Behavioral Neural Networks

Shaowei Ke* Chen Zhao† Zhaoran Wang‡ Sung-Lin Hsieh§¶

December 7, 2020

Abstract

We provide an axiomatic foundation for a class of neural-network models applied to decision-making under risk, called neural-network expected utility (NEU) models. Motivated by classic experimental findings, we weaken the independence axiom in a novel way. We show how to use simple neurons, referred to as behavioral neurons, in NEU models to capture behavioral effects, such as the certainty effect and reference dependence. Empirically, we show that some simple NEU model with natural interpretation predicts better than existing theories, such as expected utility theory and cumulative prospect theory out of sample, and that behavioral neurons help improve NEU models' performance.

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

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

‡Department of Industrial Engineering and Management Sciences, Northwestern University. Email: zhaoran.wang@northwestern.edu.

§Department of Economics, University of Michigan. Email: slhsieh@umich.edu.

¶We are grateful to Tilman Börgers, Yan Chen, In-Koo Cho, David Dillenberger, Kfir Eliaz, Andrew Ellis, Wayne Gao, Itzhak Gilboa, Faruk Gul, Peter Klibanoff, Annie Liang, Jay Lu, Yusufcan Masatlioglu, Paulo Natenzon, Pietro Ortoleva, Wolfgang Pesendorfer, Matthew Shapiro, Marciano Siniscalchi, and seminar participants at Microsoft Research, HKUST, PSE, Econometric Society Summer Meeting, Maryland, UPenn, CETC, D-TEA, Northwestern, Emory, Yale, LSE, and Caltech for helpful discussions. Ke thanks the University of Pennsylvania and the Cowles Foundation at Yale University for hospitality and support when some of the work on this paper was undertaken, and gratefully acknowledges the financial support of the Michigan Institute for Teaching and Research in Economics (MITRE).

1 Introduction

Over the last decade, machine-learning models have demonstrated strong predictive power in many decision problems. For example, when we shop at Amazon, Amazon recommends products to us with the help of machine learning. For a machine-learning model to perform well in this regard, the model must make good predictions about the likelihood that a consumer will buy a product after it is recommended. Machine learning can make good predictions even when the decision problem is complex. In 2015, a machine-learning computer program, AlphaGo, became the first computer program to defeat a human professional Go player, and two years later it defeated the number one ranked player in the world.

The fact that a machine-learning model predicts well in some decision problems, however, does not necessarily make it a good model of how people make decisions. For example, suppose there is a true model that describes how a decision maker behaves. From the formula of the true model, one may gain insights about the decision process and choice behavior. A machine-learning model may approximate the true model in some decision problems well—and therefore predict well in those problems—but those insights from the true model may be lost in the approximation.

Nonetheless, is it possible that of the numerous machine-learning models with the potential to predict well, some are indeed good models of how people make decisions? First, if we can find such a machine-learning model, economists may understand the model better and feel more comfortable applying it in economics, such as in the demand component of an industrial-organization structural model.

Second, given the recent success of machine learning when large datasets are available, it is quite likely that as we accumulate more economic data about people's behavior, such a machine-learning model (together with well-developed methods to train it) may significantly outperform traditional economic models in prediction and help us identify behavioral phenomena that would be difficult to identify using traditional methods.[1]

[1]For instance, traditionally, economists often come up with a conjecture about a behavioral pattern, then

To search for such a machine-learning model, we first need to decide what we mean by a good model of how people make decisions. Let us take the expected utility model as an example. Economists think of it as a good model of how people make decisions for at least two reasons: (i) it is characterized by simple and reasonable axioms imposed on people's choice behavior and (ii) the model provides plausible interpretation of how people make choices.

Therefore, in this paper, we provide a simple axiomatic characterization of a neural-network model—one of the most popular machine-learning models—applied to decision-making under risk. We call this the *neural-network expected utility* (NEU; pronounced "new") model. The axioms are motivated by empirical evidence against expected utility theory, and the NEU model provides plausible interpretation of people's choice behavior. Empirically, we show that some simple NEU models that are easy to interpret perform better than expected utility theory and cumulative prospect theory out of sample. Moreover, we find that what economists have learned about decision-making helps improve the NEU model's performance significantly, at least when the dataset is not large enough.

Consider a classic environment with uncertainty. A decision maker has a preference over risky prospects, which will be called lotteries henceforth. A NEU representation of the decision maker's preference takes a lottery p, which is a vector of probabilities, as the input and outputs the utility of p through a feedforward neural network. Figure 1 offers an example. This NEU function has two *hidden layers*. Each hidden layer has two *neurons* (indicated by boxes). The j-th neuron in the i-th layer does two things. First, it aggregates the values of its child neurons (or the input) using some affine function $\tau_i^{(j)}$. Second, it compares the aggregated value with a threshold and delivers the maximum of the two to the next layer. It is without loss of generality to use 0 as the uniform threshold, since we

conduct experiments to examine it. This approach is important and useful, but its effectiveness hinges on researchers' (prior) domain knowledge about the underlying data-generating process. By contrast, when large datasets are available, machine-learning models such as the neural-network model are good at discovering crucial features or patterns of data effectively without too many inputs from experts' domain knowledge about the data-generating process. Thus, it may be useful in helping us identify behavioral patterns that we have little prior knowledge about.

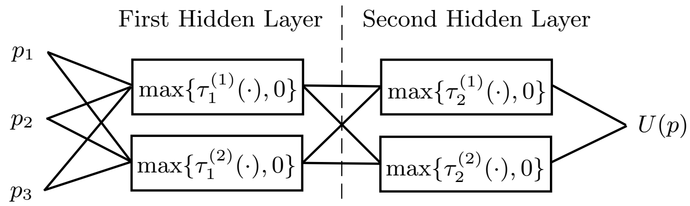

FIGURE 1: In this example, there are three prizes. For any lottery p, p_1, p_2, and p_3 indicate the probabilities of the three prizes. This NEU function has two hidden layers, and each layer has two neurons. Each affine $\tau_1^{(j)}$ is from the set of lotteries (a subset of $\mathbb{R}^3$) to $\mathbb{R}$, 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.

can always add an arbitrary constant to $\tau_i^{(j)}$. The maximum function here captures the activation of a neuron: A neuron is activated if the aggregated value of its child neurons' values is sufficiently high. Finally, the utility of p is an affine aggregation of the values of neurons in the last layer.

In general, we allow for any finite number of neurons and hidden layers. Moreover, because a function defined on the set of lotteries is affine if and only if it is an expected utility function, the NEU function reduces to an expected utility function when the number of hidden layers is zero.

The NEU function has a simple interpretation. In expected utility theory, the decision maker has a unique risk attitude captured by the expected utility function. A decision maker whose preference can be represented by a NEU function may consider multiple risk attitudes plausible, captured by the expected utility functions of first-hidden-layer neurons. For instance, she may have one neuron that activates when the expected value of prizes is high, and may have another neuron that activates whenever the downside risk is high. She also may not be decisive about how to aggregate those risk attitudes; that is, she may have multiple ways in mind to aggregate the risk attitudes (captured by the affine functions of second-hidden-layer neurons). Then, she continues to be indecisive about how to aggregate the aggregations from the previous step until she finally applies the last affine aggregation to obtain the evaluation of the lottery.

We show how to construct simple neural-network structures in a NEU function to capture well-known behavioral phenomena such as the certainty effect, reference dependence, etc. We call these structures *behavioral neurons*. They will be useful in our empirical analysis.

1.1 Behavioral Foundation

Of expected utility theory's axioms, a normatively appealing and yet descriptively controversial one is *Independence*. Our axiomatic characterization of the NEU function naturally stems from an observation common in several well-known empirical findings against *Independence*. Take the Allais paradox as an example. The key component of this paradox is a pair of lotteries that consists of a risk-free one and a risky one. Decision makers are so inclined to choose the risk-free lottery that *Independence* is violated.

Our observation from the Allais paradox has not been studied much: The lotteries in the Allais paradox must be sufficiently far apart. To see this, if the risky lottery is almost risk-free, the fact that the other lottery is risk-free will no longer be that attractive; that is, the certainty effect will be cancelled out. Similar observations also apply to other evidence against *Independence* (see Section 2.2).

Therefore, for behavioral effects to influence the decision maker's evaluation of lotteries so that *Independence* is violated, lotteries in the decision problems often need to be far apart. A natural way to relax *Independence* is then to allow it to fail when lotteries are distant, but require that it hold locally. However, if *Independence* holds locally everywhere, it will hold globally. Thus, we can only hope that some weaker version of *Independence* holds locally.

We introduce a novel and simple way to weaken *Independence*. Essentially, we require that for any lotteries p and q, the independence property holds *only with respect to* p and q locally (see Section 2.2 for details). In our representation theorem, by replacing *Independence* with *Weak Local Bi-Independence*, together with *Weak Order* and *Continuity*, we show that the decision maker's preference has a NEU representation.

1.2 Empirical Analysis

We are interested in understanding how the NEU function performs empirically and how complex the neural network must be to explain and predict people's choice behavior well. Arguably, if a rather complex neural network is required, the interpretation that the NEU function offers might be too complex to be interesting or insightful.

We analyze the NEU model using the training and testing datasets provided by the Choice Prediction Competition 2018 (see Plonsky, Apel, Ert, Tennenholtz, Bourgin, Peterson, Reichman, Griffiths, Russell, Carter, Cavanagh, and Erev (2019)). After aggregating the individual choice data, each data point consists of a description of two lotteries and the fractions of experiment participants who choose the first lottery and the second. We use the training dataset to estimate a model, and then compute its mean square error in the testing dataset (testing error), which measures the model's performance.

We begin by taking expected utility theory and cumulative prospect theory (see Tversky and Kahneman (1992)) as the benchmark. Our first observation is that we must parametrize these models to avoid *overfitting*, because the dataset is not sufficiently large. If we use a (general) expected utility function to fit the training dataset, it will overfit and have a large testing error. By contrast, if we use the constant-absolute-risk-aversion (CARA) expected utility function, this problem can be largely avoided. To our surprise, we also observe that the CARA expected utility model demonstrates high predictive power: It outperforms cumulative prospect theory under standard parametrization.

We also need to parametrize the NEU function to prevent overfitting. Since the affine functions of the first hidden layer of a NEU function are (general) expected utility functions, a natural idea is to require that those functions be CARA expected utility functions. This parametrization turns out to remove too much flexibility from the NEU function, as its performance is essentially identical to the CARA expected utility benchmark.

Recall that we use behavioral neurons to capture well-documented behavioral effects. These behavioral neurons may provide useful flexibility for the NEU function, but are ruled out under the CARA parametrization. Therefore, we require that the first hidden layer of the NEU function consist of the following three types of neurons: (i) neurons that evaluate CARA expected utility, (ii) neurons that capture the certainty effect, and (iii) neurons that capture reference dependence. The second and third types are the behavioral neurons introduced in our theoretical analysis. Then, additional standard hidden layers may be concatenated with this first hidden layer.

We consider different configurations of the network structure and find that the two-hidden-layer NEU function with all three types of neurons used in the first layer has the lowest testing error, which is significantly lower than the benchmark. Hence, a reasonably complex NEU function that has natural interpretation seems to have better predictive power than NEU functions that are too simple (with only one hidden layer) or too complex. This shows that economists' domain knowledge in decision-making is useful for making predictions, especially when the datasets are not sufficiently large.[2]

1.3 Related Literature

The class of NEU functions is identical to the class of continuous piecewise-linear functions. The latter has played an important role in decision theory. Ellis and Masatlioglu (2020) propose a regional preference model as the micro-foundation for salient thinking à la Bordalo, Gennaioli, and Shleifer (2012). 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/regions that preserve bi-independence pairwise from the preference with the help of *Weak Local Bi-Independence*.

In the Anscombe–Aumann choice domain, Siniscalchi (2006) characterizes continuous

[2]This approach is similar to feature engineering in machine learning, except that we require that the engineered features still be consistent with our axioms. The idea of constructing variables/features based on behavioral effects as the input of machine-learning models has appeared in Erev, Ert, Plonsky, Cohen, and Cohen (2017) and Plonsky, Erev, Hazan, and Tennenholtz (2017). Our behavioral neurons are different from their features, since many of their features are incompatible with our axioms.

piecewise-linear functions that satisfy *Certainty-Independence* from Gilboa and Schmeidler (1989) to identify the set of plausible priors. Chandrasekher, Frick, Iijima, and Le Yaouanq (2020) characterize the dual-self expected utility representation by dropping *Uncertainty Aversion* from Gilboa and Schmeidler. Their representation, which is equivalent to two other representations in Ghirardato, Maccheroni, and Marinacci (2004) and Amarante (2009), is a continuous piecewise-linear function that satisfies *Certainty-Independence* from Gilboa and Schmeidler if the number of priors is finite. This also means that its special case, such as Gilboa and Schmeidler, is also continuous piecewise-linear when the number of priors is finite. Assuming that the number of priors is finite, these characterizations of (special cases of) continuous piecewise-linear functions differ from ours. First, roughly speaking, our representation is a dual counterpart of theirs. Second, some form of *Certainty-Independence* is often required in their proofs. Our characterization relies on different techniques and does not require a dual version of *Certainty-Independence*.

Our paper belongs to the literature on non-expected utility. There are three popular directions to relax *Independence* (and hence, the linearity of the expected utility representation in probabilities). Maccheroni (2002); Cerreia-Vioglio (2009); Dillenberger (2010); Chatterjee and Krishna (2011); and Cerreia-Vioglio, Dillenberger, and Ortoleva (2015) consider convex preferences such that any convex combination of two lotteries is weakly preferred to the worse of the two. In the probability simplex, these preferences exhibit convex indifference curves. Dekel (1986); Chew (1983); and Gul (1991) consider preferences that satisfy betweenness; that is, any convex combination of two lotteries is ranked between them. Thus, the indifference curves for these preferences are linear. Quiggin (1982); Yaari (1987); and Tversky and Kahneman (1992) consider rank-dependent preferences such that probabilities of prizes are distorted depending on the ranks of prizes. In our theory, by contrast, preferences exhibit piecewise-linear indifference curves and no probability distortion is introduced.

In spirit, the notion of local independence is first introduced by Machina (1982), who assumes that the preference relation has a smooth representation that can be approximated by linear functions locally everywhere. Under the NEU representation, however, the threshold function $\max\{\cdot, 0\}$ (in Figure 1) creates kinks in the representation, and hence the smoothness assumption imposed by Machina is violated. Moreover, our weakened form of *Independence* holds in sufficiently small neighborhoods in the choice domain, while in Machina, *Independence* only holds approximately in a neighborhood, no matter how small the neighborhood is. In that sense, our notion of locality is different from Machina's.

Many papers have examined non-expected utility theory empirically, including Hong and Waller (1986); Battalio, Kagel, and Jiranyakul (1990); Harless and Camerer (1994); Starmer (2000); Wu, Zhang, and Abdellaoui (2005); Choi, Fisman, Gale, and Kariv (2007); and Bernheim and Sprenger (2020). Among them, Harless and Camerer (1994) propose a new measure of a utility model's empirical performance that takes the model's complexity into account. Similar to us, they find that in some cases, expected utility theory has the best performance. Different from our analysis, they do not examine models' predictive power using a testing dataset that is unused when estimating models. Similar to Bernheim and Sprenger (2020), when we estimate the probability weighting function of cumulative prospect theory, we find little distortion of probabilities.

A growing literature combines economic theory with computer science algorithms. Fudenberg and Liang (2019) use the decision tree algorithm to study the initial play of games. By studying games that the algorithm predicts well, but existing economic models do not, they identify a new parameter that, if introduced to the best existing economic model, improves the model's performance. Similar to their finding that hybrid models may achieve higher predictive power, we find that incorporating multiple types of behavioral neurons into the NEU model is useful. Plonsky, Apel, Erev, Ert, and Tennenholtz (2017) and Erev et al. (2017) are the first to show that introducing features constructed based on findings from behavioral economics and psychology to machine-learning models may significantly improve their performance. Unlike in our paper, they do not provide axiomatic foundations for the models they use. In addition, many of their features are incompatible with our axioms.

Therefore, our behavioral neurons are rather different from their features.

Last, our paper is also related to research that focuses on interpretable machine learning. See Murdoch, Singh, Kumbier, Abbasi-Asl, and Yu (2019) for a recent survey.

The rest of the paper is organized as follows. Section 2 presents the axiomatic characterization of the NEU representation and introduces behavioral neurons. Section 3 analyzes the NEU model in data, and Section 4 concludes.

2 Behavioral Foundation of the NEU Model

Consider a classic choice domain with (objective) uncertainty. Let $Z = \{z_1, \ldots, z_n\}$ be a nonempty finite set of prizes. The set of choice alternatives is

$$\mathcal{L} = \left\{ p \in \mathbb{R}^n_+ : \sum_{i=1}^{n} p_i = 1 \right\},$$

whose elements are lotteries (probability measures) on Z.[3] For any lottery p, p_i indicates the probability of prize z_i. Generic elements of Z are denoted by x, y, z, and generic elements of $\mathcal{L}$ are denoted by p, q, r, s, t. We use $\delta_x \in \mathcal{L}$ to denote the degenerate lottery that yields prize x with probability 1. For any $\lambda \in [0, 1]$, $\lambda p + (1 - \lambda)q$, denoted by λpq, is a lottery such that $(\lambda pq)_i = \lambda p_i + (1-\lambda)q_i$, $i \in \{1, \ldots, n\}$. For any finite set of lotteries $\{p^1, \ldots, p^m\}$, let $\overline{p^1 \ldots p^m} := \mathrm{co}(\{p^1, \ldots, p^m\})$ be the convex hull of $\{p^1, \ldots, p^m\}$. The decision maker has a preference $\succsim$ on $\mathcal{L}$. Its asymmetric and symmetric parts are denoted by $\succ$ and $\sim$, respectively.

2.1 The NEU Representation

Before introducing the NEU representation of $\succsim$ that our axioms on $\succsim$ will characterize, we define two types of functions. First, for two arbitrary natural numbers w and $\tilde{w}$, a function

[3] When Z is a metric space, $\mathcal{L}$ is endowed with the topology of weak convergence. Otherwise, the metric of $\mathcal{L}$ is the Euclidean metric.

