr}$ and $\overline{ys}$ each preserve independence, then the following statements are equivalent:*

(i) $\overline{xr}$ and $\overline{ys}$ preserve bi-independence.

(ii) For any $x', r' \in \overline{xr}$ and $y', s' \in \overline{ys}$ such that $x' \sim y'$ and $r' \sim s'$, $\lambda x'r' \sim \lambda y's'$ for any $\lambda \in (0,1)$.

(iii) There exists $\varepsilon > 0$ such that for any $y', s' \in \overline{ys}$ with $||y' - s'|| < \varepsilon$, $\overline{xr}$ and $\overline{y's'}$ preserve bi-independence.

Proof. (i)⇒(iii) is trivial. We will show (ii)⇒(i) and (iii)⇒(ii).

To show (ii)⇒(i), suppose (ii) holds but $\overline{xr}$ and $\overline{ys}$ do not preserve bi-independence. Without loss of generality, assume that $x \sim y$ and $r \succ s$ but $\lambda ys \succsim \lambda xr$ for some $\lambda \in (0,1)$. Now we show that there must exist $r' \in \overline{xr}$ and $s' \in \overline{ys}$ such that $r' \sim s'$, and $\lambda xr' \not\sim \lambda ys'$, which is a contradiction.

Case 1: $x \sim y \succsim r \succ s$. Since $\overline{xr}$ and $\overline{ys}$ each preserve independence, we have $y \succsim \lambda ys \succsim \lambda xr \succsim r$. By $x \sim y \succsim r \succ s$ and continuity, there exist $s' \in \overline{ys}$ such that $r \sim s'$. Then, since $\overline{xr}$ and $\overline{ys}$ each preserve independence, we have $\lambda ys' \succ \lambda ys \succsim \lambda xr$ and we are done.

Case 2: $r \succ s \succsim x \sim y$. Similar to case 1, we have $s \succsim \lambda ys \succsim \lambda xr \succsim x$. By $r \succ s \succsim x \sim y$ and continuity, there exist $r' \in \overline{xr}$ such that $r' \sim s$. Then, since $\overline{xr}$ and $\overline{ys}$ each preserve independence, we have $\lambda ys \succsim \lambda xr \succ \lambda xr'$, and we are done.

The case in which $r \succ x \sim y \succ s$ is impossible, since it implies $\lambda xr \succ x \sim y \succ \lambda ys$. Hence, we have established (ii)⇒(i).

Now we show (iii)⇒(ii). Suppose (iii) holds. Consider $x', r' \in \overline{xr}$ and $y', s' \in \overline{ys}$ such that $x' \sim y'$ and $r' \sim s'$. We want to show that $\lambda x'r' \sim \lambda y's'$ for any $\lambda \in (0,1)$. Since $\overline{x'r'}$ and $\overline{y's'}$ each preserve independence, it is without loss of generality to assume that $x' \sim y' \succ r' \sim s'$.

Pick $m \in \mathbb{N}$ such that $||y' - s'|| < m\varepsilon$, in which ε is given in (iii). For $k \in \{0, 1, \ldots, m\}$, let $t^k = y' + (s' - y')k/m$. It is clear that $||t^k - t^{k+1}|| < \varepsilon$. By (iii), $\overline{t^k t^{k+1}}$ and $\overline{x'r'}$ preserve bi-independence for any $k \in \{0, 1, \ldots, m\}$.

Suppose $\overline{t^0 t^k}$ and $\overline{x'r'}$ preserve bi-independence. Since $\overline{x'r'}$ preserves independence, there exists a monotone transformation f such that $U(\lambda x'r') = f \circ V(\lambda x'r') = \lambda$ for any $\lambda \in [0,1]$. Note that since V represents $\succsim$, $U = f \circ V$ also represents $\succsim$. Let $\alpha, \beta \in (0,1)$ be such that $\alpha x'r' \sim t^k$ and $\beta x'r' \sim t^{k+1}$. Since $\overline{x'(\alpha x'r')}$ and $\overline{t^0 t^k}$ preserve bi-independence, we have $U(\lambda x'(\alpha x'r')) = U(\lambda t^0 t^k)$ for any $\lambda \in [0,1]$. Thus,

$$U(\lambda t^0 t^k) = U((\lambda + (1-\lambda)\alpha)x'r')) = \lambda + (1-\lambda)\alpha.$$

