Essays on Decision Theory

by

Jiayun Dong

Department of Economics
Duke University

Date: ______________________
Approved:

____
Todd Sarver, Advisor

____
David McAdams

____
Philipp Sadowski

____
Curtis Taylor

____

Dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Department of Economics
in the Graduate School of
Duke University

2020

ABSTRACT

Essays on Decision Theory

by

Jiayun Dong

Department of Economics
Duke University

Date: ______________________

Approved:

____

Todd Sarver, Advisor

____

David McAdams

____

Philipp Sadowski

____

Curtis Taylor

____

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics in the Graduate School of Duke University

2020

Abstract

Decision problems could involve comparisons of alternatives according to different aspects. An alternative may be more desirable than another alternative in some of the aspects, yet less desirable in others. The multi-dimensional comparisons between alternatives requires a high degree of rationality – the ability to make systematic trade-offs across aspects.

Limited ability to make such trade-offs could lead to the default bias, the tendency of choosing the default regardless of the presence of a better one. In dynamic decision problems, default bias causes a novel trade-off between the current consumption and the flexibility to switch in the future. An asset replacement problem and a contract design problem involving decision makers with default bias are studied.

A decision maker may resort to market information to better understand the right trade-off to make in multi-aspects comparisons between alternatives. When the market composition of products that features different aspects carry relevant information regarding the trade-offs, the decision makers choices between two products could be affected by the presence of other alternatives. Learning from markets explains violations of the properties of classic random utility maximization models.

Normatively appealing decision rules are proposed for each of the scenarios of multi-dimensional comparisons. A choice rule imposes behavioral conditions that the choice data must satisfy if the data is generated by such a rule. The essays in the dissertation identify the behavioral conditions that characterize the proposed choice rules: When the decision makers behaviors satisfy those conditions, the behaviors can be interpreted as if they are made according to the choice rule.

The dissertation distinguishes itself from some of the existing literature in that the set of attributes that a decision maker considers in her comparisons is not assumed to

be observable. Instead, they can be uniquely identified from the choice behaviors. It extends the current understanding of some behavioral anomalies to a richer domain and proposes novel explanations of them.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Well-defined preferences[1] of economic agents are assumed and taken as primitive of economic analysis in almost all economic literature. By assuming well-defined preferences, researchers have implicitly assumed that the economic agent is always able to compare the desirability of two alternatives in a systematic and stable manner.

However, one may argue with real life examples that such comparisons are not always easy to make. Moreover, they are sometimes inconsistent and contradicting with each other. This is more likely to happen when one evaluate alternatives according to multiple dimensions, such as price and quality. Such observations violate a fundamental defining property of preferences, completeness, that a decision maker is always able to form a well-defined ranking between alternatives. The essays of the dissertation pertain to decision environments in which there are potentially violations of completeness and a decision maker may feel it difficult to compare alternatives but have to find a way to make the comparisons. We review some of the related works before proceeding to the essays.

1.1 Preference Reversals

Lichtenstein and Slovic (1971) is among the first works that document observations of preference reversals in experiments. They offer a lottery ("P-bet") that offers

[1]Or utility function in many applied analysis

a moderate amount of prize with relatively larger probability and another lottery ("\$-bet") that offers a large amount of prize with relatively small probability. Two different ways are used to elicit the preference of a decision maker. The direct way is to ask the decision maker about which lottery is preferred. The indirect way is to ask the decision maker about the amount of money she is willing to take instead of the lottery that makes her equally satisfied, referred to as the certainty equivalent of the lottery. The comparison between the certainty equivalents should be consistent with the comparison between the lotteries themselves if there is a well-defined preference. Lab experiments reveal inconsistent results using these two different ways to elicit preferences.

Among many other explanations, Slovic and Lichtenstein (1983) attributes the phenomenon to the different ways of evaluating the alternatives adopted by the decision maker under the direct and indirect ways of eliciting preferences, suggesting that the ranking of desirability among alternatives are not invariant of the procedure to elicit them, which undermines the validity of modeling a decision maker's taste with well-defined preferences.

Preference reversals are not limited to comparing prospects with uncertainty. In deterministic environments, a decision maker may still demonstrate inconsistencies in her choices. In decision theory, a common way to elicit preferences is to use the following revealed preferences argument: If a decision maker would like to choose an alternative instead of other available ones, she must prefer it to other available alternatives. The existence of a well-defined preference imply that if one alternative is chosen when another alternative is present in the choice set, then the latter cannot be chosen in any choice set whenever the former is present. This conjecture is called the weak axiom of revealed preference.