$\tau : A \subseteq \mathbb{R}^w \to \mathbb{R}^{\tilde{w}}$ is *affine* if there exist a $\tilde{w}$-by-w matrix β and a vector $\gamma \in \mathbb{R}^{\tilde{w}}$ such that $\tau(a) = \beta a + \gamma$ for any $a \in A$. As usual, the vector-valued function τ can be written as $(\tau^{(1)}, \ldots, \tau^{(\tilde{w})})$, in which $\tau^{(j)}$ is the jth component of τ. Each $\tau^{(j)}$ is affine if and only if τ is affine. Second, a function $U : \mathcal{L} \to \mathbb{R}$ is an *expected utility function* if there exists a function $u : Z \to \mathbb{R}$ such that for any lottery p,

$$U(p) = \sum_{i=1}^{n} p_i u(z_i).$$

The function u is called a *Bernoulli index*.

A useful observation is that every real-valued affine function on $\mathcal{L}$ is an expected utility function with a Bernoulli index $u(z_i) = \tau(\delta_{z_i})$. The converse also holds: Every expected utility function is affine. The NEU representation of $\succsim$ is defined as follows.

Definition 1 *A function $U : \mathcal{L} \to \mathbb{R}$ is a NEU function if there exist*

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

(ii) functions $\theta_i : \mathbb{R}^{w_i} \to \mathbb{R}^{w_i}$, $i = 1, \ldots, h$, such that for any $b \in \mathbb{R}^{w_i}$, $\theta_i(b) = (\max\{b_1, 0\}, \ldots, \max\{b_{w_i}, 0\})$, and

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

$$U(p) = \tau_{h+1} \circ \theta_h \circ \tau_h \circ \cdots \circ \theta_2 \circ \tau_2 \circ \theta_1 \circ \tau_1(p). \tag{1}$$

We say that $\succsim$ has a NEU representation if there exists a NEU function $U : \mathcal{L} \to \mathbb{R}$ such that $p \succsim q \iff U(p) \geqslant U(q)$.

Each function $\theta_i \circ \tau_i$ is called the i-th *hidden layer* whose width is w_i, and $(\theta_i \circ \tau_i)^{(j)} = \max\{\tau_i^{(j)}(\cdot), 0\}$ is called a *neuron*. Thus, the i-th hidden layer has w_i neurons, and equation (1) characterizes a network of neurons with h hidden layers. Figure 1 provides an example of a NEU function.

Mathematically, to evaluate a lottery p, each neuron in the NEU function first aggregates its child neurons' values (see Figure 1) in an affine fashion. Next, 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. Note that it is without loss of generality to normalize all thresholds to 0, since we can always add an arbitrary constant to each affine function $\tau_i^{(j)}$. The neurons in the first hidden layer aggregate the input of the NEU function, p_i's, directly, and the values of neurons in the last (h-th) hidden layer are aggregated into the utility of p.

To interpret the NEU function, recall that in expected utility theory the decision maker has a unique risk attitude characterized by an expected utility function. By contrast, a decision maker whose preference has a NEU representation may consider multiple risk attitudes plausible, characterized by affine functions—which are expected utility functions—of first-hidden-layer neurons. 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. Next, she may not be sure about how to aggregate those risk attitudes. She may consider multiple ways to aggregate the risk attitudes (captured by the affine functions of second-hidden-layer neurons) plausible. The risk attitudes that enter into the aggregations are those significant enough to trigger activation. This process continues until she aggregates the values of last-hidden-layer neurons to obtain the evaluation of the lottery.

Before introducing the axioms, let us point out that θ_i, called the *activation function*, may take other functional forms in general. However, the form we focus on in Definition 1, also known as the rectified linear unit, is considered to be the most popular activation function and to have strong biological motivations.[4] Thus, our representation theorem in Section 2.3 will reinforce the support for rectified linear units from a new perspective: NEU functions with rectified linear units have a simple and reasonable behavioral foundation.

[4]See Hahnloser, Sarpeshkar, Mahowald, Douglas, and Seung (2000); Hahnloser, Seung, and Slotine (2003); and LeCun, Bengio, and Hinton (2015), among others.

2.2 Axioms and the Representation Theorem

Expected utility theory shows that $\succsim$ on $\mathcal{L}$ satisfies the following axioms if and only if it has an expected utility representation—that is, there exists an expected utility function such that $p \succsim q \iff U(p) \geqslant U(q)$.

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

Axiom 2 *(Continuity) For any $p \in \mathcal{L}$, $\{q \in \mathcal{L} : q \succsim p\}$ and $\{q \in \mathcal{L} : p \succsim q\}$ are closed in $\mathcal{L}$.*

Axiom 3 *(Independence) For any $p, q, r \in \mathcal{L}$ and $\lambda \in (0,1)$, $p \succsim q \Rightarrow \lambda pr \succsim \lambda qr$ and $p \succ q \Rightarrow \lambda pr \succ \lambda qr$.*

The main idea of *Independence* is simple—if p is better than q, mixing p and r with any probability should also be better than mixing q with r with the same probability. This idea can be expressed in an equivalent way that will be useful in our paper.

Axiom 4 *(Bi-Independence) For any $p, q, r, s \in \mathcal{L}$ and $\lambda \in (0,1)$, if $p \succsim q$, then $r \succsim s \Rightarrow \lambda pr \succsim \lambda qs$ and $r \succ s \Rightarrow \lambda pr \succ \lambda qs$.*

If we require $r = s$, *Bi-Independence* implies *Independence*. Conversely, by applying *Independence* twice, we can obtain *Bi-Independence*.

Of these axioms, (*Bi-*)*Independence* is the most controversial. A well-known violation comes from the Allais paradox. 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	87%: $1M 3%: $0 10%: $1.5M	87%: $0 13%: $1M	90%: $0 10%: $1.5M

However, let $p = \delta_{1\mathrm{M}}$, $q = \frac{3}{13}\delta_0 + \frac{10}{13}\delta_{1.5\mathrm{M}}$, $r = \delta_{1\mathrm{M}}$, and $s = \delta_0$. The first pair of lotteries becomes $0.13pr$ and $0.13qr$, and the second pair becomes $0.13ps$ and $0.13qs$. (*Bi-*)*Independence* requires that $0.13pr \succsim 0.13qr$ if and only if $0.13ps \succsim 0.13qs$. Therefore, the Allais paradox violates (*Bi-*)*Independence*.

It is well known from the Allais paradox that people are biased toward certainty ($\delta_{1\text{M}}$ in the first pair). We want to point out a new observation that is crucial to our theory: The lotteries in the Allais paradox must look *sufficiently different*. For example, if the right-hand lottery $0.13qr$ in the first pair becomes almost degenerate, decision makers will not be much biased toward $\delta_{1\text{M}}$, and hence the certainty effect will likely be cancelled out in the first pair.[5] To see this, now suppose we have $0.013pr$ and $0.013q'r$ in the first pair and $0.013ps'$ and $0.013q's'$ in the second, in which $q' = \frac{3}{13}\delta_{0.5M} + \frac{10}{13}\delta_{1.5M}$ and $s' = \delta_{0.5\text{M}}$:

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

We make the distributions of each pair of lotteries closer to each other by switching q to q' and adopting a much smaller weight, 0.013. Now it seems much less likely that the degenerate lottery can cause significant violations of (*Bi-*)*Independence*.

Similar observations can be found in other well-known violations of (*Bi-*)*Independence*, such as the common ration effect (see Machina (1987) for a summary). Confronting the following lotteries, most decision makers choose the left-hand lottery from the first pair but

[5]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 right-hand lottery from the second:

First pair		Second pair	
60%: \$2,000	45%: \$2,500	12%: \$2,000	9%: \$2,500
40%: \$0	55%: \$0	88%: \$0	91%: \$0

Intuitively, in the first pair, the difference between 60% and 45% is more salient, but in the second, the difference between \$2,000 and \$2,500 is more salient. Because the lotteries in the second pair are generated by mixing the first pair with δ_0 using the same probability $\frac{1}{5}$, this example violates (*Bi-*)*Independence*.[6]

In this example, first, the lotteries in the first pair must be sufficiently far apart, and second, the two lotteries in the first pair must be sufficiently distant from the two lotteries in the second pair. Suppose now that the two lotteries in the first pair are instead mixed with δ_0 with probability $\frac{29}{30}$ to generate the second pair. The mixed lotteries will become $58\% \times \delta_{2000} + 42\% \times \delta_0$ and $43.5\% \times \delta_{2500} + 56.5\% \times \delta_0$, which is much less likely to generate the significant choice reversal observed in the original example.

Therefore, it seems that to trigger psychological effects in an asymmetric way, as above, to violate (*Bi-*)*Independence*, at least some lotteries in the decision problems must be far apart. To put this differently, if we hope to stick to (*Bi-*)*Independence* as much as we can due to its normative appeal, the above observation suggests that although (*Bi-*)*Independence* may fail when lotteries are distant, perhaps we can assume that it still holds locally.

An immediate difficulty arises. If we assume that (*Bi-*)*Independence* always holds locally, (*Bi-*)*Independence* will hold globally. Therefore, we can only hope that some weaker version of (*Bi-*)*Independence* holds locally everywhere. Below, we introduce a novel way to weaken (*Bi-*)*Independence*.

[6] Although the degenerate lottery δ_0 is used in this example, it is important. We may mix the first pair of lotteries with a non-degenerate lottery and create a similar violation of (*Bi-*)*Independence*.

Definition 2 *A subset L of $\mathcal{L}$ preserves independence with respect to $p \in \mathcal{L}$, denoted by $L \perp p$, if for any $q, r \in L$ and $\lambda \in (0,1)$, $q \succsim r \Rightarrow \lambda pq \succsim \lambda pr$ and $q \succ r \Rightarrow \lambda pq \succ \lambda pr$.*

This definition allows $L \subseteq \mathcal{L}$ to preserve independence with respect to some $p \notin L$, and L need not be convex. A subset of $\mathcal{L}$ is said to be a neighborhood of a lottery p if it is an open convex set that contains p. The axiom below is not our main axiom, but will be implied by and useful in understanding our main axiom.

Axiom 5 *(Weak Local Independence) Every $p \in \mathcal{L}$ has a neighborhood L_p such that $L_p \perp p$.*

Weak Local Independence does not imply that *Independence* holds on L_p. We say that a subset L of $\mathcal{L}$ *preserves independence* if for any $p, q, r \in L$ and $\lambda \in (0,1)$ such that $\lambda pr, \lambda qr \in L$, $q \succsim r \Rightarrow \lambda pq \succsim \lambda pr$ and $q \succ r \Rightarrow \lambda pq \succ \lambda pr$.[7] Clearly, when L is convex, L preserves independence if and only if $L \perp p$ for any $p \in L$.

Consider a simple example with $Z = \{x, y\}$, in which case every lottery p can be identified with a number in $[0,1]$ indicating the probability of x. Suppose $\succsim$ can be represented by

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

This decision maker's utility is increasing in the probability of x in most parts of the domain, but is biased toward certainty: The utility of δ_y is higher than the lottery $p = 0.01$. It can be verified that no neighborhood of $p = 0.01$ preserves 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 lotteries are far apart. A local and weakened version of *Bi-Independence*, by contrast, can avoid this issue.

[7] The idea of the axiom *L-Independence* in Dillenberger, Krishna, and Sadowski (2018) is related to our definition, whereby L preserves independence.

Axiom 6 *(Weak Local Bi-Independence) For any* $p, q \in \mathcal{L}$*, if* $p \succsim q$*, then* p *and* q *have neighborhoods* L_p *and* L_q*, respectively, such that for any* $r \in L_p$ *and* $s \in L_q$*,* $r \succsim s \Rightarrow \lambda pr \succsim \lambda qs$ *and* $r \succ s \Rightarrow \lambda pr \succ \lambda qs$*.*

By letting $p = q$ in *Weak Local Bi-Independence*, we obtain *Weak Local Independence*. Note that we do not require that L_p and L_q preserve bi-independence. We say that two subsets L_1, L_2 of $\mathcal{L}$ *preserve bi-independence* if for any $i \in \{1, 2\}$, $p, r \in L_i$, $q, s \in L_{3-i}$, and $\lambda \in (0, 1)$ such that $\lambda pr \in L_i$ and $\lambda qs \in L_{3-i}$, if $p \succsim q$, then $r \succsim s \Rightarrow \lambda pr \succsim \lambda qs$ and $r \succ s \Rightarrow \lambda pr \succ \lambda qs$.

Clearly, if L_p and L_q preserve bi-independence for any $p, q \in \mathcal{L}$, (*Bi-*)*Independence* holds globally. Hence, *Weak Local Bi-Independence* weakens *Bi-Independence* in a similar fashion to how *Weak Local Independence* weakens *Independence*. We fix some arbitrary p and q, and then require that *Bi-Independence* hold with respect to p and q, respectively.

Weak Local Bi-Independence rules out many kinds of choice behavior. We will illustrate this in Section 2.4 via an example that is not allowed by *Weak Local Bi-Independence* but is consistent with *Weak Local Independence*. Below is the representation theorem.

Theorem 1 *The preference* $\succsim$ *has a NEU representation if and only if* $\succsim$ *satisfies Weak Order, Continuity, and Weak Local Bi-Independence.*

The axioms in expected utility theory characterize linear functions on $\mathcal{L}$. Our axioms characterize piecewise-linear functions on $\mathcal{L}$, as will be explained in Section 2.4. Piecewise-linear functions have been important in decision theory (see Section 1.3) and their behavioral characterization is often more challenging than it might appear (see Siniscalchi (2006)).

In addition to helping us understand behavioral implications of the NEU representation, our axiomatic characterization helps us select a particular class of models from a myriad of machine-learning models: We know that many machine-learning models are useful, but which might be a good economic model of how people make decisions (under risk)? Our axioms suggest the neural-network models.

Moreover, how we apply the neural-network model to decision-making under risk is novel due to the axioms: If we ask a statistician or computer scientist to apply the neural-network model to decision-making under risk, they are likely to do this differently. Our NEU model takes a lottery as the input and gives the *unobservable* utility of the lottery as the output. Then, facing multiple lotteries, the NEU model assigns utility to each of them and predicts that the lottery with the highest utility will be chosen. By contrast, given the binary-choice preference data, a statistician or computer scientist will likely take *two* lotteries as the input and apply a neural network to them jointly, rather than applying the same neural network to each lottery separately. In the end, some output layer will predict the chance that each of the two lotteries is chosen. This kind of neural-network model is different from the NEU model and may violate our axioms, including, at least, transitivity in *Weak Order*.

2.3 Behavioral Neurons

We introduce a few examples to show how a NEU representation captures the main idea of well-known empirical findings. These examples will be useful in our empirical analysis.

2.3.1 Behavioral Neurons of Certainty Effects

Suppose $Z = \{z_1, z_2, z_3\}$. Recall that decision makers are often biased toward certainty. Figure 2 presents a NEU function in which the first neuron captures standard expected utility evaluation, while the other three neurons capture the bias toward certainty for z_1, z_2, and z_3, respectively.[8] We call the neurons that capture the bias toward certainty *certainty-effect neurons*.

The first neuron in Figure 2, $\max\{V(p), -\infty\}$, does not compare the outcome of aggregation to zero, as required for a NEU function. This is for simplicity and without loss of generality, because $\max\{V(p), -\infty\} = V(p) = \max\{V(p), 0\} - \max\{-V(p), 0\}$.[9]

[8] 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.

[9] Alternatively, let $\underline{V} = \min_{p \in \mathcal{L}} V(p)$. If $\underline{V} \geqslant 0$, $\max\{V(p), -\infty\} = \max\{V(p), 0\}$. Otherwise, we replace

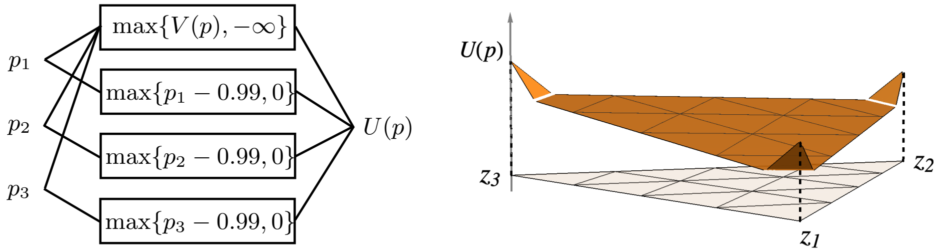

FIGURE 2: In the first neuron, V is an expected utility function, which is affine. If $p_i > 0.99$ for some $i \in \{1, 2, 3\}$, a neuron that captures the bias toward certainty will be activated. Finally, $U(p)$ is equal to some weighted sum of all neurons' values.

2.3.2 Behavioral Neurons of Reference Dependence

The second example is related to reference dependence. Suppose Z is some finite set of numbers (monetary prizes). As pointed out by Kahneman and Tversky (1979) and many other papers, prizes are often evaluated relative to a reference point, and people treat gains and losses (i.e., prizes better than and worse than the reference point, respectively) differently. In addition, it is also documented that the difference disappears when prizes do not deviate much from the reference point (see, for example, Ert and Erev (2013)).

A NEU function with one hidden layer can capture these ideas. Let there be two neurons in the first hidden layer. Suppose v is a Bernoulli index. The first neuron, $V(p) = \max\left\{\sum_i p_i v(z_i), -\infty\right\}$, again computes the expected utility of p. The second neuron captures loss aversion relative to prize x with a threshold ε:

$$V_l(p) = \max\left\{\sum_{i=1}^{n} p_i \max\{x - z_i, 0\} - \varepsilon, 0\right\}.$$

We call such neurons *reference-dependence neurons.* Note that $\sum_i p_i \max\{x - z_i, 0\} - \varepsilon$ is an affine function of p, and the loss of prize z_i is given by $\max\{x - z_i, 0\}$. If the expected loss is larger than ε, the neuron V_l is activated. Finally, $U(p) = V(p) - \lambda V_l(p)$, with λ being the loss-aversion coefficient. Clearly, in this NEU function, loss aversion only occurs when prizes deviate from the reference point significantly.

the first neuron with $\max\{V(p) - \underline{V}, 0\} = V(p) - \underline{V}$. Then, at its parent neurons, we add $\underline{V}$ back into the affine aggregation.