Since $\overline{(\alpha x'r')(\beta x'r')}$ and $\overline{t^k t^{k+1}}$ preserve bi-independence, $U(\lambda(\alpha x'r')(\beta x'r')) = U(\lambda t^k t^{k+1})$ for any $\lambda \in [0,1]$. Thus,

$$U(\lambda t^k t^{k+1}) = \lambda\alpha + (1-\lambda)\beta.$$

Thus, U is continuous on $\overline{t^0 t^{k+1}}$ and linear on $\overline{t^0 t^k}$, $\overline{t^k t^{k+1}}$, and $\overline{x'r'}$. If U restricted to $\overline{t^0 t^{k+1}}$ has a kink at t^k, it is easy to see that $\overline{x'r'}$ and $\overline{t^0 t^{k+1}} \cap B_\varepsilon(t^k)$ cannot preserve bi-independence, which contradicts (iii). Hence, U is linear on both $\overline{t^0 t^{k+1}}$ and $\overline{x'r'}$, which implies that $\overline{t^0 t^{k+1}}$ and $\overline{x'r'}$ preserve bi-independence. Inductively, we establish that $\overline{t^0 t^m} = \overline{y's'}$ and $\overline{x'r'}$ preserve bi-independence, and thus $\lambda y's' \sim \lambda x'r'$ for any $\lambda \in (0,1)$. ■

Proof of Lemma 12

Lemma 12 *Suppose $L_1, \ldots, L_m \subseteq X$ are nonempty, connected, and open subsets such that L_i and L_j preserve bi-independence for any $i, j \in \{1, \ldots, m\}$. If $L_1 \rightleftarrows \cdots \rightleftarrows L_m$, there exist affine functions $U_i : L_i \to \mathbb{R}$, $i = 1, \ldots, m$, such that the function $U : \bigcup_{i=1}^m L_i \to \mathbb{R}$ that satisfies $x \in L_i \Rightarrow U(x) = U_i(x)$, $i = 1, \ldots, m$, represents $\succsim$ on $\bigcup_{i=1}^m L_i$.*

Proof. We say that $U : \bigcup_{i=1}^m L_i \to \mathbb{R}$ *weakly represents* $\succsim$ for $L_1 \rightleftarrows \cdots \rightleftarrows L_m$ if U represents $\succsim$ on each $L_j \cup L_{j+1}$, $j = 1, \ldots, m-1$.

We first consider the case in which $m = 2$. Since L_i and L_i preserve bi-independence, L_i preserves independence. By Lemma 1, we can find an affine function $U_i : L_i \to \mathbb{R}$ that represents $\succsim$ on L_i, $i = 1, 2$, respectively. Since $L_1 \rightleftarrows L_2$, we can find $x_h, x_l \in L_1$ and $y_h, y_l \in L_2$ such that both x_h and y_h are strictly preferred to both x_l and y_l. Since $\succsim$ on L_i can be represented by a continuous affine function and L_i is connected, $i = 1, 2$, there must exist $x \in L_1$ and $y \in L_2$ such that $x \sim y$. Since L_1 is open and there exists $x_h, x_l \in L_1$ such that $x_h \succ x_l$, we can always find x^*, x_* in a small ε-ball centered at x such that $\overline{x^* x_*} \subseteq L_1$ and $x^* \succ x \sim y \succ x_*$. Without loss of generality, let $U_1(x) = 0$. Since L_1 and L_2 are open, by continuity, there exists some small enough $\alpha \in (0,1)$ such that $r \in L_1$ implies $\alpha r x \in L_1$ and $x^* \succ \alpha r x \succ x_*$, and $s \in L_2$ implies $\alpha s y \in L_2$ and $x^* \succ \alpha s y \succ x_*$. Then, standard arguments imply that for each $s \in L_2$ there exists a unique $\lambda_s \in (0,1)$ such that $\alpha s y \sim \lambda_s x^* x_*$. Define for each $s \in L_2$

$$\hat{U}_2(s) = \frac{1}{\alpha} U_1(\lambda_s x^* x_*).$$

Take any $s, s' \in L_2$. Since L_2 preserves independence, we have $s \succsim s' \iff \alpha s y \succsim \alpha s' y \iff \hat{U}_2(s) \geqslant \hat{U}_2(s')$. Hence, $\hat{U}_2$ represents $\succsim$ on L_2. For any $\lambda \in (0,1)$ such that $\lambda s s' \in L_2$, since L_1 and L_2 preserve bi-independence,

$$\alpha(\lambda s s')y = \lambda(\alpha s y)(\alpha s' y) \sim \lambda(\lambda_s x^* x_*)(\lambda_{s'} x^* x_*),$$

which implies that

$$\begin{aligned}
\hat{U}_2(\lambda s s') &= \frac{1}{\alpha}U_1((\lambda\lambda_s + (1-\lambda)\lambda_{s'})x^*x_*) \\
&= \frac{1}{\alpha}U_1(\lambda(\lambda_s x^* x_*)(\lambda_{s'}x^*x_*)) \\
&= \frac{1}{\alpha}[\lambda U_1(\lambda_s x^* x_*) + (1-\lambda)U_1(\lambda_{s'}x^*x_*)] \\
&= \lambda\hat{U}_2(s) + (1-\lambda)\hat{U}_2(s').
\end{aligned}$$

