Boundedly Rational Backward Induction

Shaowei Ke*

Department of Economics, University of Michigan

January 2016

Abstract

We propose a model of backward induction with a decision maker who has limited ability to identify the optimal choice path and chooses with randomness. Our axioms yield a two-parameter representation of the decision maker's behavior; one characterizes her attitude towards complexity; i.e., her willingness to choose more complicated continuations over simpler ones, the other her error-proneness. We analyze comparative measures of complexity aversion and error-proneness. When complexity aversion and error-proneness disappear, our model becomes fully rational backward induction.

*Shaowei Ke, shaoweik@umich.edu; I am deeply indebted to Faruk Gul and Wolfgang Pesendorfer for their invaluable advice, guidance, and encouragement. I benefit very much from the insightful comments from Stephen Morris. I also thank Dilip Abreu, Tilman Börgers, Kfir Eliaz, Larry Epstein, Xiao Liu, Yusufcan Masatlioglu, Efe Ok, Doron Ravid, Ariel Rubinstein, and numerous seminar participants for their helpful comments.

1 Introduction

Backward induction has been used to predict decision makers' behavior in dynamic economic situations. In a dynamic situation, fully rational backward induction begins with identifying the optimal choice for the last stage, and then rollbacks to the first. The solution is taken as a prediction of how decision makers behave.

Empirical and experimental evidence has suggested that such predictions are often rather different from how people actually behave, even in simple dynamic problems.[1] The reason is simple. Think of a decision maker who needs to make a sequence of choices in a dynamic problem. Since the current-stage choice determines the continuation problems, the decision maker needs to look forward. By using the solution derived from fully rational backward induction to predict such a decision maker's behavior, we implicitly assume that when the decision maker looks forward, she is able to identify the optimal choice path, and also able to follow the path deterministically. However, research often finds that (i) forward looking is imperfect (e.g., Camerer, et al. (1993)), and (ii) decision makers' choices appear to contain randomness from our point of view (e.g., McKelvey and Palfrey (1995, 1998)). Part (i) is self-evident. As for part (ii), in practice, a decision maker may employ a heuristic or deterministic rule to make choices. However, it is unlikely for us, the modelers, to know which heuristic or rule has been employed. Hence, the decision maker's choices appear random.

Motivated by these observations, we formulate a model of a decision maker who cannot identify the optimal choice path and chooses with randomness. Our goal is not to analyze specific heuristics that work only for a particular class of problems, nor to study the decision maker's actual reasoning process.[2] Rather, we present a framework for analyzing how the decision maker's choices may vary with the presentation of the choice problem; that is, how changes further down a decision tree affect the decision maker's choice at the current stage.

[1] For example, see Camerer, et al. (1993), and Binmore, et al. (2002).

[2] For behavioral economic research that analyzes specific heuristics for specific problems, see DellaVigna (2009) for a recent review. Meanwhile, computer scientists have developed hundreds of heuristics to approximately solve dynamic problems. For examples, see Pearl (1984), and Russell and Norvig (1995).

In the resulting model, our decision maker's choice behavior differs from that of a fully rational one in two ways. Facing a decision tree, a decision maker makes a series of choices to reach an outcome.[3] A fully rational decision maker identifies each tree with its best subtree (and hence identifies the optimal choice path), and chooses the best subtree with certainty. Our decision maker behaves *as if* (i) she evaluates a decision tree by employing a general aggregator to aggregate all the subtree values, and (ii) she makes random mistakes when choosing among subtrees. The first departure enables us to identify a comparative measure of *complexity aversion*; that is, the extent to which the decision maker avoids complex subtrees. The second departure enables us to identify a comparative measure of *error-proneness*; that is, the likelihood that the decision maker makes mistakes.

Our model is derived from simple axioms on the decision maker's choices. The primitive is a *random choice rule* that describes how the decision maker chooses among the available subtrees when facing a decision tree. Decision trees are defined recursively: Depth-1 decision trees are finite sets of outcomes, depth-2 decision trees are finite sets consisting of outcomes and depth-1 decision trees, and so on. By definition, a generic decision tree $a = \{a_1, \ldots, a_n\}$ is a set of subtrees. Implicitly, we assume that the modeler can observe the decision tree and the decision maker's behavior in a variety of decision trees repeatedly.

We present axioms that relate how the decision maker chooses in some decision tree $a = \{a_1, \ldots, a_n\}$ to how she would have chosen in each subtree $a_1, \ldots, a_n$. If a decision maker chooses a more often from $\{a, d_1, \ldots, d_n\}$ than b from $\{b, d_1, \ldots, d_n\}$ for all $d_1, \ldots, d_n$; that is, if

$$P(\{a\}, \{a, d_1, \ldots, d_n\}) \geq P(\{b\}, \{b, d_1, \ldots, d_n\})$$

for all $d_1, \ldots, d_n$, we say that the decision maker prefers a to b (see Figure 1). This terminology is appropriate: an error-prone decision maker cannot reveal her preference deterministically but can reveal it statistically. Our first axiom, *Independence*, allows us to identify a complete preference relation from the decision maker's error-prone choices.

[3]The resulting representation can be extended to the case with payoffs along the path.

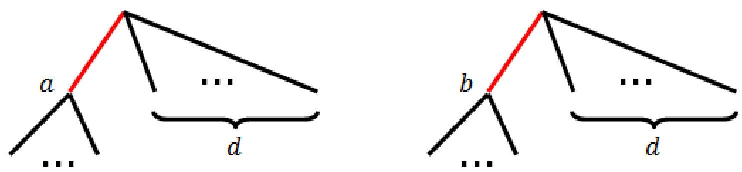

Figure 1: If for any set of subtrees d, the decision maker chooses a more often in the left-hand-side decision tree than b in the right-hand-side decision tree, then the decision maker reveals statistically that she prefers a to b.

The second axiom *Dominance* states that the decision maker chooses $\{a, d_1, \ldots, d_n\}$ over some other subtrees more often than $\{b, d_1, \ldots, d_n\}$ over those subtrees if and only if she prefers a to b. *Independence* and *Dominance* together allow us to identify the decision maker's true objectives from her imperfect attempts at achieving them. The next two axioms describe the manner in which our decision maker can depart from rationality.

Stochastic Set Betweenness requires that if the decision maker prefers a to b and a, b have no common subtrees, then $a \cup b$ should be ranked between a and b in her preference. See Figure 2 for an example. A fully rational decision maker prefers a to b if and only if the best subtree in a is better than the best subtree in b, in which case $a \cup b$ and a have the same best subtree and hence are indifferent. For a decision maker who has limited ability to identify the optimal choice path, *Stochastic Set Betweenness* allows a to be strictly preferred to $a \cup b$ because a appears simpler and better than $a \cup b$.

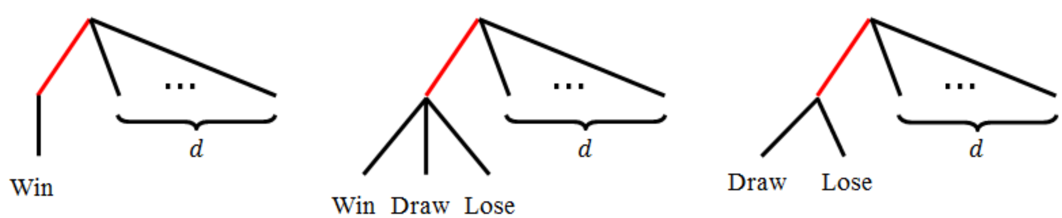

Figure 2: *Stochastic Set Betweenness* implies that the decision maker chooses $\{win\}$ over d more often in the first decision tree than $\{win, draw, lose\}$ over d in the second decision tree, which is in turn chosen more often than $\{draw, lose\}$ over d in the last decision tree.

Finally, *Preference for Accentuating Swaps* implies that if the decision maker prefers an *outcome* x to y, then she also prefers the *depth*-2 decision tree $a = \{\{x, w_1, \ldots, w_m\}, \{y, z_1, \ldots, z_n\}\}$

to $b = \{\{y, w_1, \dots, w_m\}, \{x, z_1, \dots, z_n\}\}$, as long as $0 \le m \le n$. To see what this means, note that in the decision tree b, the outcome y is more visible than x, because y is presented at a smaller subtree $\{y, w_1, \dots, w_m\}$, and x is presented at a larger subtree $\{x, z_1, \dots, z_n\}$. By swapping x for y, we transform b into a. Such a swap renders the better outcome x more visible, while the original tree b emphasizes the inferior outcome y. Accentuating the better outcome in this fashion improves the original tree and hence weakly increases the probability that the decision maker chooses the tree. Note that this axiom only applies to depth-2 trees. See Figure 3 for a concrete example.

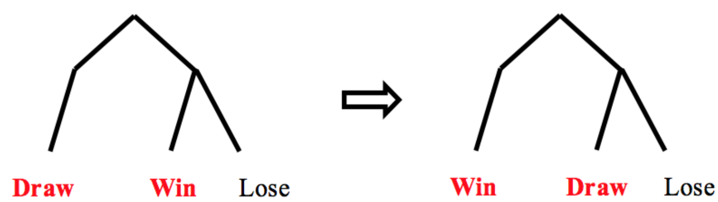

Figure 3: The left-hand-side decision tree is swapped into the right-hand-side one. After the swap, **Win** becomes more salient and **Draw** becomes less so. Thus the decision maker chooses the right-hand-side decision tree with higher probability.

Theorem 1 establishes that these four axioms, together with other technical conditions, yield the following representation of the random choice rule: there exists a *value function* V and $\lambda > 0$ such that

$$P(\{a_i\}, a) = \frac{V(a_i)^\lambda}{\sum_{j=1}^n V(a_j)^\lambda}$$

for any decision tree $a = \{a_1, \dots, a_n\}$. Thus, the random choice rule P is a Luce rule (see Luce (1959)). For a fixed λ, subtrees with higher Luce values are chosen more often. Moreover, there is a function f such that V satisfies

$$V(a) = f^{-1}\left(\frac{1}{n} \sum f(V(a_j))\right) \tag{1}$$

for all $a = \{a_1, \dots, a_n\}$. The aggregator (the right-hand side of (1)) relates tree a's value to its subtree values. Intuitively, the aggregator is a general notion of *average*, while in fully rational backward induction, the aggregator is the maximum function ($V(a) = \max V(a_j)$).[4] Depending on f, our aggregator ranges from the maximum to the minimum. We call a random choice rule that has the representation described above a *Boundedly-Rational Backward-Induction Rule* (BBR). When a decision maker's behavior follows a BBR, she acts as if she assigns an average of the subtree values to evaluate a tree. Then, when she actually chooses, sometimes she mistakenly chooses low-value subtrees. The propensity of making mistakes depends on λ.

The two parameters λ and f quantify the extent to which the decision maker's behavior differs from that of a fully rational decision maker. To see this, first consider the choice between an outcome and a decision tree. Outcomes are the simplest choice objects in our setting. For two decision makers, if decision maker 2 always chooses the outcome over tree more often than decision maker 1 who faces the same problem, then decision maker 2 is said to be more complexity-averse than decision maker 1. In Theorem 2, we show that this holds if and only if f_2 is a concave transformation of f_1. Therefore, the concavity of f describes the decision maker's attitude toward complexity. Next, consider two decision makers who have the same value function and f. Decision maker 1 is characterized by a BBR with parameters λ_1 and decision maker 2 with $\lambda_2 \leq \lambda_1$. Note that the two decision makers share the same ranking of decision trees. Given any decision tree $\{a, b\}$, decision maker 1 chooses a more often than b if and only if 2 chooses a more often than b. However, since $\lambda_2 \leq \lambda_1$, decision maker 2 makes more mistakes; that is, she always chooses the preferred tree less often. Theorem 3 extends this observation to develop a comparative measure of error-proneness.

We take limits of the comparative measure of complexity aversion and error-proneness to find BBR's limiting cases. Fully rational backward induction is a limiting case where both complexity aversion and error-proneness disappear. We present another useful limiting case in which the decision maker chooses deterministically; that is, she never makes a mistake given her valuation of trees. However, she remains complexity-averse; that is, she may choose

[4] This average is called *Kolmogorov-Nagumo* mean. Simple average, quardratic mean, power means, and symmetric CES function are its special cases.

a simpler subtree over a complex one even though had she chosen the complex one, she could have ended up with better outcomes.

Lastly, we apply the BBR to discuss some simple examples. In particular, we discuss (i) when adding a subtree to a node of a decision tree increases the value of the tree, even though the decision tree seems to become more complex, (ii) the effectiveness of the presentation strategy that singles out an outcome from other outcomes, and (iii) the effectiveness of the presentation strategy that repeats an outcome in multiple subtrees of a decision tree.