2.4 Sketch of the Proof

Tarela and Martínez (1999) show that a continuous piecewise-linear function has a lattice representation; that is, there exists affine functions $U_1, \ldots, U_m$ and $s_1, \ldots, s_k \subseteq \{1, \ldots, m\}$ such that the continuous piecewise-linear function can be represented as $\max_{1 \leqslant j \leqslant k} \min_{i \in s_j} U_i$. Wang and Sun (2005) and Arora, Basu, Mianjy, and Mukherjee (2018) prove that a lattice representation can in turn be converted into a neural-network model with rectified linear units, which becomes a NEU function when the domain is $\mathcal{L}$. Therefore, to establish the sufficiency of the axioms, we only need to show that $\succsim$ can be represented by a continuous piecewise-linear function, defined as follows.

Definition 3 *A function $U : \mathcal{L} \to \mathbb{R}$ is continuous piecewise-linear if U is continuous and there exist finitely many closed sets in $\mathcal{L}$ whose union is $\mathcal{L}$ such that U is affine on each of those closed sets.*

Our proof contains three main parts. First, we identify the interior of all regions over which the preference will be represented by an affine function. By *Weak Local Independence* (implied by *Weak Local Bi-Independence*), every lottery p has a neighborhood L_p that preserves independence with respect to p. Pick $q \in L_p$, we can find $L_q \subseteq L_p$ that preserves independence with respect to q. Inductively, we can find $n-1$ lotteries inside L_p such that the polytope formed by these $n-1$ lotteries and p preserves independence. Figure 3 illustrates the construction when $n = 3$.

Hence, we can find a polytope containing p that preserves independence for every lottery p. It may seem that this is almost sufficient for us to construct a piecewise-linear representation, but if this is true, we only need *Weak Local Independence* rather than *Weak Local Bi-Independence*. Figure 4 illustrates why *Weak Local Bi-Independence* is crucial with a pathological example that only satisfies *Weak Local Independence*.

This suggests that when constructing polytopes, we need more than *Weak Local Independence*. Exploiting *Weak Local Bi-independence*, we show that if the neighborhood in

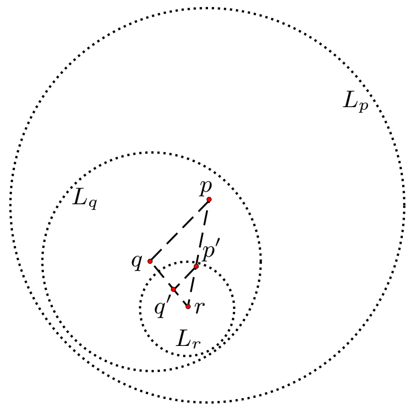

FIGURE 3: Let $n = 3$. Without loss of generality, assume that $L_r \subseteq L_q \subseteq L_p$. It is clear that the polytope $\overline{p'q'r}$ preserves independence with respect to p, q, r. We then show that $\overline{p'q'r}$ preserves independence with respect to p', q', r, which implies that $\overline{p'q'r}$ also preserves independence. The last step is to show that $\overline{pqr}$ preserves independence based on the fact that $\overline{p'q'r}$ preserves independence with respect to p, q.

each step of the construction is sufficiently small, each edge of the polytope will preserve bi-independence with *any* open convex subset that preserves independence. This ensures that every pair of polytopes constructed in this way preserves bi-independence. Now, intuitively, each of these polytopes must be part of one (linear) region. We take the interior of the union of these polytopes and denote it by $\mathcal{L}_o$.

Next, we identify each region via Zorn's lemma. First, we consider the set of all functions that map $\mathcal{L}_o$ into subsets of $\mathcal{L}$ that preserve independence individually and bi-independence pairwisely. By Zorn's lemma, we are able to find a maximal element among these functions that assigns $p \in \mathcal{L}_o$ a maximal region that satisfies the required properties. The image of this maximal function identifies all of the regions. Moreover, we show that for *Weak Local Independence* to hold everywhere, the number of regions must be finite.

The third and last part is to construct a continuous piecewise-linear representation of the preference. First, by *Weak Order* and *Continuity*, we obtain a continuous utility representation of the preference. Second, we show that if a collection of subsets of $\mathcal{L}$ preserve independence individually and bi-independence pairwisely, one can construct a piecewise-linear representation on the union of these subsets. The proof of this step is closely related to Chapter 2.4 of Schmidt (1998) and Ellis and Masatlioglu (2020) (see their Appendix B).

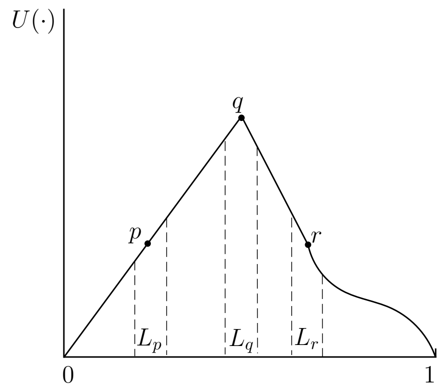

FIGURE 4: Let $n = 2$. Then each lottery is identified with a number in $[0, 1]$. The decision maker's utility function U is shown in the figure, in which the horizonal axis is the set of all lotteries. In this example, every lottery $p \in [0, 1]$ has a neighborhood that preserves independence with respect to p. In particular, L_r preserves independence with respect to r since U is monotone within L_r. However, p and r do not satisfy the requirement in *Weak Local Bi-independence*, since any neighborhood of r includes a nonlinear segment. In fact, none of the lotteries in $(r, 1]$ is contained in a polytope (in this case a line segment) that satisfies the bi-independence requirement.

Moreover, *Continuity* ensures that the piecewise-linear representation must also be continuous. Finally, we perform monotone transformations on the continuous utility representation to construct the continuous piecewise-linear representation on $\mathcal{L}$. In particular, for each batch of regions that overlap in utility values, we transform the continuous representation into a continuous piecewise-linear representation while keeping the utility values unchanged elsewhere. Since the piecewise-linear representation is unique only up to a positive affine transformation, we can exploit the two degrees of freedom to ensure that no discontinuity is introduced to the utility function by the transformation. Thus, the resulting utility function will be continuous and affine within each region we identified above.

3 Empirical Analysis

Although the NEU representation offers an interpretation of how the decision maker evaluates lotteries, if the neural network is too complex, the interpretation may not be intuitive or insightful. To see how well the NEU function performs empirically and how complex the neural network must be to explain and predict people's choice behavior well, we analyze the

NEU representation empirically.

3.1 Data Description and Training (Stochastic) Utility Models

We use the training and testing datasets provided by the aggregate-behavior track of the Choice Prediction Competition 2018[10] (see Plonsky et al. (2019)). The datasets come from several experiments conducted at the Hebrew University of Jerusalem and Technion–Israel Institute of Technology. Each participant in the experiments faces 750 binary choice problems over lotteries, in which a lottery is instantiated by the description of a probability vector defined over its support, a nonempty finite set of monetary prizes.[11] In each binary choice problem, a participant must choose one lottery of the two.

The 750 binary choice problems each participant faces consist of 30 different problems presented in a random order, and each of the 30 different problems is repeated 25 times consecutively. Different participants may face different binary choice problems.

In total, there are 270 different binary choice problems. Of these, 30 are designed to replicate 14 well-known behavioral phenomena, including the certainty effect, the reflection effect, overweighting of small probabilities, etc.[12] The other binary choice problems are generated randomly according to some rules.

Henceforth, when we say a binary choice problem, we mean one of the 270 different ones. Of these, 210 are in the training dataset and 60 in the testing dataset. The 30 replication binary choice problems are in the training dataset.

We aggregate individual choice data for each of the 270 binary choice problems; that is, for each binary choice problem, we calculate the fraction of participants choosing each lottery. We call these fractions *choice probabilities*. A data point contains information about the two lotteries (as the covariate) and the choice probabilities (as the response).

[10] The datasets are publicly available at `https://cpc-18.com`.

[11] Since the total number of prizes is finite, we continue using our notations from Section 2 for lotteries and prizes.

[12] The behavioral biases are successfully replicated, but the magnitude is smaller than in the original studies that document the biases. See Erev et al. (2017) for more details.

Some data from the experiments are not suitable for our analysis. First, our theory has nothing to say about binary choice problems that involve ambiguity (lotteries' probabilities are not specified) and little to say about binary choice problems in which realizations of lotteries are correlated. We exclude those binary choice problems from our analysis. Second, recall that each binary choice problem is repeated 25 times for each participant. After the first 5 repetitions, a participant observes feedback—that is, realizations of the lotteries from previous repetitions. Our theory has little to offer about how choice behavior will be affected by feedback. Therefore, we ignore the information about whether choices are made with feedback provided or not. Eventually, the training dataset contains 169 data points and the testing dataset contains 45 data points.

We estimate/train a model using the training dataset, and then evaluate its performance on the testing dataset. To estimate a model, we take a standard approach to combine it with the logit model (see Train (2003)). For example, suppose we want to analyze the expected utility model. Take any expected utility function $U : \mathcal{L} \to \mathbb{R}$. Given any data point with lotteries p and q, we use the probability that $U(p) + \varepsilon_p > U(q) + \varepsilon_q$ to predict the choice probability of p over q, in which ε_p and ε_q are independently and identically distributed random variables following the Gumbel distribution.[13] Following Plonsky et al. (2019), we use the mean squared error (MSE) of predicted choice probabilities (compared to actual choice probabilities) as the metric to evaluate the performance of U. Then, we find the expected utility function that minimizes the MSE using the training dataset, called the *training MSE*. Last, we take the estimated expected utility function to compute the MSE using the testing dataset, called the *testing MSE*, to measure the expected utility model's performance.[14] Obviously, the procedure above applies to other models, such as the NEU model. We replace the expected utility function U with a NEU function.

Note that by doing so, we turn a deterministic choice model described by maximizing

[13]The literature using this approach to estimate expected and non-expected utility models is immense. See Harrison, List, and Towe (2007) and Noussair, Trautmann, and van de Kuilen (2014) among others.

[14]More details about how we train the NEU model will be provided later.

some utility function into a stochastic one that maximizes some utility function under random noises. Our axiomatic characterization is for the deterministic choice model rather than the stochastic one. However, extending our characterization to the stochastic choice model is straightforward and well understood in the literature. We take a stochastic choice function as the primitive, and derive a stochastic preference from it. Then, for example, we impose the axioms in Section 2.2 to the stochastic preference. The resulting model will essentially be the logit version of the NEU model. We only need to add axioms that give us the logit model. See Ke (2018) for a related exercise.

Finally, also note that we could have first divided the datasets by participants' types and then estimated a model for each type using the approach described above. However, as will be explained in Footnote 17, this turns out to worsen the overfitting problem at least for the benchmark. Therefore, we choose not to do so.

3.2 The Benchmark and Necessity of Parametrization

Two classic theories will be used as our benchmark, expected utilty theory and cumulative prospect theory (CPT). Our first observation is that under the current data, we must parametrize the model, including the expected utility model, to avoid overfitting.

Let us illustrate this using the expected utility model. The expected utility model is $U(p) = \sum_{i=1}^{n} p_i u(z_i)$ for each lottery p, in which u is the Bernoulli index. Also consider the CARA expected utility model:[15] For each lottery p,

$$U_{\text{CARA}}(p) = \sum_{i=1}^{n} p_i u(z_i) \text{ with } u(z_i) = \frac{\beta}{\alpha}(e^{\alpha z_i} - 1). \tag{2}$$

As usual, $-\alpha$ measures the decision maker's risk aversion. As α approaches 0, $u(z_i)$ converges to βz_i, which is a risk-neutral Bernoulli index. The parameter $\beta \in \mathbb{R}_+$ is a normalization parameter that is necessary in discrete choice estimation.

[15]We do not consider the alternative and equally popular expected utility model, the constant-relative-risk-aversion (CRRA) expected utility model, mainly because some prizes are negative and hence are not well defined for the CRRA Bernoulli index.

Combining these with the logit model, as explained in the previous subsection, we can find the best expected utility function and the best CARA expected utility function. The training and testing MSE$\times$100 for the CARA expected utility model are 2.28 and 1.98, respectively, with essentially zero standard deviations. Compared with the CARA expected utility model, the expected utilty model's training MSE is certainly lower, since the expected utility model is more general, but its testing MSE turns out to be about 10 times higher.[16]

This is due to overfitting. A more general model can explain more phenomena (have a lower training MSE), but that does not imply that it will predict well.[17] The classic example is to use a polynomial to fit a dataset generated by a linear function plus noises. Therefore, we will need some parametrization for the CPT model and the NEU model as well.

Next, we examine the CPT benchmark, which is arguably the most popular non-expected utility model. Our parametrization is standard. The probability weighting function is

$$\pi(a) = \frac{\delta a^{\gamma}}{\delta a^{\gamma} + (1-a)^{\gamma}}$$

for any probability $a \in [0,1]$. The value function takes the CARA form in both the gain and loss regions, with a loss-aversion coefficient weakly larger than 1. We allow the reference point to be endogenously estimated. Note that due to the convexity of the value function in the loss region, even when probabilities are not distorted the CARA expected utility model is not a special case of the CPT model. These two models only intersect at the risk-neutral case (without probability distortion).

We find that CPT's training MSE$\times$100 is 2.255, and testing MSE$\times$100 is 1.996 (standard

[16] If we have a large amount of data, the expected utility model should outperform the CARA expected utility model, but given the current dataset, the expected utility model's training MSE$\times$100 is 1.07 and its testing MSE$\times$100 is 19.47. Note that there are prizes that appear in the testing dataset but not in the training dataset. For those prizes, their Bernoulli indices cannot be estimated. However, even if we exclude the binary choice problems that contain prizes that only appear in the testing dataset, the testing MSE$\times$100 of the expected utility model only reduces to 17.82, which is still much higher than that of the CARA expected utility model.

[17] We also examine a generalization of the above CARA expected utility model by allowing the CARA parameter to depend on participants' genders. The resulting testing MSE$\times$100 is 2.03 for female and 2.14 for male. Therefore, the overfitting problem gets worse. This is because when we divide the datasets by gender into two halves, the data in each half becomes noisier.

deviations are 0.022 and 0.099, respectively). Hence, under the current dataset, CPT does not seem to outperform the (CARA) expected utility model in terms of predictive power, although its performance is significantly better than the risk-neutral expected utility model.[18] One potential reason is that the current dataset does not focus on the kind of lotteries involved in the fourfold pattern of risk attitudes. In any case, our estimation suggests little probability distortion, which is consistent with recent findings by Bernheim and Sprenger (2020). Therefore, for the rest of the paper, we use the CARA expected utility model as the benchmark, whose testing MSE×100 is 1.98.

3.3 The NEU Model and Behavioral Neurons

The dataset is also too small for the general NEU model. We need to find parametrization that helps overcome overfitting and, at the same time, keeps its flexibility in the right direction. The first idea comes from the observation that the affine functions in the first hidden layer are expected utility functions. Therefore, it is possible that replacing first-hidden-layer neurons with CARA expected utility functions could help. The resulting model is still a NEU model, even though we do not use the activation function in the first hidden layer (see Section 2.3.1).

It turns out that this restriction destroys too much flexibility of the NEU model. We allow the first hidden layer's width to be 15, 20, or 25; the number of hidden layers above the first to be 0, 1, or 2; and the width of the hidden layers above the first to be 15, 20, or 25. The best testing MSE×100 we obtain out of these NEU functions is 1.97, which is barely better than the CARA benchmark.

To see what kind of useful flexibility has been removed, consider the certainty-effect neurons and reference-dependence neurons from Sections 2.3.1 and 2.3.2. These are neurons

[18] The testing error is still higher than the CARA expected utility model's if the value function is parametrized via the CRRA form, or if we allow the value function to be convex in the gain region or concave in the loss region so that the CARA expected utility model is nested as a special case. We also check the testing error of the estimated CPT model from Tversky and Kahneman (1992), which is significantly higher.

in the first hidden layer and help us capture well-documented behavioral effects, but are assumed away if we focus on CARA expected utility functions for the first hidden layer.

Presumably, there may be other useful behavioral neurons and we may use statistical methods to endogenously select which behavioral neurons are best to use. This is an interesting and more general approach, but is beyond the scope of this paper. Below, we will show that by requiring that the first hidden layer consist of a CARA expected utility function, certainty-effect neurons, and reference-dependence neurons, we overcome overfitting of the NEU model without destroying its useful flexibility. This allows us to improve the NEU model's performance significantly, compared with the case with only CARA expected utility functions in the first hidden layer and the original NEU model.[19]

To explain how this is done, we need to describe how we estimate the NEU model based on the training dataset using *cross-validation*. Each NEU model has some *hyperparameters*, such as the number of hidden layers and the width of each hidden layer. Obviously, if we select hyperparameters by minimizing the training MSE, we will want bigger networks and overfit. Therefore, for each set of hyperparameters, we use the training dataset to estimate the NEU model and compute the *leave-one-out* cross-validation (LOOCV) MSE (see Chapters 5 and 7 of Hastie, Tibshirani, and Friedman (2009)). We find the hyperparameters that yield the lowest LOOCV MSE. Then, we train the NEU model under the selected hyperparameters using the training dataset and compute its training MSE. Finally, we take the trained NEU model to compute the testing MSE using the testing dataset.

We consider the NEU model in which the first hidden layer consists of three types of neurons to be described below. We call this first hidden layer a *behavioral-neuron layer*. The behavioral-neuron layer is subsequently concatenated with additional standard hidden layers (defined in Section 2.1), except that we will apply one restriction to the second-hidden-layer

[19] This approach is similar to using domain knowledge to engineer features in statistics and computer science, except that we will only consider features that are consistent with our axioms. It is well known that domain knowledge may help improve the performance of machine-learning models via feature engineering. See, among others, convolutional neural-network models (Chapter 9 in Goodfellow, Bengio, and Courville (2016)) in image recognition.

neurons due to the use of certainty-effect neurons, which we will explain shortly.