While being an appealing property, this axiom is often violated in reality. People have been long documenting the attraction effect, also known as the asymmetric dominance effect, in marketing literature, including Huber et al. (1982) and Huber and Puto (1983). Consider the following scenario as an illustrating example. Suppose there are one high-quality-high-price product and a low-quality-low-price product in the market, introducing the third product with the lowest quality but median price would improve the sales of the low-quality-low-price product. Notice that the third product is definitely worse than the low-quality-low-price product in both the quality and price dimensions, but not immediately comparable with the high-quality-high-price product. Researchers conjecture and test that the asymmetric dominance makes the dominant product more attractive.

This observation again undermines the validity of assuming a well-defined preference, because the sales in the two-product market and the three-product market reveals different preferences of at least some decision makers in the market. Again, this suggests that we might elicit different preferences under different context, meaning that eliciting preferences is not procedure invariant in the deterministic setting either.

While preference reversals have seen explained by multiple researchers, we focus on reviewing the models that argue that decision maker evaluate alternatives according to multiple dimensions to form their preferences since they are closely related to the essays of the dissertation. For other explanations, one can refer to Loomes and Sugden (1983), Tversky et al. (1990), Tversky and Thaler (1990) and Seidl (2002), among others.

1.2 Multi-dimensional Evaluation of Alternatives

The two decision problems in which the elicited preferences could vary in the ways to elicit them share a common feature: It is intrinsically difficult for the decision maker to compare the alternatives without making sophisticated calculations because of the dimension of the problems.

In the first problem, the $-bet is appealing because its prize is larger but the P-bet if appealing because its winning probability is larger. The problem would otherwise be much simpler if one lottery is better than another both in terms of the prize and winning probability. Similarly, the problem would be straightforward if one needs to compare a high-quality-low-price product with a low-quality-high-price one. The difficulty arises exactly because there are multiple dimensions to consider in the comparison and neither alternative dominates the other one in all of the dimensions.

In multi-dimensional comparisons, even a decision maker is able to compare alternatives in each individual dimension, it is still not clear what the overall evaluation would be. There are two common approaches to this problem. One can assume a reasonable aggregation rule to make alternatives comparable, suggesting that the preference is a complete order. Instead, one can also admit the fundamental difficulties in comparing alternatives and have a weaker notion of preferences that could be incomplete. Cailloux and Destercke (2017) reviews some of the popular models. In what follows we describe some models that are related to the essays in the dissertations.

1.2.1 Incomplete Preferences

As we have mentioned above, researchers sometimes adopt a weaker notion of preferences and give up completeness when there are multiple dimensions to consider in the decision making process.

In experimental settings, it is documented that people's behaviors exhibit the status quo bias, meaning that they have a tendency of staying in the status quo instead of switching to something else even when there is a potentially better alternative. Samuelson and Zeckhauser (1988) and Kahneman et al. (1991) have summarized both empirical and experimental evidence. This could be modeled as multi-dimensional considerations: The decision maker is only willing to opt out of the statue quo when the alternative dominates the status quo (Masatlioglu and Ok (2005)). The preferences elicited from such behaviors could be incomplete. It is possible that in the comparison between two alternatives, none of them is revealed to be superior to the other when it is not the status quo.

Such treatment of incomplete preferences are also analyzed by Bewley (2002) and Dubra et al. (2004). The former paper models beliefs as dimensions. A decision maker faces multi-dimensional evaluations in ambiguous decision problems when she have multiple beliefs. The latter one models linear incomplete preferences that relax the classic expected utility theory.

One of our essays extend the idea of multi-dimensional considerations and incomplete preferences to dynamic decision problems. In dynamic problems, the novel trade-off arises as today's choice becomes tomorrow's status quo. Since the status quo may limit the scope of choices in the sense that only those that dominate the status quo would be considered as appealing, it may not be optimal for the decision maker to choose the alternative that maximizes the current utility without having the future choices in mind.

1.2.2 Dimension Aggregation

On the other hand, there are scenarios in which people are assumed to be able to make comparisons about the desirability of alternatives, despite the intrinsic difficulties in making multi-dimensional comparisons. In those scenarios, preferences are assumed as complete preference relations and they can be represented using a utility function. This is particularly common when the dimensions are the states of the world.