1.1 Related Literature

Our work belongs to the research of bounded rationality and behavioral economics that aims at developing better models to describe how people choose in complex dynamic situations. Among others, Jéhiel (1995) examines the implication of limited foresight in a special class of repeated games. In his model, the agents can only look forward j steps. Jéhiel equips the agents with a specific heuristic, the average payoff from the j steps, to evaluate the continuation problems beyond the j^{th} step. The heuristic is very reasonable for the games he studies. Gabaix and Laibson (2005) study a reasoning procedure where the decision maker evaluates situations as if they end right away. Based on this heuristic, the procedure determines the optimal number of steps that the decision maker looks forward endogenously.

Although analyzing specific heuristics is important, our work does not focus on a specific heuristic. There are many reasonable heuristics, and it is difficult to know or to test which heuristic is used when we estimate or predict the decision maker's behavior. Therefore, we take a different approach. From simple dynamic-choice examples that illustrate plausible deviations from fully rational backward induction, we abstract testable axioms. Then, we characterize the unique model that is equivalent to those axioms. As a byproduct, unlike most models in the previous literature, our model is fully testable and identifiable.

In the decision theory literature, Gul, Natenzon, and Pesendorfer (2014) establish that when the choice environment is rich enough, the Luce rule is the only random choice rule that satisfies *Independence*. Our model incorporates that axiom, and extends the Luce rule to model how changes further down a decision tree affects the decision tree's Luce value. Gul, Natenzon, and Pesendorfer also study dynamic choice. The decision maker in their model can identify all the duplicates and treat the duplicates as a single choice object. In our case, duplicates should not be treated as a single choice object because the decision maker makes random mistakes when choosing. If a choice problem consists of many duplicates of some choice object, then the other options should be chosen with small probability.

An axiom similar to our *Stochastic Set Betweenness* is first proposed by Bolker (1966). He uses it to propose a generalization of expected value. Gul and Pesendorfer (2001) propose a stronger axiom *Set Betweenness* to model temptation and self-control. In their model, a decision maker may prefer a smaller choice set to a larger one because the larger one contains tempting options. Their axiom applies to the case where choice sets have nonempty intersection, while ours does not.

Fudenberg and Strzalecki (2015) formulate an interesting alternative extension of the Luce rule to dynamic problems. In their model, a choice problem is a set of current-period choices. Each current-period choice yields current-period consumption and a continuation problem for the next period. The utility of a current period choice has three components: a deterministic utility derived from backward induction (taking taste shocks into account), a random component reflecting possible taste shocks, and a choice-attitude term that depends on the number of alternatives available in the next period. The last term, when the relevant coefficient is positive, reflects the decision maker's choice aversion. When the coefficient is negative, the term captures a preference for flexibility beyond the option value associated with the continuation problem.

Fudenberg and Strzalecki are the first to axiomatically extend the Luce rule to a dynamic setting, and the first to axiomatically introduce choice aversion. One of their main findings is that choice aversion is associated with a preference for delaying decisions. Our work has a different goal. We focus on relaxing backward induction. Our axiom 5 also rules out the type of preference for delay that Fudenberg and Strzalecki consider.

The rest of the paper is organized as follows. The axioms, the representation, and our main theorem are presented in Section 2. Section 3 defines and characterizes comparative complexity aversion and error-proneness. Section 4 provides simple application examples. Section 5 concludes.

2 Model

In our model, a decision maker makes a series of choices to reach an outcome. A *decision tree* describes this choice situation. Let D_0 be the set of outcomes. A depth-1 decision tree is a nonempty finite subset of D_0. When the decision maker confronts a depth-1 decision tree $a \subset D_0$, she chooses an outcome $x \in a$. Let $D_1 := K(D_0)$ be the set of all depth-1 decision trees, where $K(\cdot)$ denotes the collection of all nonempty finite subsets of a set. Recursively, we define the set of depth-k decision trees as $D_k := K(D_{k-1} \cup D_0)$. When the decision maker confronts a depth-k decision tree, she chooses at most k times to reach an outcome. Let $D := \bigcup_{k=1}^{\infty} D_k$ be the set of all decision trees. A generic decision tree $a \in D$ is a finite set of *subtrees*. A subtree could either be an outcome or itself a decision tree. Let $\mathcal{D} := D \cup D_0$ denote the set of all decision subtrees.

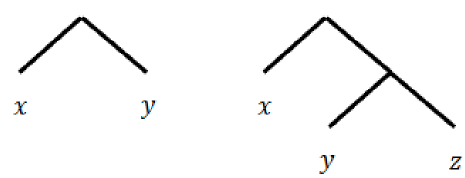

Figure 4: The left-hand-side decision tree is a depth-1 decision tree $\{x, y\} \in D_1$. The right-hand-side decision tree is a depth-2 decision tree $\{x, \{y, z\}\} \in D_2$. In both trees, x, y, z are outcomes.

Confronting a decision tree $b \in D$, the decision maker chooses among b's subtrees with randomness. Let $\mathcal{L}$ be the set of finite-support probability measures on $\mathcal{D}$ endowed with the topology of weak convergence. The probability measure $P(b) \in \mathcal{L}$ describes the probability of choosing b's subtrees. With some abuse of notation, we use $P(a, b)$ instead of $P(b)(a)$ to denote the probability that $P(b)$ assigns to the set of subtrees $a \in D$; that is, the probability that any subtree in a is chosen when the decision tree is b. We call the function $P : D \to \mathcal{L}$ a *random choice rule* (RCR) if $P(a, a) = 1$ for all $a \in D$.

We have in mind a decision maker with limited ability to identify the optimal choice path of a decision tree and may choose with randomness. The first axiom we consider is from Gul, Natenzon, and Pesendorfer (2014). It implies that if the subtrees a are chosen over c more often than b over c, then a should be chosen over d more often than b over d as well.

Axiom 1 *(Independence) For $a, b, c, d \in D$ such that $(a \cup b) \cap (c \cup d) = \emptyset$, $P(a, a \cup c) \geq P(b, b \cup c)$ implies $P(a, a \cup d) \geq P(b, b \cup d)$.*

When we say optimal choice path, implicitly we mean that the decision maker has some true objective. This axiom allows us to identify that true objective consistently, even though the decision maker's behavior is suboptimal. The decision maker may in fact prefer subtree a to b, but she cannot reveal her preference deterministically. However, if we observe that the decision maker always chooses a subtree a over subtrees d more often than a subtree b over d for all d that does not contain a, b, then the decision maker reveals statistically that she prefers a to b.

Definition 1 *For any $a, b \in \mathcal{D}$, we say that the decision maker prefers a to b (and write $a \succeq b$) if $P(\{a\}, \{a\} \cup d) \geq P(\{b\}, \{b\} \cup d)$ for all $d \in D$ such that $a, b \notin d$.*

For simplicity, several axioms below are stated in terms of the uncovered preference. They can be stated in terms of the RCR as well.

To focus on analyzing the decision maker's suboptimal choice behavior, we restrict our attention to the case where her objective does not change over time.[5] Thus, we consider the following simple monotonicity assumption. It states that replacing a subtree with a better one makes the decision tree itself better (see Figure 5).

[5] For research that focuses on changing objective, see Strotz (1955), and Laibson (1997) for examples.

Axiom 2 *(Dominance) For* $a = \{a_1, a_2, \ldots, a_n\}$, $a' = \{a'_1, a_2, \ldots, a_n\}$, $a_1 \succeq a'_1$ *implies* $a \succeq a'$, *and* $a_1 \succ a'_1$ *implies* $a \succ a'$.

The first part of *Dominance* ($a_1 \succeq a'_1$ implying $a \succeq a'$) is satisfied by a fully rational decision maker. The second part ($a_1 \succ a'_1$ implying $a \succ a'$) incorporates some departure from the fully rational behavior. To see this, suppose $a = \{a_1, a_2\}$ and $a' = \{a'_1, a_2\}$, where $a_2 \succ a_1 \succ a'_1$. A fully rational decision maker is indifferent between a and a' since they have the same best subtree a_2. In contrast, *Dominance* implies that $a \succ a'$; that is, our decision maker has some awareness of her own suboptimal behavior, and more often avoids decision trees with inferior subtrees.

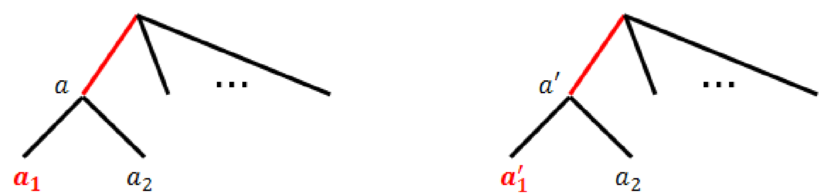

Figure 5: *Dominance* implies that the choice probability of a is higher than a' if and only if a_1 is preferred to a'_1.

The two axioms below encapsulate our model of complexity-averse and error-prone decision making. The first, *Stochastic Set Betweenness*, considers two decision trees a and b that have no subtree in common. For example, suppose a is $\{win\}$, b is $\{draw, lose\}$ and the decision maker reveals statistically that she prefers $\{win\}$ over $\{draw, lose\}$. *Stochastic Set Betweenness* requires that $\{win\}$ is chosen more often than $\{win, draw, lose\}$, which in turn is chosen more often than $\{draw, lose\}$ (see Figure 6).

Axiom 3 *(Stochastic Set Betweenness) For all* $a, b \in D$ *such that* $a \cap b = \emptyset$, $a \succeq b$ *imply* $a \succeq a \cup b \succeq b$.

When $a \succeq b$, a fully rational decision maker should be indifferent between a and $a \cup b$ since they both contain the same best subtree from a. *Stochastic Set Betweenness* allows the decision maker to strictly prefer a over $a \cup b$, reflecting her aversion to more complex

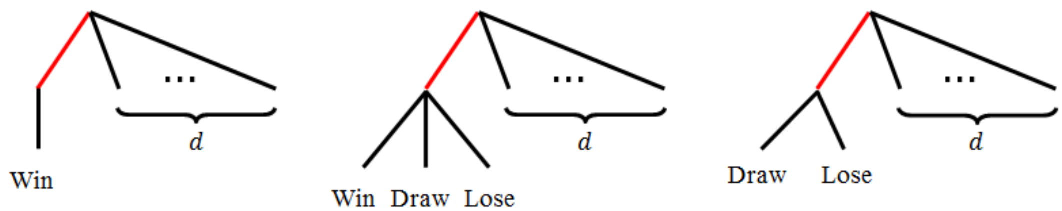

Figure 6: *Stochastic Set Betweenness* implies that the decision maker chooses $\{win\}$ over d more often in the first decision tree than $\{win, draw, lose\}$ over d in the second decision tree, which is in turn chosen more often than $\{draw, lose\}$ over d in the last decision tree.

decision trees. Note that complexity is not about the size of a decision tree. The tree $a \cup b$ is larger than both a and b, but compare to b, $a \cup b$ is better because it contains better subtrees that b does not have.

In the literature, Bolker (1966) is the first to use this type of condition, through which he derives a generalization of expected value.[6] Gul and Pesendorfer (2001) use a related axiom to model temptation and self-control. Our axiom is weaker than the Gul-Pesendorfer version since we require that a and b have empty intersection. To see why this is important, assume that $a = \{win, lose\}, b = \{win, lose^*\}$, where $lose$ and $lose^*$ are two similar unattractive outcomes. If the decision maker struggles with complex decision trees, then it may well be that $\{win, lose, lose^*\}$ is worse than both $\{win, lose\}$ and $\{win, lose^*\}$. Therefore, the Gul-Pesendorfer version of set betweenness is violated.

The next axiom is built upon a simple idea: when a depth-1 tree contains fewer outcomes, each of its outcomes commands more attention. To see what attention has to do with choice, let us first introduce a notion of a "swap." Let $|\cdot|$ denote the cardinality of a set.

Definition 2 *For $a = \{a_1, a_2, \dots, a_n\} \in D_2$ such that $x \in a_1 \backslash a_2$, $y \in a_2 \backslash a_1$, and $|a_1| \geq |a_2|$, a swap of x for y is*

$$\Delta_y^x(a) := a \backslash \{a_1, a_2\} \cup \{a_1', a_2'\}$$

where $a_1' := a_1 \backslash \{x\} \cup \{y\}$, $a_2' := a_2 \backslash \{y\} \cup \{x\}$.

[6] We thank Larry Epstein for referring this paper to us.

Note that the definition requires that $a \in D_2$ is a depth-2 tree, which also implies that x, y are outcomes. In the definition, the outcome x originally belongs to a larger subtree (a_1) than the one (a_2) containing y. We assume that the outcomes from a smaller tree command more attention. Therefore, the swap of x for y accentuates x and masks y. If x is preferred to y, we call this swap an *accentuating swap* to emphasize the fact that the better subtree x is now more visible. When we write $\Delta_y^x(a)$ to denote the swap of x for y, implicitly we have $a_1, a_2 \in d$, $x \in a_1 \backslash a_2$, $y \in a_2 \backslash a_1$, and $|a_1| \geq |a_2|$.