Every lottery p is mapped to its utility $U(p)$ via such a NEU model. However, we cannot directly train the NEU model using state-of-the-art machine-learning methods, because the output of the NEU model is utility, which is not observable in the data. To see how we train the NEU model, recall that each data point has two lotteries p and q. Hence, the exact neural-network model to be trained is as follows. First, it takes both p and q as the input. Next, it derives $U(p)$ and $U(q)$ through two *separate and identical* NEU models. Last, it uses the probability that $U(p) + \varepsilon_p > U(q) + \varepsilon_q$ as the output to predict the choice probability of p for this data point, in which ε_p and ε_q are again independent and identically distributed random variables following the Gumbel distribution. We call such a neural-network model the *logit symmetric behavioral NEU* model, which may be trained using state-of-the-art machine-learning methods. Note that this model is a function defined on $\mathcal{L} \times \mathcal{L}$, so it is not a special case of the NEU model. It is constructed based on the NEU model for estimation.

Now, we describe the behavioral-neuron layer. It consists of three types of neurons:

1. *The CARA neuron*: A CARA neuron is the function $U_{CARA} : \mathcal{L} \to \mathbb{R}$ defined in (2). When estimating its parameters α and β, α is initialized uniformly at random in $[-1, 1]$ and β is initialized uniformly at random in $[0, 1]$.

2. *The certainty-effect (CE) neuron*: For any lottery p, a CE neuron with respect to prize z_i is a function from $\mathcal{L}$ to $\mathbb{R}$ that takes the following form:

$$U^j_{\mathrm{CE}_i}(p) = \max\{p_i - \eta_j, 0\},$$

   $j = 1, 2$ (see Section 2.3.1); that is, we allow the decision maker to have two types of CE neurons, each with a possibly different η_j (the threshold parameter).

   Fixing j, we require that η_j be identical across different prizes. Otherwise, if some prize never shows up in the training dataset, we will not be able to estimate those two parameters for that prize. For the same reason, fixing j, we require that every second-hidden-layer neuron attach the same weight to the CE neurons $U^j_{CE_i}$, $i = 1, \ldots, n$, in the affine aggregation. This is the restriction on the second hidden layer we mention above. The threshold parameter η_j is initialized uniformly at random in $[0.9, 0.99]$.

3. *The reference-dependence (RD) neuron*: For any lottery p, an RD neuron is a function from $\mathcal{L}$ to $\mathbb{R}$ that takes the following form:

$$U^j_{\mathrm{RD}}(p) = \max\left\{\sum_{i=1}^{n} p_i \max\{\lambda_j z_i - \gamma_j, 0\}, \kappa_j\right\},$$

$j \in \{1, \ldots, n_{\mathrm{RD}}\}$; that is, we allow the decision maker to have n_{RD} types of RD neurons, each of which is characterized by three parameters: the loss-aversion coefficient $-\lambda_j$, the reference point $\frac{\gamma_j}{\lambda_j}$, and the threshold parameter κ_j. All three parameters are initialized at random according to the standard Gaussian distribution, and n_{RD} is left as a hyperparameter.

To summarize, in the behavioral-neuron layer, we have one CARA neuron, two types of CE neurons, and n_{RD} RD neurons. We do not turn the number of types of CE neurons into a hyperparameter only to shorten computation time. The CE neurons only affect a small area of $\mathcal{L}$, and we believe that allowing for two types of CE neurons will be adequate. We only use one CARA neuron because we have seen that having multiple CARA neurons does not seem to help. Then, additional hidden layers will be concatenated with the behavioral-neuron layer, whose number of layers and width are hyperparameters. The affine-aggregation parameters of these layers are initialized by the standard Gaussian distribution.

3.3.1 Training, Regularization, and Hyperparameter Selection

We train the NEU model using adaptive moment estimation (also known as Adam; see Kingma and Ba (2017)) with minibatches of size 10, which are randomly selected at each epoch. The learning rate is 0.0002 for parameters of the additional hidden layers and the output layer, and is 0.00002 for the parameters of the behavioral-neuron layer. To regularize the training of the logit symmetric behavioral NEU model, we use ℓ_2-norm regularization with coefficient 0.0002 (see Chapter 7 of Goodfellow, Bengio, and Courville (2016)). Meanwhile, we stop the training early, after $3,000$ epochs, in which each epoch goes through the 17 minibatches in a random order. Again, we use the MSE of the model as the metric to evaluate the performance of a logit symmetric behavioral NEU model.

We divide the logit symmetric behavioral NEU models into three groups. In the first group, the NEU model has one hidden layer; that is, it only has the behavioral-neuron layer. In the second, the NEU model has two hidden layers, and in the third, the NEU model has strictly more than two hidden layers. We divide them in this way because NEU models with one hidden layer and NEU models with two layers have rather different interpretations, and both are much easier to interpret than NEU models with more layers.

Within each group, to select the most desirable model (i.e., the most desirable set of hyperparameters) from candidate models on the training dataset of size 169, we use LOOCV. The width of the additional hidden layers (if any) concatenated with the behavioral-neuron layer may be 15, 20, or 25. The number of RD neurons may be 15, 20, or 25. In the third group, the number of the additional hidden layers concatenated with the behavioral-neuron layer may be 2 or 3.

LOOCV trains each candidate model on only 168 data points and then makes a prediction on the left-out data point. Each of the 169 data points will be left out once. Then, LOOCV selects the candidate model with the least average LOOCV MSE over the 169 choices of left-out data points.

Given the selected model (set of hyperparameters), we train the logit symmetric behavioral NEU model on the training dataset of size 169, and then the trained model is taken to the testing dataset of size 45 to compute the testing MSE.

3.3.2 Results

We report the testing MSE×100 in the following table for the three groups of NEU models along with their standard deviations after 50 repetitions of experiments:

	CARA+RD	CARA+CE	CARA+RD+CE
1 Hidden Layer	1.971 (0.113)	2.009 (0.006)	1.966 (0.133)
2 Hidden Layers	1.850 (0.176)	2.030 (0.022)	1.748 (0.221)
>2 Hidden Layers	2.217 (0.339)	2.130 (0.038)	1.879 (0.315)

In the "CARA+RD" column, we require that the number of CE neurons be zero in the behavioral-neuron layer. In the "CARA+CE" column, we require that the number of RD neurons instead be zero. In the "CARA+RD+CE" column, the setup of the behavioral-neuron layer is the same as described earlier.

The above table shows that introducing the CARA, CE, and RD neurons significantly improves the performance of the NEU model. Measured by the testing MSE, the two-hidden-layer NEU model (with the first hidden layer being the behavioral-neuron layer) has the best performance. Its testing error is significantly lower than the CARA benchmark. Note that NEU models' training MSE will always be lower than the CARA benchmark.

Moreover, it can be seen that introducing only one type of behavioral neuron together with the CARA neuron is not sufficiently helpful. It is when both CE and RD neurons are included in the behavioral-neuron layer that the NEU model significantly outperforms the CARA benchmark.

These results suggest that our domain knowledge from decision theory and behavioral economics can be useful in predictions (see Footnote 19), and that reasonably complex NEU functions with intuitive interpretation have the best performance.

4 Concluding Remarks

Observing that previous studies on violations of expected utility theory suggest that the key axiom, (*Bi-*)*Independence*, does not hold globally but may hold locally, we introduce a novel way to weaken (*Bi-*)*Independence* and require that it hold locally. The resulting representation of the decision maker's preference, the NEU representation, is a novel way to apply the neural-network model to decision-making under risk.

Our axiomatic characterization shows that among numerous machine-learning models, the neural-network model applied to our choice domain with uncertainty in the form of a NEU model has a simple and reasonable behavioral foundation. Thus, the NEU model may be particularly suitable to be applied to economic models in future research compared with other machine-learning models.

Empirically, we find that relatively simple NEU models that are easy to interpret have strong predictive power. Moreover, when the training dataset is not sufficiently large, we show that neurons constructed from our domain knowledge about decision making can be useful in improving the NEU model's performance.

We expect that the NEU model and its potential future generalizations will become more and more useful in economics, as we accumulate more economic data about people's choice behavior.

References

Amarante, M. (2009). Foundations of Neo-Bayesian Statistics. *Journal of Economic Theory 144*(5), 2146–2173.

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

Battalio, R., J. Kagel, and K. Jiranyakul (1990). Testing between Alternative Models of Choice under Uncertainty: Some Initial Results. *Journal of Risk and Uncertainty 3*(1), 25–50.

Bernheim, B. D. and C. Sprenger (2020). On the Empirical Validity of Cumulative Prospect Theory: Experimental Evidence of Rank-Independent Probability Weighting. *Econometrica* (forthcoming).

Bordalo, P., N. Gennaioli, and A. Shleifer (2012). Salience Theory of Choice Under Risk. *Quarterly Journal of Economics 127*(3), 1243–1285.

Cerreia-Vioglio, S. (2009). Maxmin Expected Utility on a Subjective State Space: Convex Preferences under Risk. Working Paper.

Cerreia-Vioglio, S., D. Dillenberger, and P. Ortoleva (2015). Cautious Expected Utility and the Certainty Effect. *Econometrica 83*(2), 693–728.

Chandrasekher, M., M. Frick, R. Iijima, and Y. Le Yaouanq (2020). Dual-self Representations of Ambiguity Preferences. Working Paper.

Chatterjee, K. and R. V. Krishna (2011). A Nonsmooth Approach to Nonexpected Utility Theory under Risk. *Mathematical Social Sciences 62*(3), 166 – 175.

Chew, S. H. (1983). A Generalization of the Quasilinear Mean with Applications to the Measurement of Income Inequality and Decision Theory Resolving the Allais Paradox. *Econometrica 51*(4), 1065–92.

Choi, S., R. Fisman, D. Gale, and S. Kariv (2007). Consistency and Heterogeneity of Individual Behavior under Uncertainty. *American Economic Review 97*(5), 1921–1938.

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

Dekel, E. (1986). An Axiomatic Characterization of Preferences under Uncertainty: Weakening the Independence Axiom. *Journal of Economic Theory 40*(2), 304–318.

Dillenberger, D. (2010). Preferences for One-Shot Resolution of Uncertainty and Allais-Type Behavior. *Econometrica 78*, 1973 – 2004.

Dillenberger, D., V. Krishna, and P. Sadowski (2018). Subjective Information Choice Processes. Working Paper.

Ellis, A. and Y. Masatlioglu (2020). Choice with Endogenous Categorization. Working Paper.

Erev, I., E. Ert, O. Plonsky, D. Cohen, and O. Cohen (2017). From Anomalies to Forecasts: Toward a Descriptive Model of Decisions under Risk, under Ambiguity, and from Experience.

Psychological Review 124(4), 369–409.

Ert, E. and I. Erev (2013). On the Descriptive Value of Loss Aversion in Decisions under Risk: Six Clarifications. *Judgment and Decision Making 8*(3), 214–235.

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.

Goodfellow, I., Y. Bengio, and A. Courville (2016). *Deep Learning*. MIT Press.

Gul, F. (1991). A Theory of Disappointment Aversion. *Econometrica 59*(3), 667–686.

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.

Hahnloser, R., H. Seung, and J. Slotine (2003). Permitted and Forbidden Sets in Symmetric Threshold-Linear Networks. *Neural Computation 15*(3), 621–638.

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.

Harless, D. and C. Camerer (1994). The Predictive Utility of Generalized Expected Utility Theories. *Econometrica 62*(6), 1251–1289.

Harrison, G. W., J. A. List, and C. Towe (2007). Naturally Occurring Preferences and Exogenous Laboratory Experiments: A Case Study of Risk Aversion. *Econometrica 75*(2), 433–458.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag.

Hong, C. S. and W. S. Waller (1986). Empirical Tests of Weighted Utility Theory. *Journal of Mathematical Psychology 30*(1), 55 – 72.

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

Ke, S. (2018). Rational Expectation of Mistakes and a Measure of Error-proneness. *Theoretical*

Economics 13(2), 527–552.

Kingma, D. and J. Ba (2017). Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980v9.

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

Maccheroni, F. (2002). Maxmin under Risk. *Economic Theory 19*(4), 823–831.

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

Machina, M. (1987). Choice Under Uncertainty: Problems Solved and Unsolved. *Journal of Economic Perspectives 1*(1), 121–154.

Murdoch, W., C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu (2019). Definitions, Methods, and Applications in Interpretable Machine Learning. *Proceedings of the National Academy of Sciences 116*(44), 22071–22080.

Noussair, C. N., S. T. Trautmann, and G. van de Kuilen (2014). Higher Order Risk Attitudes, Demographics, and Financial Decisions. *Review of Economic Studies 81*(1), 325–355.

Plonsky, O., R. Apel, I. Erev, E. Ert, and M. Tennenholtz (2017). When and How Can Social Scientists Add Value to Data Scientists? A Choice Prediction Competition for Human Decision Making. White Paper for the Choice Prediction Competition 2018.

Plonsky, O., R. Apel, E. Ert, M. Tennenholtz, D. Bourgin, J. C. Peterson, D. Reichman, T. L. Griffiths, S. J. Russell, E. C. Carter, J. F. Cavanagh, and I. Erev (2019). Predicting Human Decisions with Behavioral Theories and Machine Learning.

Plonsky, O., I. Erev, T. Hazan, and M. Tennenholtz (2017). Psychological Forest: Predicting Human Behavior. In *Proceedings of the Thirty-first AAAI Conference on Artificial Intelligence*, AAAI-17, pp. 656–662.

Quiggin, J. (1982). A Theory of Anticipated Utility. *Journal of Economic Behavior & Organization 3*(4), 323–343.

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.

Starmer, C. (2000). Developments in Non-expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk. *Journal of Economic Literature 38*(2), 332–382.

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

Train, K. (2003). *Discrete Choice Methods with Simulation.* Cambridge University Press.

Tversky, A. and D. Kahneman (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty 5*(4), 297–323.

Wang, S. and X. Sun (2005). Generalization of Hinging Hyperplanes. *IEEE Transactions on Information Theory 51*(12), 4425–4431.

Wu, G., J. Zhang, and M. Abdellaoui (2005). Testing Prospect Theories Using Probability Tradeoff Consistency. *Journal of Risk and Uncertainty 30*(2), 107–131.

Yaari, M. E. (1987). The Dual Theory of Choice under Risk. *Econometrica 55*(1), 95–115.

Appendix

Proof of Theorem 1: The sufficiency of the axioms follows from Tarela and Martínez (1999); Wang and Sun (2005); and Arora et al. (2018), if we can prove that $\succsim$ can be represented by a continuous piecewise-linear function on $\mathcal{L}$. We prove this via a sequence of lemmas, and without mentioning this explicitly, we assume for each of them that the axioms hold. For any $L \subseteq \mathcal{L}$, let $\text{int}(L)$, $\text{cl}(L)$, ∂L and $\text{aff}(L)$ denote the interior, closure, boundary, and affine hull of L, respectively, under the subspace topology induced from the standard topology of $\mathbb{R}^n$. For $p \in \mathcal{L}$ and $\varepsilon > 0$, let $B_\varepsilon(p)$ denote the open ball centered at p with radius ε.

Since $\succsim$ satisfies *Weak Order* and *Continuity* and $\mathcal{L}$ is separable and connected, by Debreu (1954) it has a continuous utility representation $V : \mathcal{L} \to \mathbb{R}$.

Lemma 1 *Suppose $L \subseteq \mathcal{L}$ and int$(L) \neq \emptyset$. The following statements are equivalent:*

(i) L preserves independence;

(ii) if $p, q, r, s \in L$ and $p - q = \lambda(r - s)$ for some $\lambda > 0$, $p \succsim q \iff r \succsim s$.

Proof. (ii)$\Rightarrow$(i) is immediate. We prove that (i)$\Rightarrow$(ii). Suppose L preserves independence and $p, q, r, s \in L$ satisfy $p - q = \lambda(r - s)$ for some $\lambda > 0$. Clearly, $q + \lambda r = p + \lambda s$. Pick any $t \in \text{int}(L)$ and let $p' = \alpha pt$, $q' = \alpha qt$, $r' = \alpha rt$, $s' = \alpha st$ such that $\overline{p'q'r's'} \subseteq L$. Let $t_1 = \frac{1}{\lambda+1}q'r'$ and $t_2 = \frac{1}{\lambda+1}p'r'$. Note that by construction, $t_1 = \frac{1}{\lambda+1}p's'$. Since L preserves independence, $p \succsim q \iff p' \succsim q' \iff t_2 \succsim t_1 \iff r' \succsim s' \iff r \succsim s$. ■

Lemma 2 *Suppose $L \subseteq \mathcal{L}$ preserves independence and $int(L) \neq \emptyset$. Then, there exists some affine function $U : cl(L) \to \mathbb{R}$ that represents $\succsim$ on $cl(L)$.*

Proof. Pick any $r \in \text{int}(L)$. By definition, there exists some $B_\varepsilon(r) \subseteq L$. Since L is bounded, there exists some $\alpha \in (0, 1)$ such that $p \in L \Rightarrow \alpha pr \in B_\varepsilon(r)$. Since L preserves independence, $B_\varepsilon(r)$ must also preserve independence. Therefore, $\succsim$ restricted on $B_\varepsilon(r)$ satisfies *Weak Order*, *Continuity*, and *Independence*. Because $B_\varepsilon(r)$ together with the mixture operation forms a mixture space, the mixture space theorem implies that there exists an affine function $\tilde{U} : B_\varepsilon(r) \to \mathbb{R}$ of $\succsim$ on $B_\varepsilon(r)$. Without loss of generality, let $\tilde{U}(r) = 0$. Next, for each $p \in L$, define $U(p) = \tilde{U}(\alpha pr)/\alpha$. This function U must represent $\succsim$ on L, because for any $p, q \in L$, $p \succsim q \iff \alpha pr \succsim \alpha qr \iff \tilde{U}(\alpha pr) \geqslant \tilde{U}(\alpha qr) \iff U(p) \geqslant U(q)$. Note that U is affine: For any $p, q, \lambda pq \in L$,

$$U(\lambda pq) = \frac{1}{\alpha}\tilde{U}(\alpha(\lambda pq)r) = \frac{1}{\alpha}\tilde{U}(\lambda(\alpha pr)(\alpha qr)) = \lambda U(p) + (1 - \lambda)U(q).$$

Therefore, U is an affine function on a bounded subset of $\mathbb{R}^n$, which must be uniformly continuous. Then, there exists a unique extension of U from L to $\text{cl}(L)$: For any $p \in \text{cl(L)}$, take an arbitrary sequence $\{p^k\}_{k=1}^\infty$ in L that converges to p, $U(p) = \lim_{k\to\infty} U(p^k)$. Clearly $U : \text{cl}(L) \to \mathbb{R}$ is still affine.