Thus, $\hat{U}_2$ is affine and we can find some positive affine transformation to convert U_2 into $\hat{U}_2$. Without loss of generality, let $U_2 = \hat{U}_2$. Note that since $x \sim y$,

$$U_2(y) = \frac{1}{\alpha}U_1(\lambda_y x^* x_*) = \frac{1}{\alpha}U_1(x) = 0 = U_1(x).$$

Take any $x' \in L_1$ and $y' \in L_2$. We want to verify that $x' \succsim y' \iff U_1(x') \geqslant U_2(q')$. Because L_1 and L_2 preserve bi-independence and $x \sim y$, $x' \succsim y' \iff \alpha x'x \succsim \alpha y'y$. According to the definition of α, we can let $\gamma \in (0,1)$ be the unique number such that $\gamma x^* x_* \sim \alpha x' x$. Since $U_1(x) = U_2(y) = 0$,

$$U_1(x') = \frac{1}{\alpha}U_1(\alpha x' x) = \frac{1}{\alpha}U_1(\gamma x^* x_*),$$

and

$$U_2(y') = \frac{1}{\alpha}U_1(\lambda_{y'}x^*x_*)$$

where $\lambda_{y'}x^*x_* \sim \alpha y' y$. Then,

$$x' \succsim y' \iff \alpha x' x \succsim \alpha y' y \iff \gamma \geqslant \lambda_{y'} \iff U_1(x') \geqslant U_2(y').$$

These observations also imply that if $x \in L_1 \cap L_2$, $U_1(x) = U_2(x)$. Then, we can define a function $U : L_1 \cup L_2 \to \mathbb{R}$ such that $x \in L_i \Rightarrow U(x) = U_i(x)$, $i = 1, 2$. The above arguments show that U represents $\succsim$ on $L_1 \cup L_2$. Clearly, any positive affine transformations of U also represent $\succsim$ on $L_1 \cup L_2$.

Now we proceed to prove the lemma for any $m > 2$. By applying the procedure above and performing positive affine transformations inductively, we can find affine functions $U_1, \ldots, U_m$ such that $U : \bigcup_{i=1}^m L_i \to \mathbb{R}$ weakly represents $\succsim$ for $L_1 \rightleftarrows \cdots \rightleftarrows L_m$. We want to prove that U represents $\succsim$ on $\bigcup_{i=1}^m L_i$. Recall $\succsim$ has a continuous representation V.

Step 1: We prove that if for some $i \in \{2, \ldots, m-1\}$, $L_{i-1} \rightleftarrows L_{i+1}$, U must weakly represent $\succsim$ for $L_1 \rightleftarrows L_2 \rightleftarrows \cdots \rightleftarrows L_{i-1} \rightleftarrows L_{i+1} \rightleftarrows L_{i+2} \rightleftarrows \cdots \rightleftarrows L_m$. To prove this, we only need to verify that U represents $\succsim$ on $L_{i-1} \cup L_{i+1}$. Because L_{i-1}, L_i, and L_{i+1} are connected, $V(L_{i-1}), V(L_i)$ and $V(L_{i+1})$ are all intervals. By $L_{i-1} \rightleftarrows L_i \rightleftarrows L_{i+1} \rightleftarrows L_{i-1}$, $V(L_{i-1}) \cap V(L_i) \cap V(L_{i+1})$ contains some nonempty open interval of $\mathbb{R}$. In other words, we can find some $x \in L_{i-1}$, $y \in L_i$, $r \in L_{i+1}$, $\alpha \in (0,1)$, and $B_\varepsilon(y) \subseteq L_i$ such that $V(x) = V(y) = V(r)$ and $V(\alpha x'x), V(\alpha r'r) \in V(B_\varepsilon(y))$ for any $x' \in L_{i-1}$ and $r' \in L_{i+1}$.