Imagine that we have K dimensions and that an alternative x is represented by the vector $\{v(x,i)\}_{i=1}^K$, where $v(x,i)$ is the value of x in dimension i. This can be interpreted as the subjective evaluation of the outcome of x in state i. We may anticipate that x is preferred to y if $v(x,i) > v(y,i)$ for all $i = 1, c \ldots, K$. Note that this holds in our example of the attraction effect and this is exactly the idea underlying models of incomplete preferences.

Assuming a complete preference relation implies that we also have a function u such that $u(x) > u(y)$ if and only if x is preferred to y. This implies that we have a "bridge" between evaluating alternatives using v's and the u's. In particular, there exists an aggregator $h : \mathbf{R}^K \to \mathbf{R}$ so that $u(x) = h(v(x,1), \cdots, v(x,K))$. To preserve the consistency between u's and v's, the aggregaror h is monotone.

The aggregator h can take many forms. The aggregator with the mildest condition might just be taking summations. that is, $h(v(x,1), \cdots, v(x,K)) = \sum_{i=1}^K v(x,i)$. This aggregator is proposed by Kreps (1979). This could be interpreted as the states having equal weights and the decision maker using the expected utility to evaluate alternatives. The limitation of Kreps' model is that he has imposed conditions that are too weak such that he cannot identify beliefs separately from the evaluations in each dimension.

Dekel et al. (2001) and the subsequent paper Dekel et al. (2007) enrich the primitive of the model and is able to identify the belief. They have characterized an additive EU representation in which they can identify the subjective belief of the decision maker about the states of the world separately from the evaluation of the alternatives in each dimension. That is, there is a probability belief q defined on the states with $q_i \geq 0$ and $\sum_{i=1}^{K} q_i = 1$ such that $h(v(x,1), \cdots, v(x,K)) = \sum_{i=1}^{K} q(i)v(x,i)$.

While the dimensions are interpreted as the states of the world and the function q is interpreted as a subjective belief above, in deterministic choice problems one can alternatively interpret the dimensions as the criteria that the decision maker has in mind and the function q as some subjective decision weights assigned to the criteria. One of our essays is based exactly on this idea and conjectures that the decision maker needs to learn about the best way to aggregate the dimensions based on the set of alternatives she is choosing from.

Chapter 2

Default Bias in Dynamic Decision Problems

When a menu has a default option, the default may affect the choices from the menu in various ways. In this paper, we focus on the limited ability and the limited willingness to switch from the default option, and refer to them collectively as the default bias. The first channel may arise when an individual may not be able to make an "active decision" from the menu with some probability and has to accept the default; the second channel may arise if an individual evaluates alternatives using multiple criteria and is only willing to switch from the default to alternatives that outperform it in all of these criteria. We construct the domain of infinite horizon lotteries and characterize the Default Bias Representation. In the dynamic domain, the choice made today is directly connected with the default option of the menu for tomorrow. This connection leads to the trade-off between today's consumption and the default option which may constrain future consumption. We demonstrate the usefulness of our model with two stylized applications: The first examines the impact of default bias on asset replacement decisions, and the second demonstrates how a firm may wish to use promotional offers when contracting with consumers with default bias.

2.1 Introduction

A choice problem may have a default option. While an individual may need to take an explicit action to choose a non-default option, consuming the default option does not require any action. Classic economic theory asserts that an individual simply chooses the most preferable option from a choice set, suggesting that the default option should not matter. However, there is empirical and experimental evidence that default options do affect choice behavior. For example, Carroll et al. (2009) find that the 401(k) enrollment rate is higher in the regime where employees are required to explicitly make the enrollment decision (no default) than in the alternative regime where they are by default not enrolled and can choose to opt in. Johnson and Goldstein (2003) examine the effect of policy defaults on organ donation in an experiment. They find that donation rates are higher when people are by default donors unless they opt out than when they are by default non-donors unless they opt in.

Default options may affect choices in various ways. This paper focuses on two potential channels. The first channel, which is called the *limited ability to switch*, is that an individual might not be able to make an "active decision" whenever she wants. In other words, with some probability the individual cannot switch to a non-default option and hence has to accept the default one, and this probability is independent of her preference over these options. This channel may arise due to physical constraints or cognitive biases. For example, an individual could be occupied by some event so that she becomes too busy to switch, or she could simply forget to switch. One can alternatively motivate this effect through a random switching cost, whose value is either negligible or prohibitively large.