By *Continuity*, U must represent $\succsim$ on $\text{cl}(L)$; that is, for any $p, q \in \text{cl}(L)$, $p \succ q \iff U(p) > U(q)$. We only prove that $p \succ q \Rightarrow U(p) > U(q)$. The other direction is similar. Recall that the continuous function $V : \mathcal{L} \to \mathbb{R}$ represents $\succsim$. Since $p \succ q$, $V(p) > V(q)$. Without loss of generality, let $V(p) = 1$ and $V(q) = 0$. According to the intermediate value theorem, we can find some $p', q' \in \overline{pq}$ such that $V(p') = 2/3$ and $V(q') = 1/2$.

Next, by *Continuity*, we can find a sequence $\{p^k\}$ in L that converges to p and a sequence $\{q^k\}$ in L that converges to q such that $p^k \succ p' \succ q' \succ q^k$ for each $k \in \mathbb{N}$. We already know that U represents $\succsim$ on L. Since V restricted to L also represents $\succsim$ on L, $U = f(V)$ for some strictly increasing function f. Thus, we have $U(p^k) = f(V(p^k)) > f(2/3) > f(1/3) > f(V(q^k)) = U(q^k)$ for each $k \in \mathbb{N}$, which means that $U(p) \geqslant f(2/3) > f(1/3) \geqslant U(q)$. ■

The following corollary follows from Lemma 2 immediately.

Corollary 1 *Suppose $L \subseteq \mathcal{L}$ preserves independence and int$(L) \neq \emptyset$. Then, cl(L) preserves independence.*

Lemma 3 *Suppose L is a nonempty convex subset of $\mathcal{L}$ and $p \in L$. Then, $L \perp p$ and $L \perp q$ implies $L \perp \alpha pq$ for any $\alpha \in (0, 1)$.*

Proof. Take any $r, s \in L$. We want to show that $r \succsim s$ if and only if $\beta r(\alpha pq) \succsim \beta s(\alpha pq)$ for any $\alpha, \beta \in (0, 1)$. Since $\beta r(\alpha pq) = (\alpha + \beta - \alpha\beta)\left(\frac{\beta}{\alpha+\beta-\alpha\beta} rp\right) q$ and $\beta s(\alpha pq) = (\alpha + \beta - \alpha\beta)\left(\frac{\beta}{\alpha+\beta-\alpha\beta} sp\right) q$, it suffices to show that for any $\alpha, \beta \in (0, 1)$, $r \succsim s$ if and only if

$$(\alpha + \beta - \alpha\beta)\left(\frac{\beta}{\alpha + \beta - \alpha\beta} rp\right) q \succsim (\alpha + \beta - \alpha\beta)\left(\frac{\beta}{\alpha + \beta - \alpha\beta} sp\right) q. \tag{3}$$

Since $L \perp p$, $r \succsim s \iff \frac{\beta}{\alpha+\beta-\alpha\beta} rp \succsim \frac{\beta}{\alpha+\beta-\alpha\beta} sp$. Note that $\frac{\beta}{\alpha+\beta-\alpha\beta} rp, \frac{\beta}{\alpha+\beta-\alpha\beta} sp \in L$ since L is convex and $p \in L$. Then, (3) follows from the fact that $L \perp q$. ■

Lemma 4 *Suppose L is a convex subset of $\mathcal{L}$ such that int$(L) \neq \emptyset$ and L preserves independence. Then, $L \perp p$ implies that co$(L \cup \{p\})$ preserves independence.*

Proof. Take any $q \in \text{co}(\text{int}(L) \cup \{p\}) \backslash \{p\}$. There exists $q' \in \text{int}(L)$ and $\alpha \in (0, 1]$ such that $q = \alpha q' p$. Since $q' \in \text{int}(L)$, there exists some $\varepsilon > 0$ such that $B_\varepsilon(q') \subseteq \text{int}(S)$. For any $r, s \in B_{\alpha\varepsilon}(q)$, let $r' = \frac{r-q}{\alpha} + q'$, $s' = \frac{s-q}{\alpha} + q'$. Since $||r - q||, ||s - q|| < \alpha\varepsilon$, $r', s' \in B_\varepsilon(q')$ and

$\alpha(r'-s') = r-s$. Moreover,

$$\alpha r' p = r - q + \alpha q' + (1-\alpha)p = r - \alpha q' - (1-\alpha)p + \alpha q' + (1-\alpha)p = r.$$

Similarly, $\alpha s' p = s$. Because $L \perp p$, $r \succsim s \iff r' \succsim s'$.

We first prove that $\text{co}(\text{int}(L) \cup \{p\}) \backslash \{p\}$ preserves independence. By Lemma 1, we only need to show that for any distinct $r, s, r^*, s^* \in \text{co}(\text{int}(L) \cup \{p\}) \backslash \{p\}$ 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, $\overline{rs} \subseteq \text{co}(\text{int}(L) \cup \{p\}) \backslash \{p\}$. 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, for any $k \in \{0, \ldots, m\}$, $t_k - t_{k+1} = \lambda_k(r-s)$ for some $\lambda_k > 0$, and there exist $t'_k, t'_{k+1} \in \text{int}(L)$ such that $t_k - t_{k+1} = \beta_k(t'_k - t'_{k+1})$ for some $\beta_k > 0$ and $t_k \succsim t_{k+1} \iff t'_k \succsim t'_{k+1}$. These observations imply that for any $k \in \{0, \ldots, m\}$, $t'_k - t'_{k+1} = \lambda_k(t'_0 - t'_1)$ for some $\lambda_k > 0$.

Since L preserves independence, by Lemma 1, for any $k \in \{0, \ldots, m\}$, $t'_k \succsim t'_{k+1} \iff t'_0 \succsim t'_1$. Since $L \perp p$, $t'_k \succsim t'_{k+1} \iff t_k \succsim t_{k+1}$. Then, it follows from transitivity that $r \succsim s \iff t'_0 \succsim t'_1$. Note that $r - s = \frac{\beta_0}{\lambda_0}(t'_0 - t'_1)$.

The same arguments apply to r^* and s^*: There exist some $t^*_0, t^*_1 \in \text{int}(L)$ such that $r^* \succsim s^* \iff t^*_0 \succsim t^*_1$ and $r^* - s^* = \lambda^*(t^*_0 - t^*_1)$ for some $\lambda^* > 0$. Since $r - s = \lambda(r^* - s^*)$, we know that $t_0' - t_1' = \alpha^*(t_0^* - t_1^*)$ for some $\alpha^* > 0$. By Lemma 1, $t_0' \succsim t_1' \iff t_0^* \succsim t_1^*$. Thus, $r \succsim s \iff r^* \succsim s^*$. This completes the proof that $\text{co}(\text{int}(L) \cup \{p\}) \backslash \{p\}$ preserves independence.

It is straightforward to verify that

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

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

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

Now we introduce some lemmas regarding the (bi-)independence of line segments.

Lemma 5 *If $\overline{pr} \perp p$, then $\overline{pr}$ preserves independence.*

Proof. Suppose $r \succ p$. Since $\overline{pr} \perp p$, it is clear that $\alpha r p \succ p$ for any $\alpha > 0$. To show that $\overline{pr}$ preserves independence, it suffices to show that $\alpha r p \succ \beta r p$ whenever $1 \geqslant \alpha > \beta \geqslant 0$. By way of contradiction, assume that there exists $1 \geqslant \alpha > \beta \geqslant 0$ such that $\beta r p \succsim \alpha r p$. Since $\overline{pr} \perp p$, we have $\frac{\beta}{\alpha}(\beta r p)p \succsim \frac{\beta}{\alpha}(\alpha r p)p$, which implies $\frac{\beta^2}{\alpha} r p \succsim \beta r p$. Inductively, we have $\frac{\beta^{n+1}}{\alpha^n} r p \succsim \frac{\beta^n}{\alpha^{n-1}} r p$ for any n. Then by *Continuity* and *Weak Order*, we have $p \succsim \alpha r p$, a contradiction.

The case in which $r \sim p$ follows from the observation that, due to $\overline{pr} \perp p$, $\alpha r p \sim p$ for any $\alpha \in [0, 1]$. The case in which $p \succ r$ is symmetric to the first case. ■

Lemma 6 *Let $\overline{pr}$ and $\overline{qs}$ each preserve independence. If $\overline{pr}$ and $\overline{qs}$ do not preserve bi-independence, then there exists $p', r' \in \overline{pr}$, $q', s' \in \overline{qs}$ and $\lambda \in (0,1)$ such that $p' \sim q'$, $r' \sim s'$, and $\lambda p'r' \not\sim \lambda q's'$.*

Proof. Suppose $\overline{pr}$ and $\overline{qs}$ do not preserve bi-independence. Without loss of generality, assume that $p \succsim q$, and there exists $\lambda \in (0,1)$ such that either (a) $r \sim s$ but $\lambda qs \succ \lambda pr$, or (b) $r \succ s$ but $\lambda qs \succsim \lambda pr$. In either (a) or (b), $\lambda qs \succsim \lambda pr$.

Case 1: $p \succsim r$ and $q \succsim s$. Since $\overline{pr}$ and $\overline{qs}$ each preserve independence, we have $q \succsim \lambda qs \succsim \lambda pr \succsim r$. Suppose $q \sim r$. Then $\lambda qs \sim \lambda pr$ is forced. Thus, for $\overline{pr}$ and $\overline{qs}$ to not preserve bi-independence, we need $r \succ s$ and thus $q \succ s$. Then we must have $\lambda pr \succsim r \sim q \succ \lambda qs$, which is a contradiction. Hence, $q \succ r$. By $p \succsim q \succ r \succsim s$ and *Continuity*, there exist $p' \in \overline{pr}$ and $s' \in \overline{qs}$ such that $p' \sim q$ and $r \sim s'$. Then, since $\overline{pr}$ and $\overline{qs}$ each preserve independence, we have $\lambda qs' \succsim \lambda qs \succsim \lambda pr \succsim \lambda p'r$. Suppose we are in case (a), so that $r \sim s$ and $\lambda qs \succ \lambda pr$. Then it follows that $\lambda qs' \succ \lambda p'r$ and we are done. Suppose we are in case (b), so that $r \succ s$ and $\lambda qs \succsim \lambda pr$. Then we have $\lambda qs' \succ \lambda qs \succsim \lambda p'r$, because $s' \sim r \succ s$, and we are done.

Case 2: $r \succsim p$ and $s \succsim q$. Similarly to case 1, we have $s \succsim \lambda qs \succsim \lambda pr \succsim p$. Suppose $s \sim p$. Then we have $\lambda qs \sim \lambda pr$. Thus, for $\overline{pr}$ and $\overline{qs}$ to not preserve bi-independence, we need $r \succ s$ and thus $r \succ p$. Then it follows that $\lambda pr \succ p \sim s \succsim \lambda qs$, which is a contradiction. Hence, $s \succ p$. By $r \succsim s \succ p \succsim q$ and *Continuity*, there exist $r' \in \overline{pr}$ and $q' \in \overline{qs}$ such that $r' \sim s$ and $p \sim q'$. Then, since $\overline{pr}$ and $\overline{qs}$ each preserve independence, we have $\lambda q's \succsim \lambda qs \succsim \lambda pr \succsim \lambda pr'$. Suppose we are in case (a), so that $r \sim s$ and $\lambda qs \succ \lambda pr$. Then it follows that $\lambda q's \succ \lambda pr'$ and we are done. Suppose we are in case (b), so that $r \succ s$ and $\lambda qs \succsim \lambda pr$. Then we have $\lambda q's \succsim \lambda pr \succ \lambda pr'$, since $r \succ s \sim r'$.

Case 3: $p \succsim r$ and $s \succ q$. This case is impossible, since $\lambda pr \succsim r \succsim s \succ \lambda qs$.

Case 4: $r \succ p$ and $q \succsim s$. This case is impossible, since $\lambda pr \succ p \succsim q \succsim \lambda qs$. ■

Let $d(\cdot,\cdot)$ denote the metric function of $\mathcal{L}$.

Lemma 7 *Let $\overline{pr}$ and $\overline{qs}$ each preserve independence. If $\overline{pr}$ and $\overline{qs}$ do not preserve bi-independence, then for any $\varepsilon > 0$ there exist $q', s' \in \overline{qs}$ with $d(q', s') < \varepsilon$ such that $\overline{pr}$ and $\overline{q's'}$ do not preserve bi-independence.*

Proof. By the previous lemma, without loss of generality, assume that $p \sim q \succ r \sim s$. Let $q^0 = q$, $s^0 = s$, and $t^1 = \frac{1}{2}qs$. Suppose that $\overline{pr}$ and $\overline{qt^1}$, $\overline{pr}$ and $\overline{t^1 s}$ both preserve bi-independence. Since $\overline{pr}$ preserves independence, there exists a monotone transformation f such that $U(\lambda pr) = f \circ V(\lambda pr) = \lambda$ for any $\lambda \in [0,1]$. Let $\alpha \in (0,1)$ be such that $\alpha pr \sim t^1$. Since $\overline{p(\alpha pr)}$ and $\overline{rt^1}$ preserve bi-independence, we have

$$U(\lambda p(\alpha pr)) = U(\lambda qt^1)$$

for any $\lambda \in [0,1]$. Thus

$$U(\lambda qt_1) = U((\lambda + (1-\lambda)\alpha)pr)) = \lambda + (1-\lambda)\alpha.$$

Similarly, since $\overline{(\alpha pr)r}$ and $\overline{t_1 s}$ preserve bi-independence, we have

$$U(\lambda t_1 s) = U(\lambda(\alpha pr)r) = U((\lambda\alpha)pr) = \alpha\lambda$$

for any $\lambda \in [0,1]$.

Note that $U = f \circ V$ represents $\succsim$. Then U is a continuous function that is linear on both $\overline{qt^1}$ and $\overline{t^1 s}$. If U is linear on $\overline{qs}$, then $\overline{pr}$ and $\overline{qs}$ must preserve bi-independence, a contradiction. Hence, U restricting to $\overline{qs}$ cannot be differentiable at t^1. It is then easy to see that $\overline{pr}$ and $\overline{qs} \cap B_\varepsilon(t^1)$ cannot preserve bi-independence for any $\varepsilon > 0$.

Hence, we can assume, without loss of generality, that at least one of the following statements is true: $\overline{pr}$ and $\overline{qt^1}$ do not preserve bi-independence, or $\overline{pr}$ and $\overline{t^1 s}$ do not preserve bi-independence. If the first statement is correct, let $q^1 = q$ and $s^1 = t^1$; otherwise let $q^1 = t^1$ and $s^1 = s$. Then $\overline{pr}$ and $\overline{q^1 s^1}$ do not preserve bi-independence.

Inductively, let $t^{n+1} = \frac{1}{2}q^n s^n$. Without loss of generality, we can assume that at least one of the following statements is true: $\overline{pr}$ and $\overline{q^n t^{n+1}}$ do not preserve bi-independence, or $\overline{pr}$ and $\overline{t^{n+1} s^n}$ do not preserve bi-independence. If the first statement is correct, let $q^{n+1} = q^n$ and $s^{n+1} = t^{n+1}$; otherwise let $q^{n+1} = t^{n+1}$ and $s^{n+1} = s^n$. Then $\overline{pr}$ and $\overline{q^{n+1} s^{n+1}}$ do not preserve bi-independence.

By construction, for any n, $\overline{pr}$ and $\overline{q^n s^n}$ do not preserve bi-independence, and $d(q^n, s^n) = \frac{1}{2^n} d(q,s)$. The claim is established. ■

Lemma 8 *Let L be an open convex set that preserves independence and let $\overline{pr}$ preserve independence. If $\overline{pr}$ and L preserve bi-independence, then $\overline{pr}$ and cl(L) preserve bi-independence.*

Proof. There is nothing to prove if $L = \emptyset$. Let $L \neq \emptyset$. First, note that by Lemma 2, cl(L) preserves independence and there exists a linear U that represents $\succsim$ over cl(L). By way of contradiction suppose $\overline{pr}$ and cl(L) do not preserve bi-independence. Then, by Lemma 6 there exists $p', r' \in \overline{pr}$, $q, s' \in \text{cl}(L)$ and $\lambda \in (0,1)$ such that $p' \sim q \succ r' \sim s$ and $\lambda p' r' \not\sim \lambda qs$. Assume that $\lambda p' r' \succ \lambda qs$, since the other case is symmetric. Note that $\overline{p'r'}$ and $\overline{qs}$ each preserve independence. Thus, by standard continuity arguments, there exist $\hat{p}, \hat{r} \in \overline{p'r'}$ and $q', s' \in \overline{qs}$ such that $q \succ \hat{p} \sim q' \succ \hat{r} \sim s' \succ r$ and $\lambda \hat{p}\hat{r} \succ \lambda q' s'$. It is clear that $U(q'), U(s') \in U(L)$ by convexity of L. Thus, there exist $\hat{q}, \hat{s} \in L$ such that $\hat{q} \sim q' \sim \hat{p}$ and $\hat{s} \sim s' \sim \hat{r}$. Since cl$(L)$ preserves independence, we have $\lambda \hat{p}\hat{r} \succ \lambda q' s' \sim \lambda \hat{q}\hat{s}$. Thus, $\overline{pr}$ and L cannot preserve bi-independence, which is a contradiction. ■

Given p and q, we say that neighborhoods L_p and L_q *preserve weak bi-independence*, if L_p and L_q satisfy the condition in *Weak Local Bi-Independence.*

Lemma 9 *For any p, there exists $\varepsilon > 0$ such that for any $r \in B_\varepsilon(p)$ and any open convex set L that preserves independence, $\overline{pr}$ and L preserve bi-independence.*

Proof. By way of contradiction, suppose for any n there exist $r^n \in B_{\frac{1}{n}}(p)$ and an open convex set L_n that preserves independence such that $\overline{pr^n}$ and L_n do not preserve bi-independence.