Take any $x' \in L_{i-1}$ and $r' \in L_{i+1}$. Since L_{i-1} and L_{i+1} preserve bi-independence and $x \sim r$,

$$x' \succsim r' \iff \alpha x'x \succsim \alpha r'r \iff V(\alpha x'x) \geqslant V(\alpha r'r).$$

Recall that $V(\alpha x'x), V(\alpha r'r) \in V(B_\varepsilon(y))$, which means that we can find some $y_x, y_r \in B_\varepsilon(y)$ such that $y_x \sim \alpha x'x$ and $y_r \sim \alpha r'r$. Since U represents $\succsim$ on $L_{i-1} \cup L_i$ and $L_i \cup L_{i+1}$, respectively, $U(y_x) = U(\alpha x'x)$ and $U(y_r) = U(\alpha r'r)$. Then,

$$\alpha x'x \succsim \alpha r'r \iff U(y_x) \geqslant U(y_r) \iff U(\alpha x'x) \geqslant U(\alpha r'r) \iff U(x') \geqslant U(r'),$$

where the last equivalence follows from $U(x) = U(y)$ and $U(y) = U(r)$.

Step 2: We prove that if $L_1 \rightleftarrows L_m$, there must exist some $i \in \{2, \dots, m-1\}$ such that $L_{i-1} \rightleftarrows L_{i+1}$. If $m = 3$ there is nothing to prove. Suppose $m \geqslant 4$. Let $v_i^h := \sup_{x \in L_i} V(x)$ and $v_i^l := \inf_{x \in L_i} V(x)$ for any $i \in \{1, \dots, m\}$. By definition, $v_i^h > v_i^l$ for any $i \in \{1, \dots, m\}$, and whenever $L_j \rightleftarrows L_k$ for some $j, k \in \{1, \dots, m\}$, $(v_j^l, v_j^h) \cap (v_k^l, v_k^h) \neq \varnothing$.

Suppose for any $i \in \{2, \dots, m-1\}$, $L_{i-1} \not\rightleftarrows L_{i+1}$; that is, either $v_{i-1}^h \leqslant v_{i+1}^l$ or $v_{i+1}^h \leqslant v_{i-1}^l$. If $v_{i-1}^h \leqslant v_{i+1}^l$ holds for every $i \in \{2, \dots, m-1\}$, we must have $L_1 \not\rightleftarrows L_m$. This is clear if m is odd. Suppose m is even. Since $L_1 \rightleftarrows L_2 \rightleftarrows L_3$, it must be the case that $v_2^h > v_3^l > v_1^h$. Hence, for any even $m > 2$, $v_m^l \geqslant v_2^h > v_1^h$, which implies that $L_1 \not\rightleftarrows L_m$. Similarly, it cannot be the case that $v_{i+1}^h \leqslant v_{i-1}^l$ holds for every $i \in \{2, \dots, m-1\}$.

For $m = 4$, the arguments above implies that if $L_{i-1} \not\rightleftarrows L_{i+1}$ for all i, then either (i) $v_1^h \leqslant v_3^l$ and $v_4^h \leqslant v_2^l$, or (ii) $v_3^h \leqslant v_1^l$ and $v_2^h \leqslant v_4^l$. The two cases are symmetric, so we will focus on the former. Since $L_3 \rightleftarrows L_4$, it is clear that $v_3^l < v_4^h$, which implies that $v_1^h < v_2^l$. This contradicts the fact that $L_1 \rightleftarrows L_2$.

Now let $m \geqslant 5$. Then by $L_{i-1} \not\rightleftarrows L_{i+1}$ for all i, there must be some $j \in \{3, \dots, m-2\}$ such that $\max\{v_{j+2}^h, v_{j-2}^h\} \leqslant v_j^l$ or $v_j^h \leqslant \min\{v_{j-2}^l, v_{j+2}^l\}$. We focus on the former case, since the latter is similar. Because $L_{i-2} \rightleftarrows \cdots \rightleftarrows L_{j+2}$, it must be the case that $v_{j-1}^l < v_{j-2}^h \leqslant v_j^l < v_{j-1}^h$ and $v_{j+1}^l < v_{j+2}^h \leqslant v_j^l < v_{j+1}^h$. Then, $(v_{j-1}^l, v_{j-1}^h) \cap (v_{j+1}^l, v_{j+1}^h) \neq \varnothing$, and thus, $L_{j-1} \rightleftarrows L_{j+1}$.