The second channel, referred to as the *limited willingness to switch*, is related to how an individual evaluates the non-default options in comparison with the default one. She may have multiple evaluation criteria, and is willing to switch to a non-default option only when it dominates the default one according to all the criteria. To motivate this effect, imagine that the individual evaluates the changes in several aspects that arises from switching from the default to another option. If the individual is loss averse, she pays more attention to the aspects in which she is worse off than the other aspects in which she is better off. As she becomes extremely loss averse, her switching behaviors exhibit the limited willingness to switch. This interpretation reveals a connection between this channel and the status quo bias discussed by Kahneman et al. (1991), which is also explained by loss aversion.

We analyze the behavioral implications of default bias in dynamic decision problems. Unlike the standard dynamic choice literature, where a choice made today determines a menu for tomorrow, in this paper the choice made today also determines the default option associated with tomorrow's menu. In other words, the individual takes how the default option will affect tomorrow's choice from the menu into consideration when she makes her choice today. Linking today's consumption choices to tomorrow's default option introduces a novel economic trade-off. Imagine an individual is deciding whether to replace an asset today. The decision to keep the asset today determines that the default option for tomorrow is the depreciated asset. Anticipating default bias, she has to take into consideration the possibility that she cannot replace the depreciated asset regardless of her preference tomorrow. We will discuss this problem in the application section of the paper in order to highlight this trade-off and demonstrate the usefulness of the dynamic approach.

To study the intertemporal trade-offs induced by default bias, we need a richer domain of preferences: In this paper, we consider the set of lotteries that randomly yield a *consumption level* c for the current period and a continuation *decision prob-*

lem (z, r), which consists of a *menu* of lotteries z to choose from, with the associated *default lottery* r. Notice that the lotteries in the menu z will also yield a consumption level and a new decision problem for the future. Consequently, the domain of preferences is recursive. We extend the method in Gul and Pesendorfer (2004) and prove the existence of this recursive domain.

We model a default biased individual as if she is maximizing the following value function, referred to as the *Default Bias Representation (DB representation)*, and the main result of the paper characterizes this representation.

$$V(p) = \int \left(u(c) + \delta \max_{q \in z:\ F(q) \geq F(r),\ \forall F \in \mathcal{F}} [\mu V(q) + (1-\mu)V(r)] \right) \mathrm{d}p(c, z, r).$$

The DB representation suggests that the value of a lottery p is the expected sum of the utility from today's consumption level c and the discounted utility from tomorrow's decision problem (z, r). Default bias affects tomorrow's problem through the two channels argued above, namely the limited ability and willingness to switch. We argued that an individual who is affected by the limited ability to switch is able to make an active choice in a decision problem with some probability. This is modeled in the DB representation by the convex combination of the values $V(q)$ and $V(r)$. That is, with probability μ a lottery q of the menu z (which may or may not be the default) is chosen and with probability $1-\mu$ the default r must be chosen since there is no active decision. Even in the former situation, the lottery q may not be the best lottery of the menu z. Recall that the limited willingness to switch arises when an individual is willing to switch to a non-default option only when it dominates the default option according to all aspects. The set of options that she is willing to switch to is modeled in the DB representation by the sub-menu $\{q \in z : \ F(q) \geq F(r),\ \forall F \in \mathcal{F}\}$. The sub-menu consists of the lotteries in the menu z that are better than the default r

according to all the *criteria* $F \in \mathcal{F}$. When the individual is able to switch, the DB representation predicts that she switches to the best option among those which she is willing to switch to.

The DB representation is characterized by the tuple of the parameters $(u, \delta, \mu, \mathcal{F})$, among which μ is related to the ability to switch and $\mathcal{F}$ is related to the willingness to switch. Naturally when μ takes a larger value, the individual has greater ability to switch and when $\mathcal{F}$ becomes smaller, the individual has greater willingness to switch. We will provide comparative statics results that identify the behaviors that can be used to compare the degrees to which individuals are affected by the limited ability and willingness to switch, and connect these behavioral comparisons to the values of the parameters.