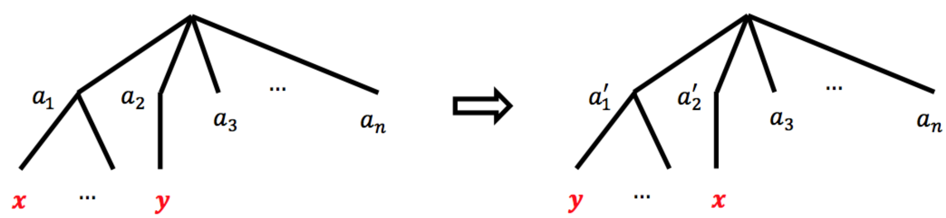

Figure 7: A swap of x for y converts a into $\Delta_y^x(a)$.

Axiom 4 *(Preference for Accentuating Swaps) If $a \in D_2$ and $x \succeq y$, then $\Delta_y^x(a) \succeq a$.*

To understand *Preference for Accentuating Swaps*, consider a depth-2 decision tree $a = \{a_1, a_2\}$ where $a_1 = \{lose, win\}$, $a_2 = \{draw\}$. For a fully rational decision maker, it does not matter which outcome is presented at which part of the tree; that is, she is indifferent between a and $\{\{win\}, \{draw, lose\}\}$. In contrast, when a boundedly rational decision maker looks forward facing the decision tree a, the outcomes in a_1 command less attention than the outcome in a_2, because there are more outcomes competing for attention in a_1 than in a_2. Suppose *win* is preferred to *draw*. An accentuating swap of *win* for *draw* makes the better outcome *win* more salient and the worse outcome less (see Figure 8). Therefore, the swapped decision tree appears to be better, and should be chosen more often than the original tree.

The remaining axioms are technical conditions that help pin down the model. The axiom below states that adding a trivial choice preceding any subtree is irrelevant (see Figure 9). As a result, the decision maker is indifferent between a and $\{a\}$.

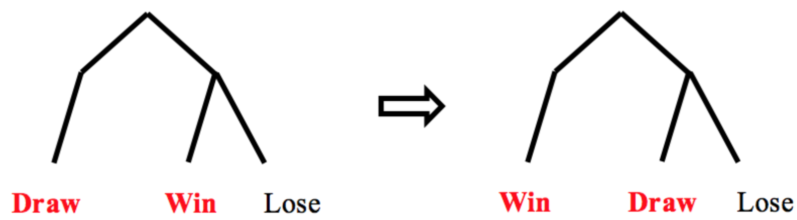

Figure 8: The left-hand-side decision tree is swapped into the right-hand-side one. After the swap, **Win** becomes more salient and **Draw** becomes less so. Thus the decision maker chooses the right-hand-side decision tree with higher probability.

Figure 9: The right-hand-side decision tree extends a into $\{a\}$ by adding a trivial choice.

Axiom 5 *(Consistency) For $a \in \mathcal{D}$, $a \sim \{a\}$.*

The last axiom is *Continuity*. The idea behind it is simple. Suppose we already have a value function that assigns values to trees. For any decision tree, we want its value to not change much when its subtree values are slightly perturbed (see Figure 10). Of course, we do not have the value function to begin with. To impose this notion of continuity, we need to define some topology for the set of decision trees D. We first define the following distance function on the set of subtrees $\mathcal{D}$. For any decision subtrees $a, b \in \mathcal{D}$, we let

$$\nu(a, b) := |P(\{a\}, \{a, b\}) - P(\{b\}, \{a, b\})|$$

be the distance between a, b. In other words, a and b are close whenever the decision maker considers them to be close substitutes. Next, analogous to the definition of the Hausdorff distance, we extend the distance function ν to the set of decision trees D as follows. For any

$c, d \in D$,

$$\mu(c,d) := \begin{cases} \max\left\{\max_{c_i \in c}\min_{d_j \in d} \nu(c_i, d_j), \max_{d_j \in d}\min_{c_i \in c} \nu(d_j, c_i)\right\} & \text{if } |c| = |d| \\ 1 & \text{if } |c| \neq |d| \end{cases}.^7 \tag{2}$$

Note that since c, d are decision trees, they are both sets of subtrees. According to the definition, c and d are close if c's subtrees and d's subtrees are pairwisely close in terms of ν. Unlike the standard Hausdorff distance, we only measure the distance between c and d that have the same cardinality. When they do not have the same cardinality, we consider them "far apart."[8] Therefore, our notion of continuity is weaker. To see we consider trees with different cardinality far apart, suppose we have three indifferent outcomes $x \sim y \sim z$. Had we not required the second line in (2), we will find that $\mu(\{x\}, \{y, z\}) = 0$ according to the first line in (2), but clearly $P(\{x\})$ and $P(\{y, z\})$ are two different probability measures.

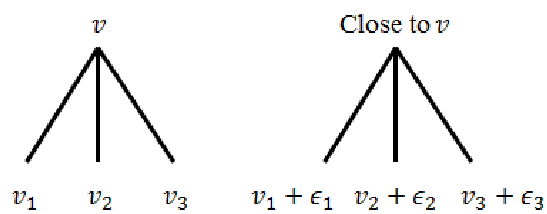

Figure 10: Suppose we already have the value function that evaluates subtrees. The continuity that we need requires that small perturbations to a decision tree's subtree values should have small impact on the decision tree's value.

Axiom 6 *(Continuity) The function P is continuous.*

In our notion of continuity, the function μ depends on P, while P is required to be continuous with respect to μ. This circularity creates no problems. As in standard metric spaces, the metric function is continuous with respect to the topology that it induces. In our case, the distance $\mu(c, d)$ depends on c, d through their subtrees. Thus, like the other axioms,

[7] Although D is a subset of $\mathcal{D}$, one can also show that $D = K(\mathcal{D})$; that is, D is the collection of all nonempty finite subsets of $\mathcal{D}$.

[8] With the other axioms, μ is a pseudometric that only violates $\mu(c, d) = 0 \Rightarrow c = d$. Without the other axioms, μ is a pseudosemimetric that might also violate the triangle inequality.

Continuity builds a connection between decision trees and their subtrees. The function P defined on depth-1 decision trees imposes a continuity requirement on P defined on depth-2 decision trees, and so on.

Our main theorem identifies the only model that can satisfy all the axioms above in a rich choice environment.

Definition 3 *An RCR P is a Boundedly-Rational Backward-Induction Rule (BBR) if there exists a value function $V : \mathcal{D} \to \mathbb{R}_{++}$, a constant $\lambda > 0$, and a strictly increasing continuous function $f : \mathbb{R}_{++} \to \mathbb{R}$ such that for any $a = \{a_1, \ldots, a_n\}$,*

$$P(\{a_i\}, a) = \frac{V(a_i)^\lambda}{\sum_{j=1}^n V(a_j)^\lambda} \tag{3}$$

for $i = 1, \ldots, n$ and

$$V(a) = f^{-1}\left(\frac{1}{n}\sum_{i=1}^n f(V(a_i))\right). \tag{4}$$

Due to (3), the BBR is a dynamic extension of the Luce rule (Luce (1959)), in which $V(a)^\lambda > 0$ is called the Luce value of a. The Luce rule/logit model has been widely used in the industrial organization literature (McFadden (1974)). Clearly from (4), V and f are not independent. However, if we restrict the domain of V to the set of outcomes D_0, then it becomes independent of the function f. In other words, through equation (4), the function f uniquely extends the valuation of outcomes to all finite decision trees. Because of this, define $V^* : D_0 \to \mathbb{R}_{++}$ to be the function such that $V^*(x) = V(x)$, $\forall x \in D_0$, for each value function V. When V, λ, f satisfy the equations above, we say that (V^*, λ, f) represents P. Now, V^*, λ, f are all independent parameters.

The choice behavior characterized by a BBR is different from that by fully rational backward induction in two ways. A fully rational decision maker identifies each tree with its best subtree, and chooses the best subtree with certainty. A decision maker whose behavior follows a BBR behaves *as if* she uses a general aggregator to aggregate the subtree values to evaluate a tree, and makes random mistakes when choosing among subtrees. Intuitively, the aggregator is some general notion of *average*.[9] A rapidly increasing (i.e. convex) f ensures that the aggregator is close to the maximum function, but the aggregator never goes above maximum or below minimum. In contrast, fully rational backward induction always requires $V(a) = \max V(a_j)$. Then, our decision maker's error-prone choice behavior follows the widely used model of mistakes, the Luce rule, as in (3). A subtree with higher value is more likely to be chosen. Higher λ induces fewer mistakes.

We illustrate how this model works through the following example.

Example 1 *Consider $a = \{x_1, x_2, x_3\}$, $b = \{x_1, \{x_2, x_3\}\}$, and $c = \{a, b\}$. Note that tree a and b contain the same set of outcomes, but they present them in different ways. Applying (4) to a, we obtain $V(a) = f^{-1}(\frac{1}{3}\sum f(V(x_i)))$. Each outcome x_i is assigned an equal weight 1/3. Applying (4) to b, we find that $V(b) = f^{-1}(\frac{1}{2}f(V(x_1)) + \frac{1}{4}f(V(x_2)) + \frac{1}{4}f(V(x_3)))$. In other words, since x_1 in b is singled out from x_2, x_3, it commands more attention than x_2, x_3. As a result, the aggregator assigns a higher weight to it. It is easy to see that if x_1 has the highest value, then $V(b) \geq V(a)$. Lastly, facing decision problem c, the decision maker chooses a with probability $\frac{V(a)^\lambda}{V(a)^\lambda + V(b)^\lambda}$, and b with probability $\frac{V(b)^\lambda}{V(a)^\lambda + V(b)^\lambda}$.*

Note that the aggregator works as if at each stage, the decision maker speads out her attention equally among the available subtrees. As in the decision tree b, x_1 and $\{x_2, x_3\}$ share the same weight 1/2, and then x_2 and x_3 split the weight 1/2 equally. This observation immediately leads to the following implication. Imagine one situation where an outcome x_n is presented deep down the decision tree $\{x_1, \{x_2, \ldots, \{x_n\}\}\}$, and another situation where x_n is presented among many other outcomes in a subtree of the decision tree $\{x_1, \{x_2, \ldots, x_n\}\}$. Intuitively, in both situations, compared to x_1, x_n seems to be much less important when the decision maker evaluates the entire decision tree. Therefore, if we replace the outcome x_n with some other outcome y, the value of the entire decision tree should not change much.

Proposition 1 *Consider a BBR (V^*, λ, f), a sequence of outcomes $\{x_i\}$ such that $V(x_i) \in$*

[9] This average is called *Kolmogorov-Nagumo* mean. Simple average, quadratic mean, power means, and symmetric CES function are all special cases of it.

$[\underline{v}, \overline{v}]$, $0 < \underline{v} \leq \overline{v}$, *and an outcome* $y \in D_0$. *Then,* $\lim_{n\to\infty} V(\{x_1, \{x_2, \ldots, \{x_n\}\}\}) - V(\{x_1, \{x_2, \ldots, \{x_{n-1}, \{y\}\}\}\}) = 0$ *and* $\lim_{n\to\infty} V(\{x_1, \{x_2, \ldots, x_n\}\}) - V(\{x_1, \{x_2, \ldots, x_{n-1}, y\}\}) = 0$.

We omit its proof. To state Theorem 1, we define a rich choice environment.

Definition 4 *We say that* (D_0, P) *is rich if* $\forall a, b \in D$, $q \in (0, 1)$, $\exists x \in D_0$ *such that* $x \notin b$ *and* $P(\{x\}, \{x\} \cup a) = q$.

Richness in our setting means that for any given probability q and any set of subtrees a, we can find an outcome x such that it is chosen with probability q when put together with a. Moreover, we can find countably many such outcomes, because the definition requires that the desired outcome x does not belong to b, for any predetermined $b \in D$. Richness is easily satisfied when the outcome set contains lotteries, as shown in the example below.

Example 2 *Let* D_0 *be all the* 50-50 *lotteries over monetary prizes. Let* δ_m *denote the degenerate lottery that yields prize* m *with probability* 1. *For each* 50-50 *lottery* $\frac{1}{2}\delta_{m_1} + \frac{1}{2}\delta_{m_2}$, *let its Luce value be* $V(\frac{1}{2}\delta_{u_1} + \frac{1}{2}\delta_{u_2}) := \exp\{\frac{1}{2}u_1 + \frac{1}{2}u_2\}$. *If* P *satisfies (3) for some fixed* λ, *then one can verify that* (D_0, P) *is rich.*

Our main result below establishes the equivalence between the axioms and the representation. Richness is required in necessity, but not sufficiency. To put it another way, when the choice environment is sparse, there may be other choice models that satisfy our axioms. However, as in Gul, Natenzon, and Pesendorfer (2014), this can be viewed merely as an artifact of the sparse choice environment.