Step 3: We prove that if there exist affine functions $U_i : L_i \to \mathbb{R}$, $i = 1, \dots, m$, such that the function $U : \bigcup_{i=1}^m L_i \to \mathbb{R}$ that satisfies $x \in L_i \Rightarrow U(x) = U_i(x)$, $i = 1, \dots, m$, weakly represents $\succsim$ for $L_1 \rightleftarrows \cdots \rightleftarrows L_m$, then U represents $\succsim$ on $\bigcup_{i=1}^m L_i$. The claim is trivial if $m = 1, 2$. Next, suppose for some $\bar{m} \geqslant 2$, the claim is true for any $m \leqslant \bar{m}$. Assume that now $m = \bar{m} + 1$. Take any $x, y \in \bigcup_{i=1}^{m} L_i$. If $x, y \in L_i$ for some $i \in \{1, \ldots, m\}$, $x \succsim y \iff U(x) \geqslant U(y)$. Therefore, for the rest of the proof of this lemma, let $x \in L_i$ and $y \in L_j / L_i$ for some distinct $i, j \in \{1, \ldots, m\}$.

First, suppose $\{x, y\} \nsubseteq L_1 \cup L_m$. Then, either $\{x, y\} \subseteq \bigcup_{i=2}^{m} L_i$ or $\{x, y\} \subseteq \bigcup_{i=1}^{m-1} L_i$. Since U weakly represents $\succsim$ for $L_1 \rightleftarrows \cdots \rightleftarrows L_m$, it also weakly represents $\succsim$ for $L_2 \rightleftarrows \cdots \rightleftarrows L_m$ and for $L_1 \rightleftarrows \cdots \rightleftarrows L_{m-1}$, respectively. By the induction hypothesis, we have $x \succsim y \iff U(x) \geqslant U(y)$.

Second, consider the case in which $\{x, y\} \subseteq L_1 \cup L_m$. Without loss of generality, let $x \in L_1$ and $y \in L_m \backslash L_1$. If $L_1 \rightleftarrows L_m$, from Steps 1 and 2, we know that there must exist some $i \in \{2, \ldots, m-1\}$ such that $L_{i-1} \rightleftarrows L_{i+1}$, and hence U weakly represents $\succsim$ for $L_1 \rightleftarrows L_2 \rightleftarrows \cdots \rightleftarrows L_{i-1} \rightleftarrows L_{i+1} \rightleftarrows L_{i+2} \rightleftarrows \cdots \rightleftarrows L_m$. Then, we know that U represents $\succsim$ on $L_1 \cup L_2 \cup \cdots \cup L_{i-1} \cup L_{i+1} \cup L_{i+2} \cup \cdots \cup L_m$, and hence that $x \succsim y \iff U(x) \geqslant U(y)$.

Hence, suppose $L_1 \not\rightleftarrows L_m$. Without loss of generality, let $v_1^l \geqslant v_m^h$. (If it is the other case, we reverse the indices of $L_1, \ldots, L_m$.) It must be the case that $x \succsim y$. Then, we only need to prove that $x \sim y \Rightarrow U(x) = U(y)$ and $x \succ y \Rightarrow U(x) > U(y)$. For any $i \in \{1, \ldots, m-1\}$, since $(v_i^l, v_i^h) \cap (v_{i+1}^l, v_{i+1}^h)$ is nonempty, $(v_i^l, v_i^h) \cup (v_{i+1}^l, v_{i+1}^h)$ is an open interval. Therefore, $\bigcup_{i=1}^{m-1} (v_i^l, v_i^h)$ is an open interval that contains $\frac{v_1^l + v_1^h}{2}$ and $\left(\bigcup_{i=1}^{m-1} (v_i^l, v_i^h)\right) \cap (v_m^l, v_m^h) \neq \emptyset$. Notice that since $\frac{v_1^l + v_1^h}{2} > v_1^l \geqslant v_m^h$, we must have $v_m^h \in \bigcup_{i=2}^{m-1} (v_i^l, v_i^h)$; that is, there exists some $r \in L_i$, $i \in \{2, \ldots, m-1\}$ such that $V(r) = v_m^h$. Note that by the induction hypothesis U represents $\succsim$ on $\bigcup_{i=1}^{m-1} L_i$ and $\bigcup_{i=2}^{m} L_i$, respectively. Since $x \in L_1, y \in L_m$ with $v_1^l \geqslant v_m^h$, if $x \sim y$, the only possibility is that $V(x) = V(y) = v_1^l = v_m^h = V(r)$. Then it follows that $U(x) = U(r) = U(y)$ and we are done. If $x \succ y$, then $V(x) \geqslant V(r) \geqslant V(y)$ and at least one of the inqualities is strict. It follows that $U(x) \geqslant U(r) \geqslant U(y)$ and at least one of the inqualities is strict. Thus, $x \succ y \Rightarrow U(x) > U(y)$. ■

Proof of Lemma 13

Lemma 13 *Suppose $L_1, L_2, L_3 \subseteq X$ are nonempty, connected, open subsets such that L_i and L_j preserve bi-independence for any $i, j \in \{1, \ldots, m\}$. If $L_1 \cap L_2 \neq \emptyset$, then*

(i) $L_1 \cup L_2$ preserves independence;

(ii) $L_1 \cup L_2$ and L_3 preserve bi-independence.