It is mentioned already that limited willingness to switch is related to the status quo bias, which has been well-studied in the literature; limited ability to switch, however, is a novel effect that has been less explored. To demonstrate the importance of this feature of our model, we consider two stylized applications in this paper. The first one is an asset replacement problem, which is motivated by Rust (1987). Suppose an individual holds an asset that depreciates throughout time. She has to decide when to replace it with a new one. She might suffer from a limited ability to switch and be unable to replace the asset in each period (in which case the depreciated asset is the default option). We show that anticipating her limited ability to switch, she may make the preemptive replacement decision, which is earlier than the optimal time if her switching ability were not limited. Arcidiacono et al. (2016) use the continuous analog of this model to estimate dynamic games and argue that it can address the computation problem associated with the curse of dimensionality in the dynamic discrete choice literature. We would argue, however, their method may introduce estimation bias, due to the preemptive replacement incentive.

In the second application, we investigate whether a firm is able to exploit consumers who suffer from limited ability to switch. The firm that uses an introductory promotion to attract new users needs to decide between automatic and explicit membership renewal after the trial period. With the automatic renewal policy, while the firm can exploit the consumer who does not like the product and forgets to quit, the consumer who anticipates this may require a larger promotion in the first place. On the other hand, with the explicit renewal policy, the consumer who likes the product may forget to subscribe after the trial period expires. The conditions under which a firm prefers each policy are analyzed.

Default bias, especially limited willingness to switch, is closely related to the status quo bias. We have interpreted this channel of the bias as the result of extreme loss aversion, which explains many well-documented empirical and experimental facts about the status quo bias. (See, for example, Samuelson and Zeckhauser (1988) and Kahneman et al. (1991)). Mathematically, the individual is modeled as if she has multiple criteria in mind, and she is only willing to switch to the options that dominate the default one according to all criteria. Masatlioglu and Ok (2005) models the status quo bias in static decision problems in the same way.

Both their work and ours are related to the literature on incomplete preferences. Given the switching criteria, one can find two lotteries p and q with neither of them dominating the other. As a result, an individual is not willing to switch to p when q is the default, and vice versa. In this sense, the switching behavior is incomplete. There are multiple ways to model the incomplete switching behavior (or in general incomplete preferences). Bewley (2002) models incomplete preferences as a result of multiple beliefs: A lottery is preferred to another if its expected value is higher for all of the beliefs. Dubra et al. (2004) propose the expected multi-utility representation that relaxes the completeness requirement in the expected utility theorem. We follows the latter method to model limited willingness to switch in the paper.

While our model is close to Masatlioglu and Ok (2005) in terms of the incomplete switching behaviors, by working with the dynamic domain, we can further compare the values of decision problems: Even when an individual is willing to choose the same non-default option from the same menu, paired with different defaults, she might still prefer one decision problem to the other (because the default options are different and she may have to choose the default when she is unable to make an active choice). This comparison is made based on preferences from the *ex ante* perspective, which is possible here in the dynamic domain and is not the focus of Masatlioglu and Ok (2005). Moreover, as we have argued before, the dynamic trade-off between today's consumption and tomorrow's default option merges in our framework, which is again not their focus.

This paper is also related to the literature about consideration sets, including Masatlioglu et al. (2012), and Manzini and Mariotti (2014). These models typically assume that individuals pay attention only to a subset of the choice set, called the consideration set. The first paper analyzes deterministic consideration sets, while the second has random consideration. In the current paper, the limited willingness to switch leads to a sub-menu which can be interpreted as a consideration set. Moreover, as the result of the limited ability to switch, the consideration set is stochastic: It is either the sub-menu or the singleton menu containing the default option only.

Finally, the idea that an individual makes active choices only in the periods when the opportunities arrive is related to a broader literature. Staggered prices models (for example, Calvo (1983)) in macroeconomcics assume that price revisions can happen only when "a random signal is lit-up". Arcidiacono et al. (2016) uses the random arrival of the "decision opportunities" to alleviate the computational issue in estimating dynamic games. Ambrus and Ishii (2015) show that coordination games have a unique prediction when players can only change their actions at "idiosyncratic random times" and they are "locked into the previously selected actions" in between. Our model characterizes the behavior foundation for these types of models.

The remaining sections are organized as follows. Sections 2 describes the domain of preferences. Section 3 characterizes the utility representation. Section 4 makes behavioral comparisons and connects them to the parameters of the DB representation. Section 5 contains two applications regarding the limited ability to switch, namely asset replacement and consumer exploitation.