Theorem 1 *If* (D_0, P) *is rich, then an RCR* P *satisfies Axioms 1–6 if and only if it is a BBR.*

Sufficiency is routine. As for necessity, first, by Theorem 1 in Gul, Natenzon, and Pesendorfer (2014), *Independence* and richness ensure the existence of V^λ such that the Luce formula (3) holds. We can pick $\lambda = 1$.[10] The more challenging part of the proof is relating $V(a)$ to the $V(a_i)$'s for any $a = \{a_1, \ldots, a_n\}$ so that (4) holds.

The construction of the function f is similar to how one calibrates a vNM utility function from the data on certainty equivalents for 50-50 gambles (see Machina (1987)). Choose any $a, b \in \mathcal{D}$ such that $V(b) > V(a)$. Set $f(V(a)) := 0$ and $f(V(b)) := 1$. Let

$$f(V(\{a,b\})) := \frac{1}{2}f(V(a)) + \frac{1}{2}f(V(b)) = \frac{1}{2}.$$

To see why this is similar to the calibration of a vNM utility function, think of $V(a)$ and $V(b)$ as monetary prizes x and y, $V(\{a,b\})$ as the certainty equivalent of the 50-50 gamble between x and y, and f as the vNM utility function. Then, the equation above is similar to stating that the utility of the certainty equivalent is equal to the expected utility expression on the right-hand side.

Next, consider $\{a, \{a,b\}\}$ and set

$$f(V(\{a,\{a,b\}\})) = \frac{1}{2}f(V(\{a,b\})) + \frac{1}{2}f(V(a)) = \frac{1}{4}.$$

Similarly, consider $\{b, \{a,b\}\}$ and set $f(V(\{b,\{a,b\}\})) = \frac{1}{2}f(V(\{a,b\})) + \frac{1}{2}f(V(b)) = \frac{3}{4}$. We can continue in this fashion and define f on some subset of the reals.

Note that this construction works only because of our axioms. For example, if the representation is to hold, we must have $b \succ \{a,b\} \succ a$, because in (4) f is strictly increasing. This is guaranteed by *Stochastic Set Betweenness* and *Dominance*. More importantly, consider two decision trees $\{\{a,b\},\{c,d\}\}$ and $\{\{a,c\},\{b,d\}\}$. If P is a BBR, it must be true that

$$\{\{a,b\},\{c,d\}\} \sim \{\{a,c\},\{b,d\}\} \tag{5}$$

[10] At this moment, λ may seem redundant. The reason why we include λ is that it helps us produce a much simpler way to define error-proneness, and much nicer way to take limits of complexity aversion and error-proneness.

because

$$V(\{\{a,b\},\{c,d\}\}) = V(\{\{a,c\},\{b,d\}\}) = f^{-1}\left(\frac{1}{4}f(V(a)) + \cdots + \frac{1}{4}f(V(d))\right).$$

Dominance, richness, and *Preference for Accentuating Swaps* ensure that (5) holds. To see why, consider $\{\{a,b\},\{c,d\}\}$ and suppose $b \succeq c$. First, by richness, find outcomes $x_1, \ldots, x_4$ such that $x_1 \sim a, \ldots, x_4 \sim d$. By *Dominance*, $\{\{a,b\},\{c,d\}\} \sim \{\{x_1,x_2\},\{x_3,x_4\}\}$ and $\{\{a,c\},\{b,d\}\} \sim \{\{x_1,x_3\},\{x_2,x_4\}\}$. According to *Preference for Accentuating Swaps*, since $|\{x_3,x_4\}| \geq |\{x_1,x_2\}|$, a swap of x_2 for x_3 should be weakly preferred to $\{\{x_1,x_2\},\{x_3,x_4\}\}$; that is, $\{\{x_1,x_3\},\{x_2,x_4\}\} \succeq \{\{x_1,x_2\},\{x_3,x_4\}\}$. However, we can swap x_2 and x_3 back, and apply the axiom again to conclude that $\{\{x_1,x_2\},\{x_3,x_4\}\} \succeq \{\{x_1,x_3\},\{x_2,x_4\}\}$. Thus, we have (5).

Recursively, we define f on a countable subset of $\mathbb{R}_{++}$ such that f satisfies

$$f(V(\{a,b\})) = \frac{1}{2}f(V(a)) + \frac{1}{2}f(V(b)).$$

Dominance implies that this subset must be dense in V's image. Hence, together with *Continuity*, f can be extended to V's image.

The construction so far only deals with binary decision trees. In the last step, we show that (4) holds not only for binary trees, but also for all finite decision trees under the same f function. All the axioms are needed to complete this last step.

Proposition 2 below establishes the uniqueness of the BBR representation. In particular, the proposition shows that V^λ is unique up to a positive scalar multiplication, and fixing V^* and λ, f is unique up to a positive affine transformation. From here on, for simplicity, when (D_0, P) is rich and P is a BBR, we say that P is a rich BBR.

Proposition 2 *Suppose P is a rich BBR. If (V^*, λ, f) represents P, then (U^*, χ, g) also represents P if and only if there exist $\alpha_1, \alpha_2 > 0$ and $\beta \in \mathbb{R}$ such that $V^*(x)^\lambda = \alpha_1 U^*(x)^\chi$ and $f(\sqrt[\lambda]{\alpha_1 u^\chi}) = \alpha_2 g(u) + \beta$.*

Note that if $V^* = U^*$ and $\lambda = \chi$, then the uniqueness condition of f becomes $f(u) = \alpha_2 g(u) + \beta$. Also note that if both (V^*, λ, f) and (U^*, χ, g) represent P, then $V^*(x)^\lambda = \alpha_1 U^*(x)^\chi$ implies that $V(a)^\lambda = \alpha_1 U(a)^\chi$.

3 Complexity Aversion and Error-Proneness

Our model describes a decision maker whose behavior falls short of fully rational backward induction on two dimensions, assigning correct values to trees and choosing the best subtree with certainty. The two types of imperfections allow us to obtain two comparative measures to quantify the extent to which a BBR deviates from fully rational backward induction. The first one, complexity aversion, describes the extent to which the decision maker avoids complex subtrees. The second one, error-proneness, describes the likelihood that the decision maker makes mistakes. We characterize these two comparative measures and present their limiting cases.

3.1 Complexity Aversion

Confronting a depth-1 decision tree $a \in D_1$, the decision maker chooses an outcome $x \in a \subset D_0$. An outcome is the least complex choice object in our framework.

Consider two decision makers, labeled 1 and 2, who exhibit the same choice behavior when confronting any depth-1 decision tree. Suppose that compared to decision maker 1, decision maker 2 is always less likely to choose a nondegenerate subtree over an outcome. Then, we say that decision maker 2 is more *complexity-averse* than decision maker 1.

It should be noted that we do not attempt to provide a particular notion of complexity. Rather, we provide a measure to compare different decision makers' propensity of choosing an outcome over a tree.

To formally define comparative complexity aversion, first recall that for an RCR P and a decision tree $a \in D$, $P(a) \in \mathcal{L}$ is the probability measure that describes how the decision maker chooses among a. We say that an RCR P_1 and RCR P_2 coincide on depth-1 decision trees if $P_1(a)$ and $P_2(a)$ are identical for all $a \in D_1$. Let $\succeq_i$ be the preference that P_i induces.

Definition 5 *The RCR P_2 is more complexity-averse than P_1 if P_1 and P_2 coincide on depth-1 decision trees, and for any $x \in D_0$, $a \in D$, $a \succeq_2 x$ implies $a \succeq_1 x$.*

We say that the function f_2 is more concave than f_1 if $f_2 = g \circ f_1$ for some strictly increasing and concave function g. The following theorem establishes that the concavity/curvature of f is the comparative measure of a decision maker's complexity aversion.

Theorem 2 *Suppose the RCRs P_1 and P_2 are rich BBRs. Then, P_2 is more complexity-averse than P_1 if and only if there exist (V^*, λ, f_1) and (V^*, λ, f_2) that represent P_1 and P_2 respectively such that f_2 is more concave than f_1.*

Theorem 2 implies that the function f in a BBR describes a decision maker's complexity aversion the same way that a utility function describes a decision maker's risk aversion. A decision tree becomes a closer substitute for its best outcome if f is less concave, and vice versa. The aggregator converges to $\min V(a_i)$ as f gets more concave, and it converges to $\max V(a_i)$ case as f gets more convex. Note that complexity aversion is not the same as being averse to trees with more subtrees, despite that one of our axioms, *Preference for Accentuating Swaps*, is closely related to the size of trees.

Some BBR has a constant measure of complexity aversion. These BBRs can potentially be useful in the applications, just as the CARA or CRRA utility functions. They are characterized by the following simple choice behavior.

Definition 6 *Suppose w, x, y, z are outcomes. We say that a BBR P is homogeneous if $P(\{x\}, \{w, x\}) \geq P(\{y\}, \{y, z\})$ implies $P(\{x\}, \{x, \{w, x\}\}) \geq P(\{y\}, \{y, \{y, z\}\})$.*

The definition says that with a homogeneous BBR P, if x is chosen more frequently *from* $\{w, x\}$ than y *from* $\{y, z\}$, then x should also be chosen more frequently *over* $\{w, x\}$ than y *over* $\{y, z\}$. Proposition 3 below shows that such BBRs have the following representation.

Definition 7 *An RCR P is a Constant-Complexity-Averse (CCA) BBR if there exists a function $V : \mathcal{D} \to \mathbb{R}_{++}$, $\lambda > 0$ and $\gamma \in \mathbb{R}$ such that for any $a = \{a_1, \ldots, a_n\}$,*

$$P(\{a_i\}, a) = \frac{V(a_i)^\lambda}{\sum_{j=1}^n V(a_j)^\lambda}, i = 1, \ldots, n$$

and either

$$V(a) = \left(\frac{1}{n} \sum_{i=1}^{n} [V(a_i)]^\gamma \right)^{1/\gamma}$$

or ($\gamma = 0$)

$$V(a) = \sqrt[n]{\prod_{i=1}^{n} V(a_i)}.$$

Hence, the value of a decision tree a is the γ-power mean of a's subtree values. The following result establishes that the homogeneity condition is equivalent to constant complexity aversion.

Proposition 3 *A rich BBR P is homogeneous if and only if it is a CCA BBR.*

When V, λ, γ satisfy the equations above, we say that (V^*, λ, γ) represents the CCA BBR P. We use the term CCA to describe such BBRs because their f functions are similar to the CRRA utility functions with domain $\mathbb{R}_{++}$. If we mimic the definition of relative risk aversion and apply it to the f function of a CCA BBR, we know that

$$-v \frac{f''(v)}{f'(v)} = 1 - \gamma.$$

Recall that f_2 is more concave than f_1 if and only if $-\frac{f_2''}{f_2'} \geq -\frac{f_1''}{f_1'}$, if both f_1 and f_2 are twice differentiable. Since $v \in \mathbb{R}_{++}$, it is clear that if $\gamma_1 \geq \gamma_2$, the RCR P_2 is more complexity-averse than P_1. Thus, the CCA BBRs with the same V^*, λ are ordered with respect to the parameter γ.

3.2 Error-Proneness

Under richness and *Independence*, an RCR is a Luce rule. Therefore, facing a binary choice problem $\{a, b\} \in D$, if a decision maker chooses a with lower probability than b, then we know that $V(b) \geq V(a)$ and hence $b \succeq a$. When comparing two decision makers, 1 and 2, who both prefer b over a, if decision maker 2 always chooses a with higher probability, then we say that decision maker 2 is more error-prone. Formally, we define it as follows.

Definition 8 *The RCR P_2 is more error-prone than P_1 if there exists a function $h : (0, \frac{1}{2}] \to \mathbb{R}_{++}$ such that $h(p) \leq p$, $h(\frac{1}{2}) = \frac{1}{2}$, and*

$$P_1(\{a\}, \{a, b\}) = h(P_2(\{a\}, \{a, b\})) \tag{6}$$

for $\forall\{a, b\} \in D$ with $P_2(\{a\}, \{a, b\}) \leq \frac{1}{2}$.

The equation (6) together with $h(\frac{1}{2}) = \frac{1}{2}$ and $h(p) \leq p$ immediately implies that $a \succeq_2 b$ if and only if $a \succeq_1 b$. Moreover, since $h(p) \leq p$, (6) implies that decision maker 2 is always more likely to choose the inferior tree than decision maker 1. The theorem below characterizes our notion of error-proneness.