By *Weak Local Independence* (implied by *Weak Local Bi-Independence*), for n large enough, $B_{\frac{1}{n}}(p) \perp p$, which implies that $\overline{pr^n}$ preserves independence by Lemma 5. Then by Lemmas 6 and 7, there exist $\hat{p}^n, \hat{r}^n \in \overline{pr^n}$ and $\hat{q}^n, \hat{s}^n \in L_n$ such that $\hat{p}^n \sim \hat{q}^n \succ \hat{r}^n \sim \hat{s}^n$, $d(\hat{q}^n, \hat{s}^n) < \frac{1}{n}$, and $\overline{\hat{p}^n\hat{r}^n}$ and $\overline{\hat{q}^n\hat{s}^n}$ do not preserve bi-independence.

It is clear that $\hat{p}^n, \hat{r}^n$ converges to p as n goes to infinity. Since $\mathcal{L}$ is compact, the sequence $\{\hat{q}^n\}$ has a subsequence that converges to q. We assume, without loss of generality, that the subsequence is $\{\hat{q}^n\}$ itself and that $d(\hat{q}^n, q)$ is monotonically decreasing in n. In addition, by *Continuity*, $V(\hat{q}^n) = V(\hat{p}^n)$ for all n implies that $V(q) = V(p)$.

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

There exists m such that $d(q, \hat{q}^m) < \delta - \frac{1}{m}$. Then it follows that $d(q, \hat{s}^m) < \delta - \frac{1}{m} + \frac{1}{m} = \delta$. Hence $\hat{q}^m, \hat{s}^m \in B_\delta(q)$. There also exists k such that $n \geqslant k$ implies $\hat{p}^n, \hat{r}^n \in B_\varepsilon(p)$. Let $N = \max\{m, k\}$. Then $\hat{p}^N, \hat{r}^N \in B_\varepsilon(p)$, $\hat{q}^N, \hat{s}^N \in B_\delta(q)$. Furthermore, $L_N \cap B_\delta(q)$ is nonempty and open, and preserves independence. By Lemma 4, we have that $\mathrm{co}((L_N \cap B_\delta(q)) \cup \{q\})$ preserves independence. Thus, $\overline{\hat{q}^N\hat{s}^N q}$ preserves independence.

By construction, since $\hat{p}^N, \hat{r}^N \in \overline{pr^N}$, and $\overline{pr^N}$ preserves independence, we either have $\hat{p}^N \succ \hat{r}^N \succsim p$ or $p \succsim \hat{p}^N \succ \hat{r}^N$. The two cases are symmetric so we will only prove the first case.

Let $\hat{p}^N \succ \hat{r}^N \succsim p$. It follows that $\hat{q}^N \succ \hat{s}^N \succsim q$. Then by *Continuity* there exists $s \in \overline{\hat{q}^N q}$ such that $s \sim \hat{s}^N$. Since $\overline{\hat{q}^N\hat{s}^N q}$ preserves independence, it is easy to see that $\overline{\hat{q}^N s}$ and $\overline{\hat{p}^N\hat{r}^N}$ do not preserve bi-independence, which implies that $\overline{\hat{q}^N q}$ and $\overline{\hat{p}^N p}$ do not preserve bi-independence. If $B_\varepsilon(p)$ and $B_\delta(q)$ preserve weak bi-independence, since $p \sim q$ and $\hat{p}^N \sim \hat{q}^N$, we have $\lambda\hat{p}^N p \sim \lambda\hat{q}^N q$ for all $\lambda \in [0, 1]$. Let

$$U(\lambda\hat{p}^N p) = U(\lambda\hat{q}^N q) = \lambda$$

for any $\lambda \in [0,1]$. Since $\overline{\hat{q}^N q}$ and $\overline{\hat{p}^N p}$ each preserves independence, and $\hat{p}^N \sim \hat{q}^N \succ p \sim q$, U represents $\succsim$ on $\overline{\hat{q}^N q} \cup \overline{\hat{p}^N p}$. It follows that $\overline{\hat{q}^N q}$ and $\overline{\hat{p}^N p}$ preserve bi-independence, which is a contradiction. ■

Thus, for each $p \in \mathcal{L}$, there exists $\varepsilon_p > 0$ such that $B_{\varepsilon_p}(p) \perp p$ and that for any $r \in B_{\varepsilon_p}(p)$ and any open convex set L that preserves independence, $\overline{pr}$ and L preserve bi-independence. In fact, by Lemma 5 and 8, $\overline{pr}$ and $\text{cl}(L)$ preserve bi-independence. Hereafter, for any p, identify $B_{\varepsilon_p}(p)$ with L_p.

Lemma 10 *For any $p \in \mathcal{L}$, there exist $p^1, \ldots, p^n \in \mathcal{L}$ such that $p \in \Delta := \overline{p^1 \ldots p^n}$, $int(\Delta) \neq \emptyset$ and Δ preserves independence. Furthermore, for any open convex set L that preserves independence, $\overline{p^i p^j}$ and $cl(L)$ preserve bi-independence for any i, j.*

Proof. Let $p^1 := p \in \mathcal{L}$. Then, recursively for $i = 1, \ldots, n-1$, let p^{i+1} be an arbitrary point in $(\bigcap_{j \leq i} L_{p^j}) \backslash \text{aff}(\{p^1, \ldots, p^i\})$. Since each L_{p^i} is open and $\text{aff}(\{p^1, \ldots, p^i\})$'s dimension is at most $i-1$, such p^{i+1}'s always exist. By construction, the dimension of $\Delta := \overline{p^1 \ldots p^n}$ is equal to $\dim(\mathcal{L})$, the dimension of $\mathcal{L}$, and Δ has nonempty interior.

Pick some $\alpha \in (0,1)$ such that for any $j = 1, \ldots, n-1$, $q^i := \alpha p^n p^i \in \bigcap_{i=1}^n L_{p^i}$. Clearly, Δ' also has nonempty interior. By construction, $\Delta' = \overline{q^1 \ldots q^{n-1} p^n}$ preserves independence with respect to p^i for $i = 1, \ldots, n$ because $\Delta' \subseteq \bigcap_{i=1}^n L_{p^i}$. Since $p^n \in \Delta'$, it follows from Lemma 3 that $\Delta' \perp q^i$ for $i = 1, \ldots, n$. Applying Lemma 3 again, we know that Δ' preserves independence with respect to every lottery in Δ', which implies that Δ' preserves independence. Then, applying Lemma 4 iteratively, we know that $\text{co}(\Delta' \cup \{p^{n-1}\})$ preserves independence, $\text{co}(\Delta' \cup \{p^{n-1}, p^{n-2}\})$ preserves independence, and so on. Since $\Delta = \text{co}(\Delta' \cup \{p^1, \ldots, p^{n-1}\})$, Δ preserves independence.

Furthermore, by construction we have $p^i \in L_{p^j}$ for any $i > j$. Thus, for any open convex set L that preserves independence, $\overline{p^i p^j}$ and $\text{cl}(L)$ preserve bi-independence. ■

Let $\mathcal{D}$ be the collection of all possible Δ's constructed using the procedure in Lemma 10.

Let $\mathcal{L}_o := \bigcup_{\Delta \in \mathcal{D}} \text{int}(\Delta)$. It is clear that

$$\mathcal{L} = \bigcup_{\Delta \in \mathcal{D}} \Delta = \bigcup_{\Delta \in \mathcal{D}} \text{cl}(\text{int}(\Delta)) \subseteq \text{cl}\left(\bigcup_{\Delta \in \mathcal{D}} \text{int}(\Delta)\right) = \text{cl}(\mathcal{L}_o) \subseteq \mathcal{L}.$$

Thus, $\mathcal{L}_o$ is an open dense subset of $\mathcal{L}$. For any $p \in \mathcal{L}_o$, pick $\Delta_p \in \mathcal{D}$ such that $p \in \text{int}(\Delta_p)$.

Lemma 11 *For any $\Delta, \Delta' \in \mathcal{D}$, Δ and Δ' preserve bi-independence.*

Proof. Since Δ and Δ' both preserve independence and have nonempty interior, Lemma 2 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, we only need to prove that for any $p, r \in \Delta$, $q, s \in \Delta'$, and $\lambda \in (0, 1)$ such that $\lambda pr \in \Delta$, $\lambda qs \in \Delta'$, and $p \succsim q$, we have $r \succsim s \Rightarrow \lambda pr \succsim \lambda qs$ and $r \succ s \Rightarrow \lambda pr \succ \lambda qs$. The case with $q \succsim p$ is similar.

Since $\Delta = \overline{p^1 \dots p^n}$ and $\Delta' := \overline{q^1 \dots q^n}$ for some $p^1, \dots, p^n \in \mathcal{L}$ and $q^1, \dots, q^n \in \mathcal{L}$, without loss of generality, let $U(p^1) = \min_i U(p^i)$, $U(p^n) = \max_i U(p^i)$, $U'(q^1) = \min_i U'(q^i)$, and $U'(q^n) = \max_i U'(q^i)$. Clearly, $U(\lambda pr) \in [U(p^1), U(p^n)]$ and $U'(\lambda qs) \in [U'(q^1), U'(q^n)]$. The cases with $p^1 \sim p^n$ or $q^1 \sim q^n$ are straightforward. Therefore, assume that $p^n \succ p^1$ and $q^n \succ q^1$. Without loss of generality, let $U(p^n) = U'(q^n) = 1$ and $U(p^1) = U'(q^1) = 0$. Standard arguments imply that there exist unique $\alpha, \alpha', \alpha'', \beta, \beta', \beta'' \in [0, 1]$ such that $\alpha p^1 p^n \sim p$, $\alpha' p^1 p^n \sim r$, $\alpha'' p^1 p^n \sim \lambda pr$, $\beta q^1 q^n \sim q$, $\beta' q^1 q^n \sim s$, and $\beta'' q^1 q^n \sim \lambda qs$. Then,

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

Similarly,

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

The procedure in Lemma 10 implies that either $p_1 \in L_{p_n}$ or $p_n \in L_{p_1}$, and either $q_1 \in L_{q_n}$ or $q_n \in L_{q_1}$. Hence, $\overline{p^1 p^n}$ and Δ' preserve bi-independence, which implies that $\overline{p^1 p^n}$ and $\overline{q^1 q^n}$ preserve bi-independence, Therefore, since $\alpha p^1 p^n \succsim \beta q^1 q^n$, we have $\alpha' p^1 p^n \succsim \beta' q^1 q^n \Rightarrow (\lambda\alpha + (1-\lambda)\alpha')p^1 p^n \succsim (\lambda\beta + (1-\lambda)\beta')q^1 q^n$ and $\alpha' p^1 p^n \succ \beta' q^1 q^n \Rightarrow (\lambda\alpha + (1-\lambda)\alpha')p^1 p^n \succ (\lambda\beta + (1-\lambda)\beta')q^1 q^n$, establishing the lemma. ■

Definition 4 *For any two subsets L_1, L_2 of $\mathcal{L}$, we write $L_1 \rightleftarrows L_2$ if there exist $p_h, p_l \in L_1$ and $q_h, q_l \in L_2$ such that both p_h and q_h are strictly preferred to both p_l and q_l.*

The proofs of Lemmas 12–15 are similar to the proofs of some lemmas in Appendix B of Ellis and Masatlioglu (2020). Therefore, we provide the proofs in the Online Appendix.

Lemma 12 *Suppose L_1 and L_2 are nonempty connected open subsets of L that preserve independence. If $L_1 \not\rightleftarrows L_2$, there exist affine functions $U_1 : L_1 \to \mathbb{R}$ and $U_2 : L_2 \to \mathbb{R}$ such that the function $U : L_1 \cup L_2 \to \mathbb{R}$ that satisfies $p \in L_i \Rightarrow U(p) = U_i(p)$, $i = 1, 2$, represents $\succsim$ on $L_1 \cup L_2$. Moreover, any positive affine transformation of U also represents $\succsim$ on $L_1 \cup L_2$.*

Proof. See the Online Appendix. ■

Lemma 13 *Suppose L_1 and L_2 are nonempty connected open subsets of $\mathcal{L}$ that preserve independence and bi-independence. If $L_1 \rightleftarrows L_2$, there exist affine functions $U_1 : L_1 \to \mathbb{R}$ and $U_2 : L_2 \to \mathbb{R}$ such that the function $U : L_1 \cup L_2 \to \mathbb{R}$ that satisfies $p \in L_i \Rightarrow U(p) = U_i(p)$, $i = 1, 2$, represents $\succsim$ on $L_1 \cup L_2$. Moreover, any positive affine transformation of U also represents $\succsim$ on $L_1 \cup L_2$.*

Proof. See the Online Appendix. ■

Lemma 14 *Suppose L_1 and L_2 are nonempty connected open subsets of $\mathcal{L}$ that preserve independence and bi-independence, and $L_1 \cap L_2 \neq \emptyset$. The following statements are true:*

(i) There exists $u : Z \to \mathbb{R}$ such that $U(p) = \sum_{i=1}^{n} p_i u(z_i)$ for any $p \in L_1 \cup L_2$ represents $\succsim$ on $L_1 \cup L_2$.

(ii) For any affine functions $U_1 : L_1 \to \mathbb{R}$ and $U_2 : L_2 \to \mathbb{R}$ such that the function $U : L_1 \cup L_2 \to \mathbb{R}$ that satisfies $p \in L_i \Rightarrow U(p) = U_i(p)$, $i = 1, 2$, represents $\succsim$ on $L_1 \cup L_2$, there exists some $u : Z \to \mathbb{R}$ such that $U(p) = \sum_{i=1}^n p_i u(z_i)$ for any $p \in L_1 \cup L_2$.

Proof. See the Online Appendix. ■

For a finite sequence of subsets $L_1, \ldots, L_m$ of $\mathcal{L}$, we write $L_1 \rightleftarrows \cdots \rightleftarrows L_m$ if for any $i \in \{1, \ldots, m-1\}$, there exist $p_h, p_l \in L_i$ and $q_h, q_l \in L_{i+1}$ such that both p_h and q_h are strictly preferred to both p_l and q_l. 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 $L_j \cup L_{j+1}$, $j = 1, \ldots, m-1$. Note that $\rightleftarrows$ is not a transitive binary relation; that is, $L_1 \rightleftarrows L_2 \rightleftarrows L_3$ does not imply $L_1 \rightleftarrows L_3$.

Lemma 15 *Suppose $L_1, \ldots, L_m$ are nonempty connected open subsets of $\mathcal{L}$ 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 $\bigcup_{i=1}^m L_i$.*

Proof. See the Online Appendix. ■

Let $\mathcal{O} := \{L \subseteq \mathcal{L} : L \text{ is nonempty, connected, and open}\}$. Let $\mathcal{P}$ be the set of all functions $P : \mathcal{L}_o \to \mathcal{O}$ such that for any $p, q \in \mathcal{L}_o$, (i) $\text{int}(\Delta_p) \subseteq P(p)$, (ii) $P(p)$ and $\text{int}(\Delta)$ preserve bi-independence for all $\Delta \in \mathcal{D}$, and (iii) $P(p)$ and $P(q)$ preserve bi-independence. Let $P_0 : \mathcal{L}_o \to \mathcal{O}$ be a function such that $P_0(p) = \text{int}(\Delta_p)$ for any $p \in \mathcal{L}_o$. Lemma 11 implies that $P_0 \in \mathcal{P}$; that is, $\mathcal{P}$ is nonempty.

Define a binary relation $\Subset$ on $\mathcal{P}$ as follows: For any $P, Q \in \mathcal{P}$, $P \Subset Q$ if for any $p \in \mathcal{L}_o$, $P(p) \subseteq Q(p)$. It is straightforward to verify that $\Subset$ is a partial order on $\mathcal{P}$. Take any totally ordered subset of $\mathcal{P}$, $\{P_i\}_{i \in I}$, in which I is an index set. Let $P^* : \mathcal{L}_o \to \mathcal{O}$ be a function such that for any $p \in \mathcal{L}_o$, $P^*(p) := \bigcup_{i \in I} P_i(p)$. It must be true that $P^* \in \mathcal{P}$. First of all, $P^*(p)$ is open since every $P_i(p)$ is open. Second, $P^*(p)$ is connected, since every $P_i(p)$ is connected and contains $\text{int}(\Delta_p)$, which is connected. Now we show that $P^*(p)$ and $P^*(q)$ preserve bi-independence for all $p, q \in \mathcal{L}_o$. To see this, for any $\lambda \in (0,1)$, if $p', r', \lambda p'r' \in P^*(p)$ and $q', s', \lambda q's' \in P^*(q)$, by $\{P_i\}_{i \in I}$ is totally ordered by $\Subset$, there must exist some index $j \in I$ such that $p', r', \lambda p'r' \in P_j(p)$ and $q', s', \lambda q's' \in P_j(q)$. Then, the property that we want $p', r', \lambda p'r', q', s', \lambda q's'$ to satisfy to ensure that $P^*(p)$ and $P^*(q)$ preserve bi-independence follows from the fact that $P_j(p)$ and $P_j(q)$ preserve bi-independence. Similarly, $P^*(p)$ and $\text{int}(\Delta)$ preserve bi-independence for any $\Delta \in \mathcal{D}$. 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 $\mathcal{P}$ contains some $\Subset$-maximal element. With a harmless abuse of notation, denote this $\Subset$-maximal element by P^*. Because $\mathcal{L}_o$ is an open dense subset of $\mathcal{L}$, $\bigcup_{p \in \mathcal{L}_o} P^*(p)$ is also an open dense subset of $\mathcal{L}$. The next step is to prove that P^* has some nice properties. We will need the following lemma.

Lemma 16 *Let L_1, L_2, L_3 be nonempty, connected, open subsets of $\mathcal{L}$ that preserve independence. If L_i and L_3 preserve bi-independence for $i = 1, 2$ and $L_1 \cap L_2 \neq \emptyset$, then $L_1 \cup L_2$ and L_3 preserve bi-independence.*

Proof. First note that by Lemma 14, $L_1 \cup L_2$ preserves independence. Now we show that $L_1 \cup L_2$ and L_3 preserve bi-independence. Suppose $L_3 \not\rightleftarrows L_1 \cup L_2$. Then, Lemma 12 implies that $L_1 \cup L_2$ and L_3 preserve bi-independence.