2.2 Domain of Preferences

Given a compact metric space X, let $\Delta(X)$ be the set of all Borel probability measures on X, and $\mathcal{K}(X)$ be the collections of all non-empty compact subsets of X. Endowed with the weak topology, the set $\Delta(X)$ is compact and metrizable. Endowed with the Hausdorff metric, the set $\mathcal{K}(X)$ is also a compact metric space.[1]

Let C be a compact set of all possible levels of consumption in each period. For a given consumption space C, we will construct the space of *infinite-horizon lotteries with defaults* (referred to as lotteries henceforth) from which the individual chooses. This space of lotteries is denoted by Π. Then the set $\mathcal{K}(\Pi)$ is interpreted as the collection of all possible menus of lotteries. A lottery randomly yields a consumption level $c \in C$ and a decision problem for the next period. Specifically, the decision problem is a menu with a default option, denoted by (z, r), where $z \in \mathcal{K}(\Pi)$ is a set of lotteries and $r \in z$ is one of the lotteries in the menu interpreted as the default option. Given a realization (c, z, r) of the lottery, an individual consumes c in the current period, and is going to solve the decision problem (z, r) in the next period

[1] See Gul and Pesendorfer (2004) for references.

by making a choice from the menu z. The choice might be affected by the default option r.

The description above suggests that the support of a lottery is a subset of $D = \{(c, z, r) \in C \times \mathcal{K}(\Pi) \times \Pi : r \in z\}$. The restriction that $r \in z$ captures the fact that the default option must be an option in the menu. It turns out that one can construct the domain so that the space of lotteries, Π, can be identified with the set of all Borel probability measures on D, as is stated formally in Theorem 1.

Theorem 1. *The set* Π *is well-defined and compact. Moreover, there is a homeomorphism* $h : \Pi \to \Delta(D)$, *where* $D = \{(c, z, r) \in C \times \mathcal{K}(\Pi) \times \Pi : r \in z\}$.

The formal definition of Π, together with the proof of Theorem 1, is given in Appendix 2.6. The domain constructed in Theorem 1 is related to the ones in Kreps and Porteus (1978), Epstein and Zin (1989), and Gul and Pesendorfer (2004). This construction implicitly assumes that both the individual and the analyst observe the default lottery of a menu. Since the defaults may affect choices, observing the default of a menu provides information to the analyst even though she does not know how the default subjectively affects the individual in the decision making process. Similar assumptions are common in the literature. For instance, Masatlioglu and Ok (2005) assume the analyst observes the status quo of an individual when she is choosing from a menu, and Masatlioglu and Nakajima (2013) assume that the analyst observes the starting point of the individual's subjective search process when she makes a choice from the menu.

In what follows, lotteries are denoted by p, q, r, s and menus are denoted by z, z'. A degenerate lottery that yields with certainty a consumption level c and a menu z with its default option r is denoted by (c, z, r). There are no special symbols reserved for the defaults of the menus. Instead, since they are also lotteries, they are also denoted by p, q, r, s.[2]

For any two lotteries $p, q \in \Pi$ and $\lambda \in [0,1]$, let the convex combination $\lambda p + (1-\lambda)q$ be the lottery in Π that assigns probability $\lambda p(E) + (1-\lambda)q(E)$ to any Borel subset $E \subset D$. For any two menus, $z, z' \in \mathcal{K}(\Pi)$ and $\lambda \in [0,1]$, let the convex combination $\lambda z + (1-\lambda)z'$ be the set of element-wise mixtures, namely $\{\lambda p + (1-\lambda)p' : p \in z \text{ and } p' \in z'\}$.

The primitive of the model is a binary relation $\succsim$ defined on the domain Π. For lotteries $p, q \in \Pi$, we interpret $p \succsim q$ as that the individual prefers the lottery p to q in the initial period. This initial period preference is not affected by the default bias, despite the fact that the individual anticipates that her choices from the decision problems will be affected in all later periods. In this sense, the preference $\succsim$ is default-free.

Requiring that the analyst could observe the default-free preference is a strong assumption. After all, if the preferences are revealed from the individual's choice from menus (with default options), they cannot be free of the effect of defaults. However, when there is only limited ability to switch, recovering default-free preference is straightforward because default does not affect choices from menus with some probability. The analyst could easily recover default-free preference if she observes the individual's choice frequencies from menus. Recovering default-free preference when there is limited willingness to switch is also possible, yet we delay the identification exercise to the next section. In this paper, we choose default-free preferences as our primitive instead of choices from menus for simplicity.

[2]Being placed as the last element of the tuples (c, z, r) (a degenerate lottery) and (z, r) (a decision problem), the lottery r is the default of the menu z. In the main text, the letter r is reserved for the defaults of menus, wherever possible.