Theorem 3 *Suppose the RCRs P_1 and P_2 are rich BBRs. Then, P_2 is more error-prone than P_1 if and only if there exist (V^*, λ_1, f) and (V^*, λ_2, f) that represent P_1 and P_2 respectively such that $\lambda_1 \geq \lambda_2$.*

Clearly, a smaller λ corresponds to more error-prone choice behavior. As for necessity, suppose there are four subtrees a, b, c, d such that $P_2(\{a\}, \{a, b\}) = P_2(\{c\}, \{c, d\}) \leq \frac{1}{2}$. Say decision maker 2 is more error-prone than decision maker 1. Our definition of comparative error-proneness implies that $P_1(\{a\}, \{a, b\}) \leq P_2(\{a\}, \{a, b\})$ and $P_1(\{c\}, \{c, d\}) \leq P_2(\{c\}, \{c, d\})$. More importantly, it must also be true that $P_1(\{a\}, \{a, b\}) = P_1(\{c\}, \{c, d\})$, because

$$\begin{aligned} P_1(\{a\},\{a,b\}) &= h(P_2(\{a\},\{a,b\})) \\ &= h(P_2(\{c\},\{c,d\})) \\ &= P_1(\{c\},\{c,d\}). \end{aligned}$$

In other words, if decision maker 2 is more error-prone than decision maker 1, then $P_2(\{a\},\{a,b\}) = P_2(\{c\},\{c,d\})$ implies $P_1(\{a\},\{a,b\}) = P_1(\{c\},\{c,d\})$. This property yields a functional equation for which the exponential is the solution.

3.3 Limiting Cases of the BBR

By taking limits of the two measures we just derive, we can study limiting cases of the BBR. Several limiting cases of the BBR are worth noting. First, fix some V^* (some value function defined only for outcomes). To illustrate, consider a collection of BBRs parametrized by two numbers, $\lambda > 0$ and γ: the CCA BBRs (V^*, λ, γ). Consider a simple decision tree $a = \{x_2, \{x_1, x_3\}\}$. For a CCA BBR (V^*, λ, γ), we know that

$$\begin{aligned} V(\{x_1,x_3\}) &= \left(\frac{1}{2}[V(x_1)]^\gamma + \frac{1}{2}[V(x_3)]^\gamma\right)^{1/\gamma} && (7) \\ &= \left(\frac{1}{2}[V^*(x_1)]^\gamma + \frac{1}{2}[V^*(x_3)]^\gamma\right)^{1/\gamma}. && (8) \end{aligned}$$

The choice probability of $\{x_1, x_3\}$ when the decision tree is a is

$$P(\{x_1,x_3\},a) = \frac{V(\{x_1,x_3\})^\lambda}{V(x_2)^\lambda + V(\{x_1,x_3\})^\lambda}. \qquad (9)$$

When both λ and γ are arbitrarily large, the choice behavior of the BBR coincides with fully rational backward induction (with an equal-probability tie-breaking rule), because (7) implies that

$$\lim_{\gamma\to\infty} V(\{x_1, x_3\}) = \max\{V(x_1), V(x_3)\}$$

and (9) implies that

$$\lim_{\lambda\to\infty} P(\{x_1, x_3\}, a) = \begin{cases} 1, & \text{if } V(\{x_1, x_3\}) > V(x_2) \\ \frac{1}{2}, & \text{if } V(\{x_1, x_3\}) = V(x_2) \\ 0, & \text{if } V(\{x_1, x_3\}) < V(x_2) \end{cases}. \tag{10}$$

These two equations are exactly what fully rational backward induction (with an equal-probability tie-breaking rule) requires.

Another useful limiting case can be derived by letting γ be finite, while keeping λ arbitrarily large. In this limiting case, the decision maker never makes a mistake. She only chooses the subtrees with the highest value as in (10). However, she may be averse to complex subtrees deterministically. For instance, suppose $V^*(x_i) = i$ and $\gamma = -1$. Then, equation (7) becomes

$$\begin{aligned} V(\{x_1, x_3\}) &= \left(\frac{1}{2}\cdot 1 + \frac{1}{2}\cdot\frac{1}{3}\right)^{-1} \\ &= 1.5 < 2 = V(x_2). \end{aligned}$$

Hence, facing a, this decision maker chooses the safe bet x_2 with certainty, despite that had she chosen $\{x_1, x_3\}$, she would have ended with the best outcome x_3. In other words, had she not shied away from the complex subtree, she would have been better off.

In this limiting case, we can see clearly that complexity aversion is not about the size of decision trees. Suppose the decision problem is $\{x_2, \{x_2, x_3\}\}$ instead. For any γ, the decision maker will choose $\{x_2, x_3\}$ with certainty because any single outcome in $\{x_2, x_3\}$ is at least as good as x_2. By *Stochastic Set Betweenness*, $\{x_2, x_3\} \succeq x_2$. Hence, although our decision maker chooses suboptimally, she is not completely irrational.

4 Applying the BBR to Decision Trees

We apply the BBR to several simple examples in this section. Before stating the examples, let us note that so far, given a decision tree $a = \{a_1, \ldots, a_n\}$, our model predicts the probability with which the decision maker chooses each subtree a_i. It has not predicted how she will continue to choose after choosing some non-outcome subtree a_j. Thus, we have presented a theory that relates a decision maker's choice at the first stage of a decision tree to how she would have chosen had she been asked to make choices in each of its subtrees. We have not addressed the decision maker's choices after the first stage of a decision tree.

A simple way to extend our model to the subsequent-stage choices is to impose *history independence*. Suppose $b = \{b_1, \ldots, b_n\}$ and $b_1 = \{a_1, \ldots, a_m\}$. Under history independence, the probability that a_i is chosen from b through b_1 is

$$\pi(a_i, b_1, b) = P(\{b_1\}, b) \times P(\{a_i\}, b_1).$$

In general, conditional on choosing b_1 from b, the probability of choosing a_i from b_1 may also depend on $b_2, \ldots, b_n$. By assuming history independence, only the chosen subtree b_1 matters. History independence is a maintained hypothesis in the analysis throughout this section.

Below we briefly discuss a simple question and two simple examples. Our first question is: According to the BBR, when adding a subtree to a decision tree increases the value of the decision tree, despite that the size of the tree increases? This question is not only of theoretical interest, but also has practical relevance. For instance, in the marketing literature, researchers find that excluding some less appealing products from a store's assortment often boosts the sales (see Broniarczyk, et al. (1998), Simonson (1999), and Boatwright and Nunes (2001)). Similarly, in the finance literature, research finds that the 401(k) plan participation rate decreases with the number of fund options (e.g., Iyengar and Kamenica (2010)). The questions in these studies are closely related to our question.

To answer the question, let us begin with a simpler example. Suppose a decision maker is facing a set of outcomes $a = \{x_1, \ldots, x_n\}$. A principal is considering whether or not to add another outcome x_{n+1} to a. In this example, adding x_{n+1} has two effects to the value of a. Note that $V(a) = f^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} f(V(x_i))\right)$. Adding x_{n+1} to a adds $\frac{1}{n+1}f(V(x_{n+1}))$ to the argument of f^{-1}, but also reduces the weight of each x_i from $1/n$ to $1/(n+1)$. Simple algebra shows that $V(a \cup \{x_{n+1}\}) \geq V(a)$ if and only if $f(V(x_{n+1})) \geq \frac{1}{n}\sum_{i=1}^{n} f(V(x_i))$. In other words

$$x_{n+1} \succeq a \Longleftrightarrow a \cup \{x_{n+1}\} \succeq a.$$

In terms of axioms, this observations follows from *Stochastic Set Betweenness* and *Dominance*.

Therefore, in general, suppose we have a depth-k decision tree $a = \{a_1, \ldots, a_n\} \in D_k \subset D$ such that $a_i = \{a_{i,1}, \ldots, a_{i,n_i}\}$, $a_{i,j} = \{a_{i,j,1}, \ldots, a_{i,j,n_{i,j}}\}$, and so on. Under this notation, a_i is at most depth-$(k-1)$, $a_{i,j}$ is at most depth-$(k-2)$, and so on. Suppose the principal is considering whether or not to attach a subtree b to the tree $a_{i_1,\ldots,i_j}$ of a by replacing $a_{i_1,\ldots,i_j}$ with $a_{i_1,\ldots,i_j} \cup \{b\}$. Here $a_{i_1,\ldots,i_j}$ needs to be a non-outcome subtree ($a_{i_1,\ldots,i_j} \in D$). Otherwise, $a_{i_1,\ldots,i_j} \cup \{b\}$ does not make sense.

From the previous example, we immediately know that attaching b to $a_{i_1,\ldots,i_j}$ increases the value of $a_{i_1,\ldots,i_j}$ if and only if $V(b) \geq V(a_{i_1,\ldots,i_j})$. By *Dominance* and an induction argument, it can be shown that the value of a increases if and only if the value of $a_{i_1,\ldots,i_j}$ increases. Thus, we have the following proposition whose proof is omitted.

Proposition 4 *If $a_{i_1,\ldots,i_j} \in D$ and $V(b) \geq V(a_{i_1,\ldots,i_j})$, adding b to decision tree a's subtree $a_{i_1,\ldots,i_j}$ increases the value of a. If $V(b)$ is greater than the values of all $a_{i_1,\ldots,i_j} \in D$, adding b to any non-terminal node of a increases the value of a.*

Intuitively, the result says that if a subtree is good enough, then adding it to a decision tree increases the tree's value. This observation is consistent with the empirical and experimental evidence we mention above.

To illustrate how the BBR may be used as a unified framework to understand the effectiveness of some popular presentation strategies, we present two examples. First, suppose there are three outcomes, $x_1, \ldots, x_3$. By presenting them in a dynamic way such as $\{x_1, \{x_2, x_3\}\}$, x_1 is singled out from the others and hence *emphasized*. Intuitively, $\{x_1, \{x_2, x_3\}\}$ increases the choice probability of x_1, compared to presenting $\{x_1, x_2, x_3\}$.

This is indeed the case. Whenever P is a BBR, $P(\{x_1\}, \{x_1, \{x_2, x_3\}\}) > P(\{x_1\}, \{x_1, x_2, x_3\})$. The reason is simple. By *Stochastic Set Betweenness*, $V(\{x_2, x_3\}) \leq \max\{V(x_2), V(x_3)\}$. Say $V(x_3) \geq V(x_2)$. Then

$$\frac{V(x_1)^\lambda}{V(x_1)^\lambda + V(\{x_2, x_3\})^\lambda} \leq \frac{V(x_1)^\lambda}{V(x_1)^\lambda + V(x_3)^\lambda} < \frac{V(x_1)^\lambda}{V(x_1)^\lambda + V(x_2)^\lambda + V(x_3)^\lambda}.$$

Therefore, $P(\{x_1\}, \{x_1, \{x_2, x_3\}\}) > P(\{x_1\}, \{x_1, x_2, x_3\})$.

In practice, when a set of options are presented, we often observe that some of the options is emphasized in a similar fashion. In policy design, the default option can be understood as being emphasized (see Samuelson and Zeckhauser (1988), Fernandez and Rodrik (1991), Kahneman, et al. (1991), and Masatlioglu and Ok (2005)). In other words, in the decision tree $\{x_1, \{x_2, x_3\}\}$ if the decision maker dislikes the default option x_1, she moves on to choose between x_2 and x_3. Similarly, in supermarkets, some products are emphasized because they are presented at more salient places, while many others are presented on shelves in a standard way. Thus, the BBR is consistent with these observations.

Of course, there are other theories that can explain these observations. However, we do not need a new theory or framework to analyze the following different presentation strategy, if we use the BBR. Again, suppose there are three outcomes x_1, x_2, x_3. By presenting them using the decision tree $a = \{\{x_1, x_2\}, \{x_1, x_3\}\}$, x_1 appears multiple times in the decision tree and hence *repeated*. Intuitively, a should also increase the choice probability of x_1, compared to presenting $\{x_1, x_2, x_3\}$.

This is true as well, whenever P is a BBR. The probability that x_1 is chosen in a is equal to $p = P(\{\{x_1, x_2\}\}, a) \times P(\{x_1\}, \{x_1, x_2\}) + P(\{\{x_1, x_3\}\}, a) \times P(\{x_1\}, \{x_1, x_3\})$, under the assumption of history independence. To see why $p > P(\{x_1\}, \{x_1, x_2, x_3\})$, simply note that p is a weighted average of $P(\{x_1\}, \{x_1, x_2\})$ and $P(\{x_1\}, \{x_1, x_3\})$. Since $P(\{x_1\}, \{x_1, x_2\}), P(\{x_1\}, \{x_1, x_3\}) > P(\{x_1\}, \{x_1, x_2, x_3\})$, we know that $p > P(\{x_1\}, \{x_1, x_2, x_3\})$.