Proof. We first show (i). Let L_0 be an open ball such that $L_0 \subseteq L_1 \cap L_2$. Suppose $x \sim y$ for any $x, y \in L_0$. Then since L_1 and L_2 each preserve independence, by Lemma 1, we have $x \sim y$ for any $x, y \in L_1 \cup L_2$. Clearly, in this case, $L_1 \cup L_2$ preserves independence. Suppose there exists $x, y \in L_0$ such that $x \succ y$. Then by definition, $L_1 \rightleftarrows L_0 \rightleftarrows L_2$. Since

L_1 and L_2 each preserve independence, it is clear that L_i and L_j preserve bi-independence for any $i, j \in \{0, 1, 2\}$. By Lemma 12, there exist affine functions $U_i : L_i \to \mathbb{R}$, $i = 0, 1, 2$, such that the function $U : \bigcup_{i=0}^{2} L_i \to \mathbb{R}$ that satisfies $x \in L_i \Rightarrow U(x) = U_i(x)$, $i = 0, 1, 2$, represents $\succsim$ on $\bigcup_{i=0}^{2} L_i$. Note that U is well-defined only if $U_0(x) = U_1(x) = U_2(x)$ for any $x \in L_0 \subseteq L_1 \cap L_2$. Since L_0 is an open ball and U_i is affine, $i = 0, 1, 2$, it follows that U must be affine on $\bigcup_{i=0}^{2} L_i = L_1 \cup L_2$. Thus, $L_1 \cup L_2$ preserve independence.

To show (ii), we first show that if $L_1 \cup L_2 \not\rightleftarrows L_3$, then $L_1 \cup L_2$ and L_3 preserve bi-independence. Since $L_1 \cup L_2$ and L_3 are both connected, $V(L_1 \cup L_2)$ and $V(L_3)$ are two (potentially degenerate) intervals, which implies that $V(L_1 \cup L_2) \cap V(L_3)$ is an interval. If $V(L_1 \cup L_2) \cap V(L_3) = \emptyset$ it is straightforward to verify that $L_1 \cup L_2$ and L_3 preserve bi-independence. By $L_1 \cup L_2 \not\rightleftarrows L_3$, the only remaining case is when $V(L_1 \cup L_2) \cap V(L_3) = \{v\}$ for some $v \in \mathbb{R}$. Pick $x \in L_1 \cup L_2$ and $y \in L_3$ such that $V(x) = V(y) = v$. Let U_0 be an affine function that represents $\succsim$ on $L_1 \cup L_2$ and U_3 be an affine function that represents $\succsim$ on L_3. In addition, we require $U_0(x) = U_3(x)$. Then, standard arguments imply that U defined on $L_1 \cup L_2 \cup L_3$, which agrees with U_0 on $L_1 \cup L_2$ and agrees with U_3 on L_3, represents $\succsim$ on $L_1 \cup L_2 \cup L_3$. Note that U is affine on both $L_1 \cup L_2$ and L_3. Then, it is straightforward to verify that $L_1 \cup L_2$ and L_3 preserve bi-independence.