2.3 Default Bias Representation

Facing the decision problem (z, r), the individual's choice from the menu z may be affected by its default r in two different ways: (1) A feasible option $p \in z$ may not dominate the default r according to *all* the criteria. In this case, even when p is better than r *overall*, the individual may be unwilling to switch. (2) Even when the individual wants to switch to a non-default option $p \in z$ from the default r, she could be unable to switch. These two sources of the default bias are referred to as *limited willingness to switch*[3] and *limited ability to switch*.

Consider a default biased individual. In each period and for each realization $(c, z, r) \in D$, she values the current consumption c by $u(c)$ and correctly anticipates that she will solve the continuation decision problem (z, r) in the next period according to the following procedure. She will form a sub-menu containing all lotteries that she is willing to switch to. With probability μ she can actively choose the most preferred lottery from the sub-menu, and with probability $(1 - \mu)$ she is unable to switch. This type of behavior can be represented by the following function.

Definition 1. A *Default Bias Representation (DB representation)* consists of two constants $\delta \in (0, 1)$ and $\mu \in (0, 1]$, a continuous function $u : C \to \mathbb{R}$, and a closed and convex set of continuous affine functions $\mathcal{F} \subset \{F : \Pi \to \mathbb{R}\}$, so that the function $V : \Pi \to \mathbb{R}$ defined below represents the preference $\succsim$ on lotteries:

$$V(p) = \int \left[u(c) + \delta \left(\max_{q \in z:\ F(q) \geq F(r)\ \forall F \in \mathcal{F}} \mu V(q) + (1 - \mu) V(r) \right) \right] \mathrm{d}p(c, z, r),$$

[3]The willingness to switch is related to the *pairwise* comparison of a non-default option p and a default option r. Being willing to switch to p when r is the default does not necessarily imply that the individual actually chooses p from a menu: There could be another lottery in the menu that she is also willing to switch to and is preferred to p.

and

$$F(q) \geq F(r) \text{ for all } F \in \mathcal{F} \implies V(q) \geq V(r).$$

This definition requires each criterion F to be an expected utility function, and the preference represented by V can be seen as an aggregation of all criteria in the sense that the dominance of q over r according to all criteria implies the preference for q over r. Notice that the same preference could be represented by multiple sets of criteria. For example, adding V to the set $\mathcal{F}$ will not change the preference. Adding convex combinations of the criteria in F will not change the preference either. We will discuss the uniqueness of the DB representation later. In the representation, the max operator is taken over a sub-menu of lotteries induced by the default, which models the channel of limited willingness to switch. The value of a decision problem (z, r) is the weighted sum of the values of the best lottery and the default, which models the channel of limited ability to switch.

The representation implies that the switching behavior may be incomplete: There could be lotteries q and r and criteria $F_1, F_2 \in \mathcal{F}$, with $F_1(p) > F_1(q)$ and $F_2(p) < F_2(q)$, so that the individual is not willing to switch to q when p is the default, nor willing to switch to p when q is the default. However, as a special case, when $\mathcal{F}$ is singleton, the switching behavior is complete and only the channel of limited ability to switch is present. The DB representation requires $\mathcal{F} = \{V\}$ in this special case, since $F(q) \geq F(r)$ for all $F \in \mathcal{F}$ implies $V(q) \geq V(r)$.

Definition 2. A *Default Bias Representation with limited ability to switch* (DB-LAS) is a DB representation with $\mathcal{F} = \{V\}$. That is,

$$V(p) = \int \left[u(c) + \delta \left(\max_{q \in z} \, \mu V(q) + (1-\mu) V(r) \right) \right] \mathrm{d}p(c, z, r).$$

If the individual is always able to switch to the most preferred lottery q in the submenu when r is the default, then only the channel of limited willingness is present. The DB representation requires $\mu = 1$ in the special case.

Definition 3. A *Default Bias Representation with limited willingness to switch (DB-LWS)* is a DB representation with $\mu = 1$. That is,

$$V(p) = \int \left[u(c) + \delta \left(\max_{q \in z:\ F(q) \geq F(r)\ \forall F \in \mathcal{F}} V(q) \right) \right] \mathrm{d}p(c, z, r).$$

Notice again that in all of the above representations, the first period is treated as a special period: In any later period, choices are affected by the default. By contrast, defaults have no effect in the first period. Treating the first period as a special period in the representation coincides with our primitive, namely default-free preferences. We are going to discuss how to recover default-free preferences from our default-biased individual's choices from menus with defaults.