In practice, repeating an option is also common. For instance, an advertised website on Google's search results recurs at multiple pages (often up to 10 pages). Such design increases the probability of clicking the advertised website. In supermarkets, some snacks are presented not only on the shelf, but also right next to the checkout counter. Such assortment presentation strategy increases the chance that the decision makers buy the snacks. Thus, the BBR is also consistent with these observations.

5 Concluding Remarks

In dynamic problems, we usually use fully rational backward induction to describe a decision maker's choice behavior. By using fully rational backward induction, we implicitly assume that the decision maker is able to identify the optimal choice path and follow it when choosing. However, empirical and experimental research often finds that decision makers' foresight is imperfect, and their choice behavior appears random to us, the modelers.

We propose several simple behavioral axioms, from which we derive a boundedly rational backward induction model, the BBR. A decision maker whose choice behavior follows some BBR chooses as if she (i) evaluates a decision tree by aggregating all its subtree values (instead of using the maximum), and (ii) makes random mistakes when choosing. As a result, the decision maker is likely to avoid a complex subtree even if it contains the best outcome.

Based on the model, we identify comparative measures of complexity aversion and error-proneness to compare different decision makers' behavior. In particular, we find that the concavity of f in (4) characterizes the decision maker's complexity aversion, the same way that the concavity of a vNM utility function characterizes a decision maker's risk aversion, and a constant λ measures the propensity that the decision maker makes mistakes. As complexity aversion and error-proneness disappear, our model converges to fully rational backward induction in the limit.

A Appendix

Lemma 1 *For $d = \{d_1, \dots, d_n\}$ such that $b \in d_1 \backslash d_2$, $a \in d_2 \backslash d_1$, and $|d_1| = |d_2|$, $\Delta_a^b(d) \sim d$.*

Proof of Lemma 1: By richness, find outcomes $x_3, \dots, x_n$ such that $P(\{x_i\}, \{x_i, d_1\}) = q = P(\{d_i\}, \{d_1, d_i\})$, $i = 3, \dots, n$. By *Independence*, $x_i \sim d_i$, $i = 3, \dots, n$. By *Dominance*, $d \sim \{d_1, d_2, x_3, \dots, x_n\}$. We can similarly replace all subtrees of d_1 and d_2 with outcomes. Say d_1 is replaced with $c_1 \in D_1$ and d_2 is replaced with $c_2 \in D_1$ such that $d_{1,i} \sim c_{1,i} \in D_0$ and $d_{2,j} \sim c_{2,j} \in D_0$. Then, $d \sim \{c_1, c_2, x_3, \dots, x_n\} = d' \in D_2$.

Without loss of generality, say $a = d_{1,1}$ and $b = d_{2,1}$. Say $b \succeq a$. Then, $|c_1| = |d_1| \geq |d_2| = |c_2|$ implies $\Delta_{c_{1,1}}^{c_{2,1}}(d') \succeq d'$ by *Preference for Accentuating Swaps*. Let $d'' := \Delta_{c_{1,1}}^{c_{2,1}}(d') = \{c_1'', c_2'', x_3, \dots, x_n\}$, $c_1'' := c_1 \backslash \{c_{1,1}\} \cup \{c_{2,1}\}$, and $c_2'' := c_2 \backslash \{c_{2,1}\} \cup \{c_{1,1}\}$. Notice that now $c_{1,1} \in c_2'' \backslash c_1''$ and $c_{2,1} \in c_1'' \backslash c_2''$. Clearly $|c_1''| = |c_2''|$, and hence $|c_2''| \leq |c_1''|$ implies $\Delta_{c_{1,1}}^{c_{2,1}}(d'') \succeq d''$. It is not difficult to see that $\Delta_{c_{1,1}}^{c_{2,1}}(d'') = d'$. Therefore, $\Delta_{c_{1,1}}^{c_{2,1}}(d') \sim d'$. Lastly, by Dominance, it can be shown that $\Delta_{c_{1,1}}^{c_{2,1}}(d') = \Delta_a^b(d)$.

■

Proof of Theorem 1: First, we show sufficiency. Suppose (D_0, P) is rich and P is a BBR. According to (3), the RCR P is a Luce rule. Therefore, we know that (i) $V(a) \geq V(b)$ if and only if $V(a)^\lambda \geq V(b)^\lambda$, which implies $a \succeq b$, and (ii) Luce rule satisfies IIA and IIA implies *Independence* (see Luce (1959) and McFadden (1974) for the definition of the IIA condition).

Dominance is satisfied because f is strictly increasing. To see why *Continuity* is satisfied, note that a low $\nu(a, b)$ is equivalent to that $V(a)$ and $V(b)$ are close, as P is a Luce rule. For two sets $c = \{c_1, \dots, c_n\}, d = \{d_1, \dots, d_n\}$, $\mu(c, d)$ being small implies that there is a bijection $\pi : \{1, \dots, n\} \to \{1, \dots, n\}$ such that $\max_i \nu(c_i, d_{\pi(i)})$ is small too. Thus, $V(c_i)$ and $V(d_{\pi(i)})$ are close. By f's continuity, we know that $V(c)$ and $V(d)$ should be close and hence $\nu(c, d)$ is also small.

As for *Stochastic Set Betweenness*, consider any $a, b \in D$ such that $a \cap b = \emptyset$, say $a = \{a_1, \ldots, a_m\}$, $b = \{b_1, \ldots, b_n\}$. If $a \succeq b$, then $V(a) \geq V(b)$. Since $f(V(a)) = \frac{1}{m}\sum_{i=1}^{m} f(V(a_i))$, $f(V(b)) = \frac{1}{n}\sum_{i=1}^{n} f(V(b_i))$,

$$\begin{aligned} f(V(d_1 \cup d_2)) &= \frac{1}{n_1 + n_2}\left(\sum_{i=1}^{m} f(V(a_i)) + \sum_{i=1}^{n} f(V(b_i))\right) \\ &= \frac{m}{m+n} f(V(a)) + \frac{n}{m+n} f(V(b)). \end{aligned}$$

Thus, $V(a) \geq V(a \cup b) \geq V(b)$, and *Stochastic Set Betweenness* is satisfied. *Consistency* is satisfied since $V(\{a\}) = f^{-1}(f(V(a))) = V(a)$.

For $a = \{a_1, \ldots, a_n\}$ such that $x \in a_1 \backslash a_2$, $y \in a_2 \backslash a_1$, $x \succeq y$, and $|a_1| \geq |a_2|$, let $a_1' := a_1 \backslash \{x\} \cup \{y\}$ and $a_2' := a_2 \backslash \{y\} \cup \{x\}$. We have

$$\begin{aligned} |a| \cdot [f(V(\Delta_y^x(a))) - f(V(a))] &= f(V(a_1')) + f(V(a_2')) - f(V(a_1)) - f(V(a_2)) \\ &= (f(V(x)) - f(V(y)))\left(\frac{1}{|a_2|} - \frac{1}{|a_1|}\right) \geq 0. \end{aligned}$$

Therefore, *Preference for Accentuating Swaps* is satisfied.

Next, we prove necessity. When (D_0, P) is rich and P satisfies *Independence*, P must be a Luce rule (see Gul, Natenzon, and Pesendorfer (2014)); that is, there exists a function $V : \mathcal{D} \to \mathbb{R}_{++}$ that assigns each decision subtree $a \in \mathcal{D}$ a Luce value $V(a) > 0$, and for $a = \{a_1, \ldots, a_n\}$,

$$P(\{a_i\}, a) = \frac{V(a_i)}{\sum_{j=1}^{n} V(a_j)}.$$

Implicitly in the above equation, we set $\lambda = 1$. It is easy to see that $a \succeq b$ implies $V(a) \geq V(b)$.

Take any $x \in D_0$, and suppose $V(x) = v$. We first prove that $V(D_0) = \mathbb{R}_{++}$. For any $v' \in \mathbb{R}_{++}$, we can find an $y \in D_0$ such that $V(y) = v'$, because by richness, we can find $y \in D_0$ such that $P(\{y\}, \{x, y\}) = \frac{v}{v+v'}$. Then, we know that $V(y) = v$. By richness, we also know that for any v and any given finite set $a \subset D_0$, we can find $z \in D_0$ such that $V(z) = v$

and $z \notin a$.

Due to *Stochastic Set Betweenness*, for any $a \in D_1$, $V(a) \in [\min V(a_i), \max V(a_i)]$. Furthermore, by richness, for any $v \in \mathbb{R}_{++}$, we can find $x \neq y$ such that $V(x) = V(y) = v$. Thus, $\{x, y\} \in D_1$ and $V(\{x, y\}) = v$. Hence, $V(D_1) = \mathbb{R}_{++}$. We can do this for all D_k, and find that $V(\mathcal{D}) = \mathbb{R}_{++}$.

A standard induction argument shows that P satisfies *Dominance* only if the following statement holds. For $a = \{a_1, \ldots, a_n\}$ and $b = \{b_1, \ldots, b_n\}$ such that $a_i \succeq b_i$, $a \succeq b$, and if any of the former is strict, so is the latter. Let us call this statement Dominance*.

Next, we show that for all $a = \{a_1, \ldots, a_n\} \in D$, by *Dominance*, there is a sequence of symmetric and strictly increasing function M_n's such that $V(a) = M_n(V(a_1), \ldots, V(a_n))$, where $M_n : \mathbb{R}^n_{++} \to \mathbb{R}_{++}$. The previous arguments show that M_n's domain is indeed $\mathbb{R}^n_{++}$. For any $(v_1, \ldots, v_n) \in \mathbb{R}^n_{++}$, we can find $\{x_1, \ldots, x_n\}$ such that $V(x_i) = v_i$. It is guaranteed by richness that x_i's are distinct, even if $v_i = v_j$ for some i, j. Now for any $a = \{a_1, \ldots, a_n\}$ such that $V(a_i) = v_i$, it has to be true that $V(a) = V(\{x_1, \ldots, x_n\})$, because we have $V(a_i) \geq V(x_i)$ which by Dominance* implies $V(a) \geq V(\{x_1, \ldots, x_n\})$, and the other way around. Therefore, we let M_n map $(v_1, \ldots, v_n)$ to $V(\{x_1, \ldots, x_n\})$, which delivers a well-defined sequence of functions. Clearly M_n is symmetric, meaning that $M_n(v_1, \ldots, v_n) = M_n(v_{\pi(1)}, \ldots, v_{\pi(n)})$ for any bijection $\pi : \{1, \ldots n\} \to \{1, \ldots, n\}$. Furthermore, the strictness in *Dominance* implies that M_n is strictly increasing, and *Consistency* implies that $M_1(v) = v$.

Notice that by *Dominance*, $\nu(a, b) = 0$ if $\nu(a_i, b_i) = 0$ for all i. It is then straightforward to translate *Conintuity* into the following statement. For $\forall \varepsilon > 0$, $a = \{a_1, \ldots, a_n\}$, there exists a $\delta > 0$ such that for all $b = \{b_1, \ldots, b_n\}$, if $\max_i \nu(a_i, b_i) < \delta$, then $\nu(a, b) < \varepsilon$.

We show in this paragraph that M_n is continuous. Consider any $\varepsilon > 0$ and $(v_1, \ldots, v_n)$, where $V(a_i) = v_i$, $a = \{a_1, \ldots, a_n\}$. Now for $\varepsilon' = \frac{\varepsilon}{\varepsilon + 2V(a)}$, we can find a $1 > \delta' > 0$ such that if $\max_i \nu(a_i, b_i) < \delta'$, then $\nu(a, b) < \varepsilon'$. Notice that $\nu(a_i, b_i) < \delta'$ means that

$$\frac{|V(a_i) - V(b_i)|}{V(a_i) + V(b_i)} < \delta'. \tag{11}$$

If $V(a_i) \leq V(b_i)$, (11) becomes $\frac{V(b_i)-V(a_i)}{V(b_i)+V(a_i)} < \delta'$, which is equivalent to

$$V(b_i) - V(a_i) < \frac{2V(a_i)}{1/\delta' - 1} := \delta_i.$$

If $V(a_i) \geq V(b_i)$, we have

$$V(a_i) - V(b_i) < \frac{2V(a_i)}{1/\delta' + 1} < \delta_i$$

Thus, now we know that if $\max |V(a_i) - V(b_i)| < \delta := \min \delta_i$, $\nu(a, b) < \varepsilon'$. Since $\nu(a, b) < \varepsilon'$ implies

$$|V(a) - V(b)| < \frac{2V(a)}{1/\varepsilon' - 1} = \varepsilon$$

we know that M_n is continuous.