Now suppose $L_3 \rightleftarrows L_1 \cup L_2$. Let $v_i^h = \sup_{x \in L_i} V(x)$ and $v_i^l = \inf_{x \in L_i} V(x)$ for $i = 1, 2, 3$. Since both L_3 and $L_1 \cup L_2$ are nonempty, open, and connected, $L_3 \rightleftarrows L_1 \cup L_2$ implies $(v_3^l, v_3^h) \cap (\min\{v_1^l, v_2^l\}, \max\{v_1^h, v_2^h\}) \neq \emptyset$. Since $L_1 \cap L_2 \neq \emptyset$, $(\min\{v_1^l, v_2^l\}, \max\{v_1^h, v_2^h\}) = (v_1^l, v_1^h) \cup (v_2^l, v_2^h)$. It follows that $L_1 \rightleftarrows L_3$ or $L_2 \rightleftarrows L_3$. Furthermore, since $L_1 \cup L_2$ preserves independence and $(\min\{v_1^l, v_2^l\}, \max\{v_1^h, v_2^h\}) \subseteq V(L_1 \cup L_2)$, $L_1 \cap L_2 \neq \emptyset$ implies $L_1 \rightleftarrows L_2$. Hence, we can apply Lemma 12 and find affine functions $U_1 : L_1 \to \mathbb{R}$, $U_2 : L_2 \to \mathbb{R}$, and $U_3 : L_3 \to \mathbb{R}$ such that $U : L_1 \cup L_2 \cup L_3 \to \mathbb{R}$ that agrees with U_i on L_i, $i = 1, 2, 3$, represents $\succsim$ on $L_1 \cup L_2 \cup L_3$. Since $L_1 \cap L_2$ is nonempty and open, $\hat{U} : L_1 \cup L_2 \to \mathbb{R}$ that agrees with U_1 on L_1 and with U_2 on L_2 must be affine on $L_1 \cup L_2$. Thus, U is affine on $L_1 \cup L_2$ and L_3, respectively. Then, it is straightforward to verify that $L_1 \cup L_2$ and L_3 preserve bi-independence. ∎

Proof of Lemma 16

Lemma 16 *The function f_1 is strictly increasing and continuous.*

Proof. First, we show that f_1 is strictly increasing. Take $v \in V(Q_1^*)$ and $u, u' \in V(X) \backslash V(Q_1^*)$ such that $u > v > u'$. Pick $x \in Q_1^*$ and $y, y' \in X \backslash Q_1^*$ such that $V(x) = v, V(y) = u$, and $V(y') = u'$. Since Q_1^* is nondegenerate, it follows that $P \rightleftarrows P'$ for any $P, P' \in \mathcal{Q}_1$. Thus, for any $P \in \mathcal{Q}_1$, $V(P)$ is a nondegenerate interval. Since each $P \in \mathcal{Q}_1$ is open, for any $x \in Q_1^*$, there exist $x', x'' \in Q_1^*$ such that $x' \succ x \succ x''$. Hence

$$u > v > u' \Rightarrow V(y) \geqslant \sup U_1 > U_1(x) > \inf U_1 \geqslant V(y'),$$

which implies that $f_1(u) > f_1(v) > f_1(u')$. Thus, f_1 is strictly increasing on $V(X)$, and thus $f_1(V)$ represents $\succsim$ on X.

Second, we show that f_1 is continuous. Let $\{v_j\} \subseteq (V_1^l, V_1^h)$ be a sequence that converges to v. We want to show that $f_1(v_j)$ converges to $f_1(v)$. For each j, pick $y^j \in Q_1^*$ such that $V(y^j) = v_j$. If $v \in (V_1^l, V_1^h)$, then pick $y \in Q_1^*$ such that $V(y) = v$. It suffices to show that $U_1(y^j)$ converges to $U_1(y)$. This is clear, since there exists $P \in \mathcal{P}$ such that $y \in P$, and U_1 is affine on P, an open set. Now suppose $v = V_1^h$. Pick $y \in X$ such that $V(y) = v$. Without loss of generality, assume that $\{v_j\}$ is increasing. We want to show that $U_1(y_j)$ converges to $v = V_1^h = \sup U_1$. Suppose not. Then, there exists $r \in Q_1^*$ such that $r \succ y^j$ for all j. Then continuity implies that $r \succsim y$ and thus $V(r) \geqslant V(y) = V_1^h$, which is a contradiction of the fact that Q_1^* is the union of some open subsets, each of which has a nondegenerate affine representation. Hence, $\lim_{v \uparrow V_1^h} f_1(v) = V_1^h$. Similarly, if $v = V_1^l$, we have $\lim_{v \downarrow V_1^l} f_1(v) = V_1^l$. The rest is straightforward, since f_1 is simply the identity mapping outside (V_1^l, V_1^h). ■

Necessity of the Axioms in Theorem 2

Proof. Suppose the preference $\succsim$ has a CFPL representation. The fact that weak order and continuity hold is clear. Now we show that $\succsim$ satisfies weak local bi-independence. By Theorem 2.1 in Ovchinnikov (2002), there exists distinct affine functions $U_1, \ldots, U_n$ and index sets $I_1, \ldots, I_m$ such that

$$x \succsim y \iff \max_{1 \leqslant j \leqslant m} \min_{i \in I_j} U_i(x) \geqslant \max_{1 \leqslant j \leqslant m} \min_{i \in I_j} U_i(y).$$