Now suppose $L_3 \rightleftarrows L_1 \cup L_2$. It follows that $L_1 \rightleftarrows L_3$ or $L_2 \rightleftarrows L_3$. Moreover, the fact that $L_1 \cap L_2 \neq \emptyset$ implies $L_1 \rightleftarrows L_2$. Hence, we can apply Lemma 15 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 $L_1 \cup L_2 \cup L_3$. In particular, Lemma 14 implies that $\hat{U} : L_1 \cup L_2 \to \mathbb{R}$ that agrees with U_1 on L_1 and with U_2 on L_2 must be affine. Thus, U is affine on $L_1 \cup L_2$ and L_3, respectively. Then, we must have $L_1 \cup L_2$ and L_3 preserve bi-independence. ■

Lemma 17 *For any $p, q \in \mathcal{L}_o$, if $P^*(p) \cap P^*(q) \neq \emptyset$, then $P^*(p) = P^*(q)$.*

Proof. Suppose $P^*(p) \cap P^*(q) \neq \emptyset$. First note that by Lemma 14, $P^*(p) \cup P^*(q)$ preserves independence. By Lemma 16, $P^*(p) \cup P^*(q)$ and $P^*(r)$ preserve bi-independence for all $r \in \mathcal{L}_o$. Again by Lemma 16, $P^*(p) \cup P^*(q)$ and $\text{int}(\Delta)$ preserve bi-independence for any $\Delta \in \mathcal{D}$. Thus, if $P^*(p) \cap P^*(q) \neq \emptyset$, then P^* is not $\Subset$-maximal unless $P^*(p) = P^*(q)$. To see this, if $P^*(p) \neq P^*(q)$, we can define a new function $\hat{P} : \mathcal{L}_o \to \mathcal{O}$ that agrees with P^* except at p and q. Let $\hat{P}(p) = \hat{P}(q) = P^*(p) \cup P^*(q)$. Then, we have $\hat{P} \neq P^*$, $\hat{P} \in \mathcal{P}$, and $P^* \Subset \hat{P}$. ■

Next, we show that $\{P^*(p) : p \in \mathcal{L}_o\}$ is finite. For any $p \in \mathcal{L}$ and $\varepsilon > 0$, let $C_\varepsilon(p) := \{q \in \mathcal{L} : ||q - p||_\infty < \varepsilon\}$; that is, $C_\varepsilon(p)$ is the open hypercube that is centered at p with edge length 2ε.

Lemma 18 *$\{P^*(p) : p \in \mathcal{L}_o\}$ is finite.*

Proof. Suppose $\{P^*(p) : p \in \mathcal{L}_o\}$ is not finite. Let $\{P_i\}_{i \in I} := \{P^*(p) : p \in \mathcal{L}_o\}$ for some index set I. Then by the compactness of $\mathcal{L}$, there exists $p \in \mathcal{L}$ such that any neighborhood of p intersects P_i for an infinite number of i's in I. Fix some $\varepsilon > 0$ such that $C_\varepsilon(p) \subseteq L_p$. Let $J \subseteq I$ be such that $C_\varepsilon(p) \cap P_i \neq \emptyset$ if and only if $i \in J$.

First, we show that for any $\Delta \in \mathcal{D}$, there exists $i \in I$ such that $\text{int}(\Delta) \subseteq P_i$. By the denseness of $\mathcal{L}_o$, there exists $i \in I$ such that $P_i \cap \text{int}(\Delta) \neq \emptyset$. Then by Lemma 16 and the maximality of P_i, it must be the case that $\text{int}(\Delta) \subseteq P_i$.

For any $q \in C_\varepsilon(p)$, construct Δ as in Lemma 10 such that $\overline{pq} \subseteq \Delta \in \mathcal{D}$. It follows that $\text{int}(\Delta) \subseteq P_i$ for some $i \in I$, and hence, $\overline{pq} \subseteq \text{cl}(P_i)$ for some $i \in I$. Moreover, if $\overline{pq} \cap P_i \neq \emptyset$ for some $i \in I$, then $\overline{pq} \subseteq \text{cl}(P_i)$ since $P_i \cap \text{cl}(P_j) = \emptyset$ for any $i \neq j$.

The next step is to show that $C_\varepsilon(p) \cap \text{cl}(P_i)$ is the intersection of $C_\varepsilon(p)$ and a cone with vertex p for all $i \in J$. For $\tilde{p} \in \mathcal{L}$, $\tilde{\varepsilon} > 0$, and $L \subseteq \mathcal{L}$ such that $C_{\tilde{\varepsilon}}(\tilde{p}) \cap L \neq \emptyset$, let

$$\text{cone}_{\tilde{\varepsilon}}(\tilde{p}, L) := \{\tilde{q} \in C_{\tilde{\varepsilon}}(\tilde{p}) : \tilde{q} = \tilde{p} + \alpha(\tilde{r} - \tilde{p}) \text{ for some } \alpha \geqslant 0 \text{ and } \tilde{r} \in C_{\tilde{\varepsilon}}(\tilde{p}) \cap L\};$$

that is, $\mathrm{cone}_{\tilde{\varepsilon}}(\tilde{p}, L)$ is the intersection of $C_{\tilde{\varepsilon}}(\tilde{p})$ and the smallest cone with vertex $\tilde{p}$ that contains $C_{\tilde{\varepsilon}}(\tilde{p}) \cap L$. On one hand, we have $\mathrm{cone}_{\varepsilon}(p, P_i) \subseteq \mathrm{cl}(P_i)$, since $q \in C_{\varepsilon}(p) \cap P_i$ implies $\overline{pq} \subseteq \mathrm{cl}(P_i)$. Hence, $\mathrm{cone}_{\varepsilon}(p, \mathrm{cl}(P_i)) \subseteq \mathrm{cl}(\mathrm{cone}_{\varepsilon}(p, P_i)) \subseteq \mathrm{cl}(P_i)$. On the other hand, by definition, $C_{\varepsilon}(p) \cap \mathrm{cl}(P_i) \subseteq \mathrm{cone}_{\varepsilon}(p, \mathrm{cl}(P_i))$. Thus, $C_{\varepsilon}(p) \cap \mathrm{cl}(P_i) = \mathrm{cone}_{\varepsilon}(p, \mathrm{cl}(P_i))$.

Now, let $P_i^o = \mathrm{int}(\mathrm{cl}(P_i))$ for each i. It is clear that $P_i \subseteq P_i^o$. Moreover, $P_i^o \cap P_j^o = \emptyset$ for all $i \neq j$. Suppose not. Then there exist $r \in \mathcal{L}$ 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 $P_i \cap P_j = \emptyset$.

Let $\varepsilon_1 := \frac{\varepsilon}{2}$ and $C_1 := \partial C_{\varepsilon_1}(p)$; that is, C_1 is the surface of hypercube $C_{\varepsilon_1}(p)$. Clearly, $C_1 \subseteq C_\varepsilon(p)$. Now we show that $C_1 \cap P_i^o \neq \emptyset$ for all $i \in J$. For any $i \in J$, there exists $q \in C_\varepsilon(p) \cap P_i$. Therefore, $\overline{pq} \subseteq C_\varepsilon(p) \cap \mathrm{cl}(P_i) = \mathrm{cone}_\varepsilon(p, \mathrm{cl}(P_i))$. Moreover, since $q \in C_\varepsilon(p) \cap P_i^o$,

$$\mathrm{cone}_\varepsilon(p, \{q\}) \cap C_1 \subseteq \mathrm{int}(\mathrm{cone}_\varepsilon(p, \mathrm{cl}(P_i))) = \mathrm{int}(C_\varepsilon(p) \cap \mathrm{cl}(P_i)) = C_\varepsilon(p) \cap P_i^o.$$

Hence, $C_1 \cap P_i^o \neq \emptyset$ for all $i \in J$.

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 $p^1 \in A_1$ such that if L is a neighborhood of p^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 $C_{\varepsilon'}(p^1) \subseteq L_{p^1}$. Let $J_1 \subseteq J$ be such that $(C_{\varepsilon'}(p^1) \cap A_1) \cap P_i^o \neq \emptyset$ if and only if $i \in J_1$. Let $\varepsilon_2 := \frac{\varepsilon'}{2}$ and $C_2 := \partial C_{\varepsilon_2}(p^1) \subseteq C_{\varepsilon'}(p^1)$. Following the same logic as above, $(C_2 \cap A_1) \cap P_i^o \neq \emptyset$ for all $i \in J_1$. 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 $p^2 \in A_1 \cap A_2$ such that if L is a neighborhood of p^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 $p^k \in A_1 \cap \cdots \cap A_k$ such that if L is a neighborhood of p^k, then $L \cap A_1 \cap \cdots \cap A_k$ intersects P_i^o for an infinite number of i's in I. This is a contradiction, since $A_1 \cap \cdots \cap A_{n-1}$ is a singleton and $P_i^o \cap P_j^o = \emptyset$ for all $i \neq j$. $\blacksquare$

For any $p, q \in \mathcal{L}_o$, we write $p \leftrightsquigarrow q$ if $p \in P^*(q)$ or there is a finite sequence of lotteries $r^1, \ldots, r^m \in \mathcal{L}_o$ such that $p \in P^*(r^1)$, $q \in P^*(r^m)$, and $P^*(r^1) \rightleftarrows \cdots \rightleftarrows P^*(r^m)$. By definition, $\leftrightsquigarrow$ is reflexive and transitive, and hence an equivalence relation. For any $p \in \mathcal{L}_o$, let $Q(p)$ denote the equivalence class of p induced by $\leftrightsquigarrow$. Clearly $P^*(p) \subseteq Q(p)$ for any $p \in \mathcal{L}_o$.

Lemma 19 *For any $p \in \mathcal{L}_o$, $Q(p) = \bigcup_{i=1}^m P^*(q^i)$ for some $m \geqslant 1$ and $q^1, \ldots, q^m \in \mathcal{L}_o$. Moreover, there exists $i_1, i_2, \ldots, i_k$ such that $P^*(q^{i_1}) \rightleftarrows \cdots \rightleftarrows P^*(q^{i_k})$ and for any $1 \leqslant i \leqslant m$, $i_j = i$ for some $1 \leqslant j \leqslant k$.*

Proof. Take an arbitrary $p \in \mathcal{L}_o$. Define $Q_1(p) := \{P^*(p)\}$, and then recursively, $Q_i(p) := \{P^*(q) : q \in \mathcal{L}_o \text{ and } P^*(q) \rightleftarrows P \text{ for some } P \in Q_{i-1}(p)\}$. By construction, $Q_i(p) \subseteq Q_{i+1}(p)$ for all i. For each i, let $Q^i(p) := \bigcup_{P \in Q_i(p)} P$. By Lemma 18, we know that $\bigcup_{i=1}^\infty Q^i(p) = \bigcup_{i=1}^m P^*(q^i)$ for some $q^1, \ldots, q^m \in \mathcal{L}_o$. By Lemma 17, since $P^*(p) \subseteq \bigcup_{i=1}^m P^*(q^i)$, we can without loss of generality assume that $q_1 = p$. By construction, for any $j \in \{1, \ldots, m\}$, there exists $i_1, i_2, \ldots, i_{k_j} \in \{1, \ldots, m\}$ such that $P^*(p) \rightleftarrows P^*(q^{i_1}) \rightleftarrows \cdots \rightleftarrows P^*(q^{i_{k_j}}) \rightleftarrows P^*(q^j)$. Hence, according to the definition of $Q(p)$,

$$Q(p) \subseteq \bigcup_{i=1}^\infty Q^i(p) = \bigcup_{i=1}^m P^*(q^i) \subseteq Q(p). \tag{4}$$

Hence $Q(p) = \bigcup_{i=1}^m P^*(q^i)$ and the rest is straightforward. ■

It follows from the previous lemma that $V(Q(p))$ is a (possibly degenerate) interval for any $p \in \mathcal{L}_o$. Lemmas 18 and 19 immediately imply the following.

Corollary 2 *$\{Q(p) : p \in \mathcal{L}_o\}$ is finite.*

Lemma 20 *For any $p, q \in \mathcal{L}_o$ such that $p \not\leftrightsquigarrow q$, $V(Q(p)) \cap V(Q(q))$ has empty interior.*

Proof. By Lemma 19, $V(Q(p)) = \bigcup_{i=1}^m V(P^*(p^i))$ and $V(Q(q)) = \bigcup_{j=1}^{m'} V(P^*(q^j))$, where each p^i and q^j are in $\mathcal{L}_o$. By way of contradiction, assume that $V(Q(p)) \cap V(Q(q))$ has nonempty interior. It follows that $V(Q(p))$ and $V(Q(q))$ cannot be degenerate. Then, $V(P^*(p^i))$ and $V(P^*(q^j))$ must have nonempty interior for all i, j. Moreover, since m, m' are finite, there exist $k \in \{1, \dots, m\}$ and $k' \in \{1, \dots, m'\}$ such that $V(P^*(p^k)) \cap V(P^*(q^{k'}))$ has nonempty interior. Hence, by definition, $P^*(p^k) \rightleftarrows P^*(q^{k'})$. This, together with Lemma 19, implies $p \leftrightsquigarrow q$, a contradiction. ■

Since $\mathcal{L}$ is compact and connected and V is continuous, we know that $V(\mathcal{L})$ is a closed and bounded interval. Because $\mathcal{L}_o$ is dense in $\mathcal{L}$, for any $p \in \mathcal{L}$ there exists a sequence $\{p^k\} \subseteq \mathcal{L}_o$ that converges to p. Given that V is continuous, $V(p^k)$ also converges to $V(p)$. Hence, $V(\mathcal{L}) \subseteq \text{cl}(V(\mathcal{L}_o)) \subseteq \text{cl}(V(\mathcal{L})) = V(\mathcal{L})$, which implies $\text{cl}(V(\mathcal{L}_o)) = V(\mathcal{L})$.

Due to Corollary 2 and Lemma 20, there exists $m \geqslant 1$ and $p^1, \dots, p^m$ such that $V(\mathcal{L}_o) = \bigcup_{i=1}^m V(Q(p^i))$. Note that by Lemma 20, we can without loss of generality assume that $V(Q(p^i)) \cap V(Q(p^j))$ has empty interior whenever $i \neq j$. Since each $V(Q(p^i))$ is connected, it must be a (possibly degenerate) interval. For each i, let $V_i^h := \sup V(Q(p^i))$ and $V_i^l := \inf V(Q(p^i))$. If $V(\mathcal{L}_o)$ is a singleton, then the theorem is trivially true. Without loss of generality, let $V(Q(p^i))$ be nondegenerate if and only if $i \in \{1, \dots, k\}$, and assume $V_i^h \leqslant V_{i+1}^l$ for each $i \in \{1, \dots, k-1\}$.

Consider V_1 first. By Lemma 15 and Lemma 19, there exist a piecewise-affine function $U_1 : Q(p^1) \to \mathbb{R}$ that represents $\succsim$ on $Q(p^1)$. We can perform positive affine transformations to U_1 such that $\inf U^1 = V_1^l$ and $\sup U^1 = V_1^h$. Since V also represents $\succsim$ on $Q(p^1)$, there exists a strictly increasing function $f_1 : V(Q(p^1)) \to \mathbb{R}$ such that $f_1(V(p)) = U_1(p)$ for any $p \in Q(p^1)$. Extend f_1's domain to $V(\mathcal{L})$ by letting $f_1(v) = v$ for any $v \in V(\mathcal{L}) \backslash V(Q(p^1))$.

Lemma 21 *The function f_1 is strictly increasing and continuous.*

Proof. See the Online Appendix. ■

Thus, $f_1 \circ V$ is continuous on $\mathcal{L}$ and piecewise-affine on $Q(p^1)$. Recursively, for each $2 \leqslant i \leqslant k$, repeat the exercise above to construct continuous and strictly increasing function

$f_i : V(\mathcal{L}) \to \mathbb{R}$ such that $f_i \circ f_{i-1} \circ \cdots f_1 \circ V$ represents $\succsim$, and is continuous and piecewise-affine on $Q(p^i)$. In the end, we have $U = f_k \circ \cdots \circ f_1 \circ V$ represents $\succsim$, and is continuous and piecewise-affine on each $Q(p^i)$ for $i \in \{1, 2, \ldots, k\}$. Since each $V(Q(p^i))$ is a constant for $i > k$, U is piecewise-affine on $\mathcal{L}_o$. By continuity, it is clear that U is affine on $\mathrm{cl}(P^*(p))$ for any $p \in \mathcal{L}_o$, and thus U is a continuous piecewise-linear representation of $\succsim$, which concludes the proof of the sufficiency of the axioms.

See the Online Appendix for the necessity of the axioms. ■

Online Appendix

Necessity of the Axioms in Theorem 1

Now we show that the axioms are necessary. We say that $P \subseteq \mathbb{R}^n$ is a (closed convex) polytope if it is the bounded intersection of finitely many closed half-spaces in $\mathbb{R}^n$. We say that a finite collection of polytopes $P_1, \dots, P_k$ is a partition of $L \subseteq \mathbb{R}^n$ if $\bigcup_{i=1}^k P_i = L$ and $\mathrm{int}(P_i \cap P_j) = \emptyset$ for any $i \neq j$.

Suppose $\succsim$ has a NEU representation. The fact that *Weak Order* and *Continuity* hold is clear. Now we show that $\succsim$ satisfies *Weak Local Bi-Independence*. Let the NEU representation be

$$U(p) = \tau_{h+1} \circ \theta_h \circ \tau_h \circ \cdots \circ \theta_2 \circ \tau_2 \circ \theta_1 \circ \tau_1(p)$$

as in Definition 1, in which $\theta_i^{(j)} = \max\{\cdot, 0\}$ is the j-th component of θ_i. By setting

$$\tau_i^{(j)} \circ \theta_{i-1} \circ \cdots \circ \tau_1(p) = 0$$

for each i, j and taking the affine hull of each linear component of the solution, we obtain a finite set of affine hyperplanes. We denote the collection of these affine hyperplanes as $\mathcal{A}$. Thus, $\mathcal{A}$ is an arrangement of hyperplanes in $\mathcal{L}$. A region of $\mathcal{A}$ is a connected component of $\mathcal{L} \backslash (\bigcup_{H \in \mathcal{A}} H)$. Let $\mathcal{R}(\mathcal{A})$ be the collection of regions of $\mathcal{A}$. For each $L \in \mathcal{R}(\mathcal{A})$, it is easy to see that L is nonempty, open, and $\mathrm{cl}(L)$ is a polytope. Let $\mathcal{P}(\mathcal{A}) := \{\mathrm{cl}(L) : L \in \mathcal{R}(\mathcal{A})\}$. Since $\mathcal{A}$ is finite, $\mathcal{P}(\mathcal{A})$ must be finite. Hence, $\mathcal{P}(\mathcal{A})$ is a partition of $\mathcal{L}$. Moreover, within each $P \in \mathcal{P}(\mathcal{A})$, the NEU representation U coincides with an affine function.