Lemma 1 implies that for $x_i \in D_0$, $i = 1, \ldots, 4$, where $V(x_i) = v_i$, $\{\{x_1, x_2\}, \{x_3, x_4\}\} \sim \{\{x_1, x_3\}, \{x_2, x_4\}\}$. Therefore, we know that

$$M_2(M_2(v_1, v_2), M_2(v_3, v_4)) = M_2(M_2(v_1, v_3), M_2(v_2, v_4)). \tag{12}$$

By *Stochastic Set Betweenness*, for any $a, b \in D$, if $a \sim b$, then $a \sim a \cup b \sim b$; that is, if $V(a) = V(b)$, then $V(a) = V(a \cup b) = V(b)$. In particular, we know that

$$M_2(v, v) = v. \tag{13}$$

This argument can be easily generalized to $M_n(v, \ldots, v) = v$.

Consider $n = 2$. We have now shown the function M_2 is symmetric, strictly increasing, continuous, and satisfies (13) and (12). According to Aczél (1948), we know that there exist a stricly increasing continuous function $f : V(\mathcal{D}) \to \mathbb{R}$ such that $M_2(v_1, v_2) = f^{-1}(\frac{1}{2}f(v_1) + \frac{1}{2}f(v_2))$. Thus, for any $a = \{a_1, a_2\}$,

$$V(a) = f^{-1}\left(\frac{1}{2}f(V(a_1)) + \frac{1}{2}f(V(a_2))\right).$$

The idea of how to identify f for M_2 is described in Section 2. Since we also have $M_1(v) = v$, and hence $V(\{a\}) = f^{-1}(f(V(a)))$, equation (4) is true for $n = 1, 2$.

To generalize (4) to the case with $n > 2$, we first prove the following lemma.

Lemma 2 *For $a_i = \{a_{i,1}, \ldots, a_{i,m}\}$, $a = \{a_1, \ldots, d_n\}$ with $a_i \cap a_j = \emptyset$, $a \sim \bigcup_{i=1}^n a_i$.*

Proof of Lemma 2: For each $a_{i,j}$, we can find $a_{i,j}^{(t)} \in D_0$, $t = 2, \ldots, n \times m$ such that $a_{i,j} \sim a_{i,j}^{(t)}$. For simplicity, let us denote $a_{i,j}$ by $a_{i,j}^{(1)}$. By richness, we can make sure that none of the $a_{i,j}^{(t)}$ is the same as any other $a_{k,l}^{(s)}$. Define $a_i^{(t)} := \{a_{i,1}^{(t)}, \ldots, a_{i,m}^{(t)}\}$, and $a^{(t)} := \{a_1^{(t)}, \ldots, a_n^{(t)}\}$. By *Dominance**, $a^{(s)} \sim a^{(t)}$. Then, via *Stochastic Set Betweenness*, we have $\{a\} \sim \{a^{(1)}, \ldots, a^{(mn)}\}$. Since $a \sim \{a\}$ by *Consistency*, we have $a \sim \{a^{(1)}, \ldots, a^{(mn)}\} := b$.

Notice that by construction, any $a_i^{(t)}$ and $a_j^{(s)}$ can be swapped in b without changing the cardinality of any subtree. Since $|a^{(t)}| = |a^{(s)}|$, due to Lemma 1, we can perform swaps of any $a_i^{(t)}$ for any $a_j^{(s)}$ and end up with a new decision tree that is indifferent to b. In particular, we can swap for many times and obtain the following decision tree, $c = \{c_1, \ldots, c_{mn}\} \sim b$ where $c_{km+1} = \{a_{k+1}^{(1)}, \ldots, a_{k+1}^{(n)}\}$, $c_{km+2} = \{a_{k+1}^{(n+1)}, \ldots, a_{k+1}^{(2n)}\}, \ldots, c_{(k+1)m} = \{a_{k+1}^{((m-1)n+1)}, \ldots, a_{k+1}^{(mn)}\}$, $k = 0, \ldots, n-1$. Now by *Stochastic Set Betweenness*, $c_{km+i} \sim a_{k+1}$, $k = 0, \ldots, n-1$. Thus, by *Dominance* again, $b \sim \{a_1^{(1)}, \ldots, a_1^{(m)}, a_2^{(1)}, \ldots, a_2^{(m)}, \ldots, a_n^{(1)}, \ldots, a_n^{(m)}\} := d$.

We can apply Lemma 1 to d, and swap any $a_{i,j}^{(t)}$ for $a_{k,l}^{(s)}$ this time to obtain new decision trees that are indifferent to d. Again, after many swaps, we have tree $d' = \{\{a_{1,1}^{(1)}, \ldots, a_{1,1}^{(m)}\}, \{a_{1,2}^{(1)}, \ldots, a_{1,2}^{(m)}\}, \ldots, \{a_{n,m}^{(1)}, \ldots, a_{n,m}^{(m)}\}\}$. By *Stochastic Set Betweenness*, each $\{a_{i,j}^{(1)}, \ldots, a_{i,j}^{(m)}\} \sim a_{i,j}$. Hence, d' is indifferent to $\bigcup_{i=1}^n d_i$. ■

Now suppose (4) works for all $m < n$ for some $n > 2$. For any $d = \{d_1, \ldots, d_{n+1}\}$, let us find by richness distinct $x_1, \ldots, x_{n-1}$ such that none of them belongs to d and each of them is indifferent to d. By *Stochastic Set Betweennes*, $d \sim d \cup \{x_1, \ldots, x_{n-1}\}$. By Lemma 2, $d \cup \{x_1, \ldots, x_{n-1}\} \sim \{\{d_1, \ldots, d_n\}, \{d_{n+1}, x_1, \ldots, x_{n-1}\}\} := c$. Define $c_1 := \{d_1, \ldots, d_n\}$, and $c_2 := \{d_{n+1}, x_1, \ldots, x_{n-1}\}$. As $|c| = 2$ and $|c_i| = n$, noting that $V(x_i) = V(d) = V(c)$, we know that

$$\begin{aligned} f(V(c)) &= f(V(d)) \\ &= \frac{1}{2}f(V(c_1)) + \frac{1}{2}f(V(c_2)) \\ &= \frac{1}{2n}\left(\sum_{i=1}^{n+1} f(V(d_i)) + \sum_{i=1}^{n-1} f(V(x_i))\right) \\ &= \frac{1}{2n}\left(\sum_{i=1}^{n+1} f(V(d_i)) + (n-1)f(V(d))\right). \end{aligned}$$

Thus, we have

$$V(d) = f^{-1}\left(\frac{1}{n+1}\sum f(V(d_i))\right).$$

■

PROOF OF PROPOSITION 2: For sufficiency, we only show that $V^*(x)^\lambda = \alpha_1 U^*(x)^\chi$ and $f\left(\sqrt[\lambda]{\alpha_1 U(x_i)^\chi}\right) = \alpha_2 g(u) + \beta$ imply $V(a)^\lambda = \alpha_1 \times U(a)^\chi$. Suppose $a = \{x_1, \ldots, x_n\} \in D_1 \backslash D_0$. Then

$$\begin{aligned} f(V(a)) &= \frac{1}{n}\sum_i f(V(x_i)) = \frac{1}{n}\sum_i f\left(\sqrt[\lambda]{\alpha_1 U(x_i)^\chi}\right) \\ &= \beta + \frac{\alpha_2}{n}\sum_i g(U(x_i)) = \alpha_2 g(U(a)) + \beta. \end{aligned}$$

Since f and g are strictly increasing, it must be true that $V(a) = \sqrt[\lambda]{\alpha_1 U(a)^\chi}$. By an induction argument, this observation can be extended to D.

For necessity, if (U^*, χ, g) also represents P, since the Luce value is unique up to a scalar multiplication,

$$V(a)^\lambda = \alpha_1 \times U(a)^\chi$$

for all $a \in \mathcal{D}$, and $\alpha_1 > 0$. By definition, this implies that $V^*(x)^\lambda = \alpha_1 U^*(x)^\chi$.

As for f's uniqueness, consider now $x, y \in D_0$. Define $v_1 := V(x)$, $v_2 := V(y)$, and

$v_3 := V(\{x,y\})$, and similarly $u_1 := U(x)$, $u_2 := U(y)$ and $u_3 := U(\{x,y\})$. We have

$$f(v_3) = \frac{1}{2}f(v_1) + \frac{1}{2}f(v_2) \tag{14}$$

and

$$g(u_3) = \frac{1}{2}g(u_1) + \frac{1}{2}g(u_2).$$

Since we already have $V(a)^\lambda = \alpha_1 U(a)^\chi$, let us define $h(u) := f(\alpha_1 u^{\chi/\lambda})$. Now (14) becomes

$$h(u_3) = \frac{1}{2}h(u_1) + \frac{1}{2}h(u_2).$$

Thus

$$h^{-1}\left(\frac{1}{2}h(u_1) + \frac{1}{2}h(u_2)\right) = g^{-1}\left(\frac{1}{2}g(u_1) + \frac{1}{2}g(u_2)\right). \tag{15}$$

Define $t_1 := h(u_1)$ and $t_2 := h(u_2)$. (15) becomes

$$g \circ h^{-1}\left(\frac{1}{2}t_1 + \frac{1}{2}t_2\right) = \frac{1}{2}g \circ h^{-1}(t_1) + \frac{1}{2}g \circ h^{-1}(t_2).$$

Since u_1 and u_2 are arbitrary, by Jensen's inequality, it must be true that

$$g \circ h^{-1}(t) = \alpha_2' t + \beta'$$

and hence $g(u) = \alpha_2' h(u) + \beta_2'$. Since both f and g are strictly increasing, $\alpha_2' > 0$. Reorganizing the equation with $\alpha_2 := \frac{1}{\alpha_2'}$ and $\beta_2 := -\frac{\beta_2'}{\alpha_2'}$, we get

$$f(\sqrt[\lambda]{\alpha_1 u^\chi}) = \alpha_2 g(u) + \beta.$$

■

PROOF OF THEOREM 2: We first prove sufficiency. Suppose P_1 and P_2 can be represented by (V^*, λ, f_1) and (V^*, λ, f_2), respectively. Then, P_1 must coincide with P_2 on depth-1 decision trees according to (3). For any $x \in D_0$, $a = \{x_1, \ldots, x_n\} \in D_1$, let $v_i := V^*(x_i)$. Since $f_2 = g \circ f_1$,

$$\begin{aligned} f_2(V_2(a)) &= \frac{1}{n}\sum f_2(v_i) \\ g \circ f_1(V_2(a)) &= \frac{1}{n}\sum g \circ f_1(v_i). \end{aligned}$$

On the other hand, $f_1(V_1(a)) = \frac{1}{n}\sum f_1(v_i)$. By Jensen's inequality

$$\frac{1}{n}\sum g \circ f_1(v_i) \leq g\left(\frac{1}{n}\sum f_1(v_i)\right) = g(f_1(V_1(a))).$$

Therefore, $V_1(a) \geq V_2(a)$, and $a \succeq_2 x$ implies $a \succeq_1 x$. Now suppose we have proved that for some m, $a \succeq_2 x$ implies $a \succeq_1 x$ for any $x \in D_0$ and $a \in \bigcup_{i=1}^m D_i$. Note that $a \succeq_2 x \Rightarrow a \succeq_1 x$ for any $x \in D_0$ and $a \in \bigcup_{i=1}^m D_i$ implies that $V_1(a) \geq V_2(a)$ for $a \in \bigcup_{i=1}^m D_i$. Now consider $b = \{b_1, \ldots, b_n\} \in D_{m+1}$, by the induction hypothesis, we have $V_1(b_i) \geq V_2(b_i)$, and thus

$$\begin{aligned} V_1(b) &= f_1^{-1}\left(\frac{1}{n}\sum f_1(V_1(b_i))\right) \\ &\geq f_1^{-1}\left(\frac{1}{n}\sum f_1(V_2(b_i))\right) \\ &\geq f_2^{-1}\left(\frac{1}{n}\sum f_2(V_2(b_i))\right) \\ &= V_2(b). \end{aligned}$$

The second inequality is due to Jensen's inequality.

Next, we prove necessity. Since P_1 and P_2 coincide on depth-1 decision trees, and they are both Luce rules, according to Proposition 2, we can let they share the same λ, set $\alpha_1 = 1$ and find V_1 and V_2 such that $V_1(x) = V_2(x)$ for $x \in D_0$. Define $V^* := V_1^* = V_2^*$. Then, there exist (V^*, λ, f_1) and (V^*, λ, f_1) that represent P_1, P_2, respectively. Define $g := f_2 \circ f_1^{-1}$. The function g is clearly strictly increasing. We know that for any $x \in D_0$ and $a = \{x_1, \ldots, x_n\} \in D_1$, $a \succeq_2 x$ implies $a \succeq_1 x$, where we again let $v_i := V^*(x_i)$. In particular, by richness, we can find $y \in D_0$ such that $a \sim_2 y$; that is, $V_2(a) = f_2^{-1}(\frac{1}{n}\sum f_2(v_i)) = V^*(y)$, and $V_1(a) \geq V^*(y)$, which implies

$$\begin{aligned} f_1^{-1}\left(\frac{1}{n}\sum f_1(v_i)\right) &\geq f_2^{-1}\left(\frac{1}{n}\sum f_2(v_i)\right) \\ g\left(\frac{1}{n}\sum f_1(v_i)\right) &\geq \frac{1}{n}\sum f_2(v_i). \end{aligned}$$

Define $t_i := f_1(u_i)$. The inequality above becomes $\frac{1}{n}\sum g(t_i) \leq g(\frac{1}{n}\sum t_i)$, which implies that g is concave.