Since $U_1, \ldots, U_n$ are distinct, for each $i \neq j$, $\text{aff}(\{x \in X : U_i(x) = U_j(x)\})$ is either empty or defines an affine hyperplane in $\mathbb{R}^N$. We denote the collection of these affine hyperplanes as $\mathcal{A}$. Thus, $\mathcal{A}$ is an arrangement of hyperplanes in $\mathbb{R}^N$. A *region* of $\mathcal{A}$ in X is a connected component of $X \backslash (\bigcup_{H \in \mathcal{A}} H)$. Let $\mathcal{R}(\mathcal{A})$ be the collection of regions of $\mathcal{A}$ in X. For each $L \in \mathcal{R}(\mathcal{A})$, it is easy to see that L is nonempty, open, and $\text{cl}(L)$ is a polytope.[15] Let $\mathcal{P}(\mathcal{A}) := \{\text{cl}(L) : L \in \mathcal{R}(\mathcal{A})\}$. Since $\mathcal{A}$ is finite, $\mathcal{P}(\mathcal{A})$ must be finite. Clearly $\bigcup_{P \in \mathcal{P}(A)} P = X$, and for any $P \in \mathcal{P}(\mathcal{A})$ there exists k such that $\max_{1 \leqslant j \leqslant m} \min_{i \in I_j} U_i(x) = U_k(x)$ for any $x \in P$.

For any $x \in X$, let $\mathcal{A}(x) := \{H \in \mathcal{A} : x \in H\}$ and consider $\mathcal{A}' = \mathcal{A} \backslash \mathcal{A}(x)$. Clearly, there exists $L_x \in \mathcal{R}(\mathcal{A}')$ such that $x \in L_x$. It is clear that $x \in \bigcap \{P \in \mathcal{P}(\mathcal{A}) : x \in P\}$.

[15] A polytope is the bounded intersection of finitely many closed half-spaces in $\mathbb{R}^n$.

Next, we show that

$$L_x = \operatorname{int}\left(\bigcup\{P \in \mathcal{P}(\mathcal{A}) : x \in P\}\right).$$

The claim is trivially true if $\mathcal{A}(x) = \emptyset$. If $\mathcal{A}(x) \neq \emptyset$, then by construction $\mathcal{A}(x)$ is an arrangement of hyperplanes in $\mathbb{R}^N$. Moreover, $x \in \bigcap_{H \in \mathcal{A}(x)} H$. It follows that x is in any closed half-spaces defined by hyperplanes in $\mathcal{A}(x)$. Thus, $x \in P'$ for any $P' \in \mathcal{P}(\mathcal{A}(x))$. Since $x \in L_x$, we have that $x \in P' \cap L_x$ for any $P' \in \mathcal{P}(\mathcal{A}(x))$. It is clear that

$$\{L' \cap L_x : L' \in \mathcal{R}(\mathcal{A}(x))\} = \{L \in \mathcal{R}(\mathcal{A}) : L \subseteq L_x\}.$$

It follows that $x \in P$ for any $P \in \mathcal{P}(\mathcal{A})$ such that $P \subseteq \operatorname{cl}(L_x)$. Since $x \in L_x$, we have $x \notin P$ if $P \not\subseteq \operatorname{cl}(L_x)$. Hence,

$$\begin{aligned}
\operatorname{cl}(L_x) &= \operatorname{cl}\left(\bigcup\{L' \cap L_x : L' \in \mathcal{R}(\mathcal{A}(x))\}\right) \\
&= \operatorname{cl}\left(\bigcup\{L \in \mathcal{R}(\mathcal{A}) : L \subseteq L_x\}\right) \\
&= \bigcup\{P \in \mathcal{P}(\mathcal{A}) : P \subseteq \operatorname{cl}(L_x)\} \\
&= \bigcup\{P \in \mathcal{P}(\mathcal{A}) : x \in P\}.
\end{aligned}$$

Note that since L_x is the interior of a polytope, it is regular open. Thus, $L_x = \operatorname{int}(\operatorname{cl}(L_x))$ and we are done with this step.

The last step is to show that this L_x construction is exactly what we want for weak local bi-independence. Given $x, y \in X$ with $x \sim y$, by the convexity of each $P \in \mathcal{P}(\mathcal{A})$, it is clear that for any $r \in L_x$ and $s \in L_y$, $\overline{xr} \subseteq P$ and $\overline{ys} \subseteq P'$ for some $P, P' \in \mathcal{P}(\mathcal{A})$. Since U coincides with an affine function within P and P', we are done. ∎