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

Next, we show that

$$L_p = \mathrm{int}\left(\bigcup\{P \in \mathcal{P}(\mathcal{A}) : p \in P\}\right).$$

First note that since L_p is convex and open, $\mathrm{int}(\mathrm{cl}(L_p)) = L_p$. The claim is trivially true if $\mathcal{A}(p) = \emptyset$. If $\mathcal{A}(p) \neq \emptyset$, then by construction $\mathcal{A}(p)$ is an arrangement of hyperplanes on $\mathcal{L}$. Moreover, $p \in \bigcap_{H \in \mathcal{A}(p)} H$. It follows that p is in any closed half-spaces given by $\mathcal{A}(p)$. Thus, $p \in P'$ for any $P' \in \mathcal{P}(\mathcal{A}(p))$. Since $p \in L_p$, we have that $p \in P' \cap L_p$ for any $P' \in \mathcal{P}(\mathcal{A}(p))$. It is clear that

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

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

$$\begin{aligned} \mathrm{cl}(L_p) &= \mathrm{cl}\left(\bigcup\{L' \cap L_p : L' \in \mathcal{R}(\mathcal{A}(p))\}\right) \\ &= \mathrm{cl}\left(\bigcup\{L \in \mathcal{R}(\mathcal{A}) : L \subseteq L_p\}\right) \\ &= \bigcup\{P \in \mathcal{P}(\mathcal{A}) : P \subseteq \mathrm{cl}(L_p)\} \\ &= \bigcup\{P \in \mathcal{P}(\mathcal{A}) : p \in P\} \end{aligned}$$

and we are done with this step.

The last step is to show that this L_p construction is exactly want we want for *Weak Local Bi-Independence*. Given $p, q \in \mathcal{L}$, by the convexity of each $P \in \mathcal{P}(\mathcal{A})$, it is clear that for any $r \in L_p$ and $s \in L_q$, $\overline{pr} \subseteq P$ and $\overline{qs} \subseteq P'$ for some $P, P' \in \mathcal{P}(\mathcal{A})$. Since U coincides with an affine function within P and P', we conclude that $\overline{pr}$ and $\overline{qs}$ preserve bi-independence.

Omitted Proofs

Lemma 7 *Suppose L_1 and L_2 are nonempty connected open subsets of L that preserve independence. If $L_1 \not\rightleftarrows L_2$, there exist affine functions $U_1 : L_1 \to \mathbb{R}$ and $U_2 : L_2 \to \mathbb{R}$ such that the function $U : L_1 \cup L_2 \to \mathbb{R}$ that satisfies $p \in L_i \Rightarrow U(p) = U_i(p)$, $i = 1, 2$, represents $\succsim$ on $L_1 \cup L_2$. Moreover, any positive affine transformation of U also represents $\succsim$ on $L_1 \cup L_2$.*

Proof. Since L_i preserves independence, by Lemma 2, we can find some affine function

$U_i : L_i \to \mathbb{R}$ that represents $\succsim$ on L_i, $i = 1, 2$, respectively. If all lotteries in L_1 are indifferent, $L_1 \not\rightleftarrows L_2$. In this case, let $U_1 = 0$. By transitivity of $\succsim$, we can always apply some positive affine transformation to normalize U_2 such that $U_2(q) \geqslant 0$ if and only if $q \succsim p$ for any $p \in L_1$ and $q \in L_2$. The case in which all lotteries in L_2 are indifferent is symmetric.

Next, suppose that not all lotteries in L_i are indifferent for $i = 1, 2$. Since $\succsim$ has a continuous representation, L_1 and L_2 are connected, and $L_1 \not\rightleftarrows L_2$, there are only two cases left: the case in which any lottery in L_1 is weakly preferred to any lottery in L_2, and the opposite case. Since these two cases are similar, here we only examine the first. Since L_1 and L_2 are open, it follows from $L_1 \not\rightleftarrows L_2$ that $V(L_1) \cap V(L_2) = \emptyset$. Because U_2 is unique up to a positive affine transformation, we can fix U_1 and without loss of generality assume that U_2 satisfies $\sup_{q \in L_2} U_2(q) < \inf_{p \in L_1} U_1(p)$. Then, the function $U : L_1 \cup L_2 \to \mathbb{R}$ that satisfies $p \in L_i \Rightarrow U(p) = U_i(p)$, $i = 1, 2$, must represent $\succsim$ on $L_1 \cup L_2$. The last statement of the lemma is immediate. ■

Lemma 8 *Suppose L_1 and L_2 are nonempty connected open subsets of $\mathcal{L}$ that preserve independence and bi-independence. If $L_1 \rightleftarrows L_2$, there exist affine functions $U_1 : L_1 \to \mathbb{R}$ and $U_2 : L_2 \to \mathbb{R}$ such that the function $U : L_1 \cup L_2 \to \mathbb{R}$ that satisfies $p \in L_i \Rightarrow U(p) = U_i(p)$, $i = 1, 2$, represents $\succsim$ on $L_1 \cup L_2$. Moreover, any positive affine transformation of U also represents $\succsim$ on $L_1 \cup L_2$.*

Proof. Since L_i preserves independence, by Lemma 2, we can find some affine function $U_i : L_i \to \mathbb{R}$ that represents $\succsim$ on L_i, $i = 1, 2$, respectively. Suppose for some $p_h, p_l \in L_1$ and $q_h, q_l \in L_2$, both p_h and q_h are strictly preferred to both p_l and q_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 $p \in L_1$ and $q \in L_2$ such that $p \sim q$. Since L_1 is open and there exists $p_h, p_l \in L_1$ such that $p_h \succ p_l$, one can always find p^*, p_* in a small ε-ball centered at p such that $\overline{p^* p_*} \subseteq L_1$ and $p^* \succ p \sim q \succ p_*$. Without loss of generality, let $U_1(p) = 0$. Since L_1 and L_2 are open, by *Continuity*, there exists some $\alpha \in (0, 1)$ such that $r \in L_1$ implies $\alpha rp \in L_1$ and $p^* \succ \alpha rp \succ p_*$, and $s \in L_2$ implies $\alpha sq \in L_2$ and $p^* \succ \alpha sq \succ p_*$. Then, standard arguments imply that for each $s \in L_2$, there exists a unique $\lambda_s \in (0,1)$ such that $\alpha sq \sim \lambda_s p^* p_*$. Define for each $s \in L_2$

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

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

$$\alpha(\lambda ss')q = \lambda(\alpha sq)(\alpha s'q) \sim \lambda(\lambda_s p^* p_*)(\lambda_{s'} p^* p_*),$$

which implies that

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

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 $p \sim q$,

$$U_2(q) = \frac{1}{\alpha} U_1(\lambda_q p^* p_*) = \frac{1}{\alpha} U_1(p) = 0 = U_1(p).$$

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

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

and

$$U_2(q') = \frac{1}{\alpha} U_1(\lambda_{q'} p^* p_*)$$

where $\lambda_{q'} p^* p_* \sim \alpha q' q$. Then,

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

These observations also imply that if $p \in L_1 \cap L_2$, $U_1(p) = U_2(p)$. Then, we can define a function $U : L_1 \cup L_2 \to \mathbb{R}$ such that $p \in L_i \Rightarrow U(p) = U_i(p)$, $i = 1, 2$. The above arguments show that U represents $\succsim$ on $L_1 \cup L_2$. The last statement of the lemma is immediate. ■

Lemma 9 *Suppose L_1 and L_2 are nonempty connected open subsets of $\mathcal{L}$ that preserve independence and bi-independence, and $L_1 \cap L_2 \neq \emptyset$. The following statements are true:*

(i) There exists $u : Z \to \mathbb{R}$ such that $U(p) = \sum_{i=1}^n p_i u(z_i)$ for any $p \in L_1 \cup L_2$ represents $\succsim$ on $L_1 \cup L_2$.

(ii) For any affine functions $U_1 : L_1 \to \mathbb{R}$ and $U_2 : L_2 \to \mathbb{R}$ such that the function $U : L_1 \cup L_2 \to \mathbb{R}$ that satisfies $p \in L_i \Rightarrow U(p) = U_i(p)$, $i = 1, 2$, represents $\succsim$ on $L_1 \cup L_2$, there exists some $u : Z \to \mathbb{R}$ such that $U(p) = \sum_{i=1}^n p_i u(z_i)$ for any $p \in L_1 \cup L_2$.

Proof. Since L_1 and L_2 are nonempty, open, and connected, and preserve independence and bi-independence, by Lemmas 7 and 8, there exist affine functions $U_1 : L_1 \to \mathbb{R}$ and $U_2 : L_2 \to \mathbb{R}$ such that the function $U : L_1 \cup L_2 \to \mathbb{R}$ that satisfies $p \in L_i \Rightarrow U(p) = U_i(p)$, $i = 1, 2$, represents $\succsim$ on $L_1 \cup L_2$.

Take any affine U_1 and U_2 such that the function $U : L_1 \cup L_2 \to \mathbb{R}$ that satisfies $p \in L_i \Rightarrow U(p) = U_i(p)$, $i = 1, 2$, represents $\succsim$ on $L_1 \cup L_2$. Pick $p \in L_1 \cap L_2$. We can find some $\varepsilon > 0$ and $\alpha \in (0, 1)$ such that for any $q \in \mathcal{L}$, $\alpha q p \in B_\varepsilon(p) \subseteq L_1 \cap L_2$. Without loss of generality, let $U(p) = U_1(p) = U_2(p) = 0$. Define $u(z_j) = \frac{1}{\alpha} U(\alpha \delta_{z_j} p)$ for each $j \in \{1, \ldots, n\}$. We now verify that $U_i(q) = \sum_{j=1}^n q_j u(z_j)$ for any $q \in L_i$, $i = 1, 2$. To see this, take any $q \in L_i$. We have $\alpha p q \in B_\varepsilon(p)$ and $\alpha q p = q_1(\alpha \delta_{z_1} p) + q_2(\alpha \delta_{z_2} p) + \cdots + q_n(\alpha \delta_{z_n} p)$. Since $U = U_1 = U_2$ on

$L_1 \cap L_2$,

$$\alpha U_i(q) = U_i(\alpha q p) = \sum_{j=1}^{n} q_j U_i(\alpha \delta_{z_j} p) = \sum_{j=1}^{n} q_j \alpha u(z_j).$$

Therefore, $U(q) = \sum_{j=1}^{n} q_j u(z_i)$ for any $q \in L_1 \cup L_2$ represents $\succsim$ on $L_1 \cup L_2$. ■

For a finite sequence of subsets $L_1, \dots, L_m$ of $\mathcal{L}$, we write $L_1 \rightleftarrows \cdots \rightleftarrows L_m$ if for any $i \in \{1, \dots, m-1\}$, there exist $p_h, p_l \in L_i$ and $q_h, q_l \in L_{i+1}$ such that both p_h and q_h are strictly preferred to both p_l and q_l. 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 $L_j \cup L_{j+1}$, $j = 1, \dots, m-1$. Note that $\rightleftarrows$ is not a transitive binary relation; that is, $L_1 \rightleftarrows L_2 \rightleftarrows L_3$ does not imply $L_1 \rightleftarrows L_3$.

Lemma 10 *Suppose $L_1, \dots, L_m$ are nonempty connected open subsets of $\mathcal{L}$ such that L_i and L_j preserve bi-independence for any $i, j \in \{1, \dots, m\}$. If $L_1 \rightleftarrows \cdots \rightleftarrows L_m$, 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 $p \in L_i \Rightarrow U(p) = U_i(p)$, $i = 1, \dots, m$, represents $\succsim$ on $\bigcup_{i=1}^{m} L_i$.*

Proof. By applying Lemma 8 and positive affine transformations iteratively, we can find affine functions $U_1, \dots, U_m$ 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 that we have a continuous function $V : \mathcal{L} \to \mathbb{R}$ represents $\succsim$.

Step 1: We prove that if for some $i \in \{2, \dots, m-1\}$, $L_{i-1} \rightleftarrows L_{i+1}$, 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 $p \in L_{i-1}$, $q \in L_i$, $r \in L_{i+1}$, $\alpha \in (0,1)$, and $B_\varepsilon(q) \subseteq L_i$ such that $V(p) = V(q) = V(r)$ and $V(\alpha p' p), V(\alpha r' r) \in V(B_\varepsilon(q))$ for any $p' \in L_{i-1}$ and $r' \in L_{i+1}$.

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

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

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

$$\alpha p'p \succsim \alpha r'r \iff U(q_p) \geqslant U(q_r) \iff U(\alpha p'p) \geqslant U(\alpha r'r) \iff U(p') \geqslant U(r'),$$

where the last equivalence follows from $U(p) = U(q)$ and $U(q) = U(r)$.

Step 2: We prove that if $L_1 \rightleftarrows L_m$, there must exist some $i \in \{2, \ldots, m-1\}$ such that $L_{i-1} \rightleftarrows L_{i+1}$. Let $v_i^h := \sup_{p \in L_i} V(p)$ and $v_i^l := \inf_{p \in L_i} V(p)$ for any $i \in \{1, \ldots, m\}$. By definition, $v_i^h > v_i^l$ for any $i \in \{1, \ldots, m\}$, and whenever $L_j \rightleftarrows L_k$ for some $j, k \in \{1, \ldots, m\}$, $(v_j^l, v_j^h) \cap (v_k^l, v_k^h) \neq \emptyset$.

Suppose for any $i \in \{2, \ldots, m-1\}$, $L_{i-1} \not\rightleftarrows L_{i+1}$; 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 $i \in \{2, \ldots, m-1\}$, 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 $m \geqslant 2$, $v_m^l \geqslant v_2^h > v_1^h$, 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 $i \in \{2, \ldots, m-1\}$. Then by $L_{i-1} \not\rightleftarrows L_{i+1}$ for all i, there must be some $j \in \{3, \ldots, m-2\}$ 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, $(v_{j-1}^l, v_{j-1}^h) \cap (v_{j+1}^l, v_{j+1}^h) \neq \emptyset$, and it is straightforward to verify that $L_{j-1} \rightleftarrows L_{j+1}$.

Step 3: We prove that if there exist affine functions $U_i : L_i \to \mathbb{R}$, $i = 1, \ldots, m$, such that the function $U : \bigcup_{i=1}^m L_i \to \mathbb{R}$ that satisfies $p \in L_i \Rightarrow U(p) = U_i(p)$, $i = 1, \ldots, m$, 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 $p, q \in \bigcup_{i=1}^m L_i$. If $p, q \in L_i$ for some $i \in \{1, \ldots, m\}$, $p \succsim q \iff U(p) \geqslant U(q)$. Therefore, for the rest of the proof of this lemma, let $p \in L_i$ and $q \in L_j / L_i$ for some distinct $i, j \in \{1, \ldots, m\}$.

First, suppose $\{p,q\} \not\subseteq L_1 \cup L_m$. Then, either $\{p,q\} \subseteq \bigcup_{i=2}^{m} L_i$ or $\{p,q\} \subseteq \bigcup_{i=1}^{m-1} L_i$. 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 $p \succsim q \iff U(p) \geqslant U(q)$.

Second, consider the case in which $\{p,q\} \subseteq L_1 \cup L_m$. Without loss of generality, let $p \in L_1$ and $q \in L_m \backslash L_1$. 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 $p \succsim q \iff U(p) \geqslant U(q)$.

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 $p \succsim q$. Then, we only need to prove that $p \sim q \Rightarrow U(p) = U(q)$ and $p \succ q \Rightarrow U(p) > U(q)$. 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. Then, $p \sim q \Rightarrow V(p) = V(q) = v_m^h = V(r) \Rightarrow U(p) = U(r) = U(q)$. If $p \succ q$, then $V(p) \geqslant V(r) \geqslant V(q)$ and at least one of the inqualities is strict. It follows that $U(p) \geqslant U(r) \geqslant U(q)$ and at least one of the inqualities is strict. Thus, $p \succ q \Rightarrow U(p) > U(q)$. ■

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

Proof. First, we show that f_1 is strictly increasing. Take $v \in V(Q(p^1))$ and $u, u' \in V(\mathcal{L})\backslash V(Q(p^1))$ such that $u > v > u'$. Pick $p \in Q(p^1)$ and $q, q' \in \mathcal{L}\backslash Q(p^1)$ such that $V(p) = v, V(q) = u$ and $V(q') = u'$. By Lemma 14, $Q(p^1) = \bigcup_{i=1}^{m'} P^*(q^i)$ for some $q^1 \ldots, q^{m'} \in \mathcal{L}_o$. Moreover, for each i there exists $j \neq i$ such that $P^*(q^i) \rightleftarrows P^*(q^j)$. Hence, each $V(P^*(q^i))$ cannot be degenerate. Since each $P^*(q^i)$ is open, it follows that for any $p \in Q(p^1)$, there exist $p', p'' \in Q(p^1)$ such that $p' \succ p \succ p''$. Hence

$$u > v > u' \Rightarrow V(q) \geqslant \sup U_1 > U_1(p) > \inf U_1 \geqslant V(q'),$$

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

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 $q^j \in Q(p^1)$ such that $V(q_j) = v_j$. If $v \in (V_1^l, V_1^h)$, then pick $q \in Q(p^1)$ such that $V(q) = v$. It suffices to show that $U_1(q^j)$ converges to $U_1(q)$. This is clear, since there exists $p \in \mathcal{L}_o$ such that $q \in P^*(p)$, an open set, and U_1 is affine on $P^*(p)$. Now suppose $v = V_1^h$. Pick $q \in \mathcal{L}$ such that $V(q) = v$. Without loss of generality, assume that $\{v_j\}$ is increasing. We want to show that $U_1(q_j)$ converges to $v = V_1^h = \sup U_1$. Suppose not. Then, there exists $r \in Q(p^1)$ such that $r \succ q_j$ for all j. Then *Continuity* implies that $r \succsim q$ and thus $V(r) \geqslant V(q) = V_1^h$, which is a contradiction of the fact that $Q(p^1)$ is the union of some open subsets, all of which have nondegenerate image. 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). ■