■

PROOF OF PROPOSITION 3: Suppose a rich CCA BBR is homogeneous. From

$$\frac{V(x)^\lambda}{V(x)^\lambda + V(w)^\lambda} \geq \frac{V(y)^\lambda}{V(y)^\lambda + V(z)^\lambda}$$

we know that

$$V(x)/V(w) \geq V(y)/V(z).$$

Let $a := \{w, x\}$ and $b := \{y, z\}$. By definition, $V(a) = (\frac{1}{2}[V(x)]^\gamma + \frac{1}{2}[V(w)]^\gamma)^{1/\gamma}$ and $V(b) = (\frac{1}{2}[V(y)]^\gamma + \frac{1}{2}[V(z)]^\gamma)^{1/\gamma}$. Therefore

$$\begin{aligned} V(a) &= V(x)\left(\frac{1}{2} + \frac{1}{2}\left(\frac{V(w)}{V(x)}\right)^\gamma\right)^{1/\gamma} \\ &\leq V(x)\left(\frac{1}{2} + \frac{1}{2}\left(\frac{V(z)}{V(y)}\right)^\gamma\right)^{1/\gamma} \\ &= \frac{V(x)}{V(y)}V(b) \end{aligned}$$

which implies that

$$\frac{V(x)^\lambda}{V(x)^\lambda + V(a)^\lambda} \geq \frac{V(y)^\lambda}{V(y)^\lambda + V(b)^\lambda}.$$

To show necessity, note that if $P(\{x\}, a) = P(\{y\}, b)$, our condition implies that $P(\{x\}, \{x, a\}) =$

$P(\{y\}, \{y, b\})$. Since

$$\frac{V(x)^\lambda}{V(x)^\lambda + V(w)^\lambda} = \frac{V(y)^\lambda}{V(y)^\lambda + V(z)^\lambda}$$

there exists a α such that $V(x) = \alpha V(y)$ and $V(w) = \alpha V(z)$. Since

$$\frac{V(x)^\lambda}{V(x)^\lambda + V(a)^\lambda} = \frac{V(y)^\lambda}{V(y)^\lambda + V(b)^\lambda}$$

we know that $V(a) = \alpha V(b)$ too. To summarize, we have $V(x) = \alpha V(y)$ and $V(w) = \alpha V(z)$ implying

$$\begin{aligned} \alpha f^{-1}\left(\frac{1}{2}f(V(y)) + \frac{1}{2}f(V(z))\right) &= f^{-1}\left(\frac{1}{2}f(V(w)) + \frac{1}{2}f(V(x))\right) \\ &= f^{-1}\left(\frac{1}{2}f(\alpha V(y)) + \frac{1}{2}f(\alpha V(z))\right). \end{aligned}$$

By richness, the above arguments work for arbitrary α, $V(y)$ and $V(z)$. Therefore, f must be homogeneous of degree 1, and take the form $f(v) = \beta v^\gamma$ (see Wnuk (1984)).

■

Proof of Theorem 3: First, we show sufficiency. For any $\{a, b\} \in D$ such that $V_2(a) \leq V_2(b)$, we can define $h(P_2(\{a\}, \{a, b\}))$ to be $P_1(\{a\}, \{a, b\})$. It is clear that $P_1(\{a\}, \{a, b\}) \leq P_2(\{a\}, \{a, b\})$, since $\lambda_1 \geq \lambda_2$. Therefore, $h(p) \leq p$ for $p \in (0, \frac{1}{2}]$. When $V_2(a) = V_2(b)$, we can think of it as $V_2(a) \geq V_2(b)$ and the other way around. Hence, $h(1/2) = 1/2$. The only thing we need to check is that h is well-defined; that is, for $P_2(\{a\}, \{a, b\}) = P_2(\{c\}, \{c, d\})$, we have $P_1(\{a\}, \{a, b\}) = P_1(\{c\}, \{c, d\})$ as well. First note that since P_1 and P_2 share the same V^* and f, they share the same V. Then,

$$\frac{V(a)^{\lambda_2}}{V(a)^{\lambda_2} + V(b)^{\lambda_2}} = \frac{V(c)^{\lambda_2}}{V(c)^{\lambda_2} + V(d)^{\lambda_2}}$$

implies that $V(a)/V(b) = V(c)/V(d)$ and hence

$$\frac{V(a)^{\lambda_1}}{V(a)^{\lambda_1} + V(b)^{\lambda_1}} = \frac{V(c)^{\lambda_1}}{V(c)^{\lambda_1} + V(d)^{\lambda_1}}.$$

Thus, $P_1(\{a\}, \{a, b\}) = P_1(\{c\}, \{c, d\})$.

Consider necessity. Suppose P_i is represented by (V_i^*, λ_i, f_i). Since P_2 is an RCR, $P_2(\{a\}, \{a, b\}) \leq \frac{1}{2}$ if and only if $V_2(a) \leq V_2(b)$, $a, b \in \mathcal{D}$. Since $h(p) \leq p$, we know that if $V_2(a) < V_2(b)$, then $V_1(a) < V_1(b)$. Because $h(\frac{1}{2}) = \frac{1}{2}$, we know that $V_2(a) = V_2(b)$ implies that $V_1(a) = V_1(b)$. Thus, there is a strictly increasing function ϕ such that $V_1(a) = \phi(V_2(a))$, $a \in \mathcal{D}$. Now for any a, b such that $V_2(a) \leq V_2(b)$, by richness, we can find $a_\alpha, b_\alpha \in D_0$ such that $V_2(a_\alpha) = \alpha V_2(a)$ and $V_2(b_\alpha) = \alpha V_2(b)$. Notice that

$$P_2(\{a\}, \{a, b\}) = P_2(\{a_\alpha\}, \{a_\alpha, b_\alpha\}).$$

We must have

$$P_1(\{a\}, \{a, b\}) = P_1(\{a_\alpha\}, \{a_\alpha, b_\alpha\})$$

which implies that $V_1(a_\alpha) = \psi_a(\alpha)V_1(a)$, $V_2(b_\alpha) = \psi_b(\alpha)V_1(b)$ and $\psi_a(\alpha) = \psi_b(\alpha)$. Since a and b are arbitrary, there has to be a $\psi(\alpha) = \psi_a(\alpha)$ for any $a \in \mathcal{D}$. Thus

$$\begin{aligned} \phi(V_2(a_\alpha)) &= \phi(\alpha V_2(a)) \\ V_1(a_\alpha) &= \psi(\alpha)V_1(a) \\ &= \psi(\alpha)\phi(V_2(a)). \end{aligned}$$

Therefore, we have

$$\phi(\alpha v) = \psi(\alpha)\phi(v).$$

To satisfy the equation above, according to Aczél (1966, p. 144–145), $\phi(v) = \alpha_1 v^\lambda$; that is, $V_1(a) = \alpha_1 V_2(a)^\lambda$ for all $a \in \mathcal{D}$. Due to Proposition 2, we can pick a V_1 and λ_1 such that

$V_1 = V_2 := V$. Of course, $V^* := V_2^*$. Now P_i is represented by (V^*, λ_i, f_i). Since $h(p) \leq p$, we know that $\lambda_1 \geq \lambda_2$. Since we have picked V_1 such that $V_1 = V_2$, it must be true that $f_1 = f_2$.

■

References

[1] Aczél, J. (1948): "On Mean Values," *Bulletin of the American Mathematical Society*, 54(4), 392–400.

[2] Aczél, J. (1966): *Lectures on Functional Equations and Their Applications.* NewYork: Academic Press, pp. 144–145.

[3] Binmore, K., J. McCarthy, G. Ponti, L. Samuelson, and A. Shaked (2002): "A Backward Induction Experiment," *Journal of Economic Theory*, 104(1), 48–88.

[4] Boatwright, P., and J. Nunes (2001): "Reducing Assortment: An Attribute-Based Approach," *Journal of Marketing*, 65(3), 50–63.

[5] Bolker, E. (1966): "Functions Resembling Quotients of Measures," *Transactions of the American Mathematical Society*, 124(2), 292–312.

[6] Broniarczyk, S., W. Hoyer, and L. McAlister (1998): "Consumers' Perceptions of the Assortment Offered in a Grocery Category: The Impact of Item Reduction," *Journal of Marketing Research*, 166–176.

[7] Camerer, C., E. Johnson, T. Rymon, and S. Sen (1993): "Cognition and Framing in Sequential Bargaining for Gains and Losses," *Frontiers of Game Theory*, 104, 27–47.

[8] Debreu, G. (1960): "Topological Methods in Cardinal Utility Theory," *Mathematical Methods in the Social Sciences*, 1959, 16–26.

[9] DellaVigna, S. (2009): "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 47(2), 315–372.

[10] Fernandez, R., and Rodrik, D. (1991): "Resistance to Reform: Status Quo Bias in the Presence of Individual-Specific Uncertainty," *American Economic Review*, 81(5), 1146–1155.

[11] Fudenberg, D., and T. Strzalecki (2015): "Dynamic Logit with Choice Aversion," *Econometrica*, 83(2), 651–691.

[12] Gabaix, X., and D. Laibson (2005): "Bounded Rationality and Directed Cognition," Mimeo, New York University.

[13] Gul, F., P. Natenzon, and W. Pesendorfer (2014): "Random Choice as Behavioral Optimization," *Econometrica*, 82(5), 1873–1912.

[14] Gul, F., and W. Pesendorfer (2001): "Temptation and Self-Control," *Econometrica*, 82(5), 1873–1912.

[15] Iyengar, S., and E. Kamenica (2010): "Choice Proliferation, Simplicity Seeking, and Asset Allocation," *Journal of Public Economics*, 94(7), 530–539.

[16] Jéhiel, P. (1995): "Limited Horizon Forecast in Repeated Alternate Games," *Journal of Economic Theory*, 67(2), 497–519.

[17] Kahneman, D., J. Knetsch, and R. Thaler (1991): "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias," *Journal of Economic Perspectives*, 5(1), 193–206.

[18] Laibson, D. (1997): "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 112(2), 443–477.

[19] Luce, R. (1959): *Individual Choice Behavior: A Theoretical Analysis*. New York: John Wiley & Sons.

[20] Masatlioglu, Y., and E. Ok (2005): "Rational Choice with Status Quo Bias," *Journal of Economic Theory*, 121(1), 1–29.

[21] McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in Paul Zarembka, ed., *Frontiers in Econometrics*. New York: Academic Press, pp. 105–142.

[22] McKelvey, R., and T. Palfrey (1995): "Quantal Response Equilibria for Normal Form Games," *Games and Economic Behavior*, 10(1), 6–38.

[23] McKelvey, R., and T. Palfrey (1998): "Quantal Response Equilibria for Extensive Form Games," *Experimental Economics*, 1(1), 9–41.

[24] Machina, M. (1987): "Choice under Uncertainty: Problems Solved and Unsolved," *Journal of Economic Perspectives*, 50(1), 121–154.

[25] Pearl, J. (1984): *Heuristics*. Reading, Massachusetts: Addison-Wesley Publishing Company.

[26] Rubinstein, A. (1990): "New Directions in Economic Theory - Bounded Rationality," *Revista Espanola De Economie*, 7, 3–15.

[27] Russell, S., and P. Norvig (1995): *Artificial Intelligence: A Modern Approach*. New York: Prentice-Hall, pp. 92–107.

[28] Samuelson, W., and R. Zeckhauser (1988): "Status Quo Bias in Decision Making," *Journal of Risk and Uncertainty*, 1(1), 7–59.

[29] Simonson, I. (1999): "The Effect of Product Assortment on Buyer Preferences," *Journal of Retailing*, 75(3), 347–370.

[30] Strotz, R. (1955): "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies*, 23(3), 165–180.

[31] Tversky, A., and D. Kahneman (1981): "The Framing of Decisions and the Psychology of Choice," *Science*, 211(4481), 453–458.

[32] Wnuk, W. (1984): "Orlicz Spaces Cannot Be Normed Analogously to L^p-Spaces," *Indagationes Mathematicae (Proceedings)*, 87(3), 357–359.