2.3.1 Identification

While default-free preference could be recovered easily from choices from menus when defaults *only* affect the ability to switch, identifying default-free preference is not trivial when defaults also affect the willingness to switch. In this subsection, we discuss how default-free preference could be recovered from the choice data of a default-biased individual when there is limited willingness only.

Suppose the analyst observes the individual's choices from all decision problems of the type $(\{p, q, r\}, r)$[4], namely the choices from menus with less than or equal to three options. We construct the following binary relation R: For any $p, q \in \Pi$, define

[4] We do not require p, q, r to be distinct.

$p\,R\,q$ if there exists $r \in \Pi$ such that p is chosen in the decision problem $(\{p, q, r\}, r)$.[5]

We interpret $p\,R\,q$ as the fact that p is revealed to be preferred to q. However, it is possible that p is preferred to q (according to default-free preference) while p is not revealed to be preferred to q. This happens when any lottery that is worse than p according to criterion F_p is better than q according to criterion F_q. Therefore, the binary relation R could be incomplete.

Fortunately, we do have more structure imposed on the behavior of our default-biased individual. In particular, we rely on the linearly of default-free preference. For any two lotteries p and q that we cannot compare, if we can write their difference $p - q$ as the scaled difference of two lotteries that we can compare, namely $a(p' - q')$ for some $a > 0$, then we know the ranking over p and q is the same as that over p' and q', due to the linearity of the preference.

We now discuss the sufficient condition that makes the described identification strategy possible. Suppose that the individual has a finite set of criteria $\mathcal{F}$, and there exists a pair q and r with $F(q) > F(r)$ for all $F \in \mathcal{F}$. Then for any lottery p, there exists a sufficiently small $\alpha > 0$ such that the lottery $q' = \alpha q + (1 - \alpha)q$ also satisfies $F(q') > F(r)$ for all $F \in \mathcal{F}$. This is because the criteria are all continuous and linear. With this observation, it is not difficult to show that for any p and q, we can find p' and q' satisfying $p - q = a(p' - q')$, such that $F(p') > F(r)$ and $F(q') > F(r)$ for all $F \in \mathcal{F}$. As a result, the choice from the decision problem $(\{p', q', r\}, r)$ would tell the preference over p and q.

Note that for any lottery r, the finite set of criteria $\mathcal{F}$ yields a cone $\{p \in \Pi : F(p) > F(r),\ \forall F \in \mathcal{F}\}$. If the criteria are not linearly dependent, then the cone is not empty. (See for example, Theorem 5.70 in Aliprantis and Border (2006).)

[5] We assume there is no indifference here for simplicity. Indifference can be taken into account easily if the individual breaks ties by randomizing (with strictly positive probabilities) among the most preferred lotteries.

Therefore, we conclude that if a default-biased individual has a finite set of criteria and if the criteria are linearly independent, then the analyst is able to recover the default-free preference from choice data.

2.3.2 Axioms

Consider the following behavioral conditions.

Axiom 1 (Weak Order). *The binary relation $\succsim$ defined on Π is complete and transitive.*

Axiom 2 (Continuity). *For any $p \in \Pi$, the sets $\{p' \in \Pi : p' \succ p\}$ and $\{p' \in \Pi : p \succ p'\}$ are open.*

Axiom 3 (Independence). *For any $p, q, r \in \Pi$ and $\alpha \in (0,1)$, $p \succ q$ implies $\alpha p + (1-\alpha) r \succ \alpha q + (1-\alpha) r$.*

These first three are common axioms and they have the standard interpretations.

Axiom 4 (Intertemporal Separability). *For any degenerate lotteries (c, z, r), $(c', z', r') \in \Pi$,*

$$\frac{1}{2}(c, z, r) + \frac{1}{2}(c', z', r') \sim \frac{1}{2}(c, z', r') + \frac{1}{2}(c', z, r).$$

The left-hand-side and the right-hand-side lotteries have the same marginal distributions of the current consumptions and the future decision problems: Both lotteries yield c and c' with equal probabilities and (z, r) and (z', r') with equal probabilities. However, they pair the consumptions and the continuation problems differently, so that they have different *joint* distributions. An individual may evaluate the current consumption and the future decision problem separately, ignoring the possible correlation between them. The Intertemporal Separability axiom captures exactly this behavior. This axiom is similar to the separability axiom in Gul and Pesendorfer (2004).

Axiom 5 (Conditional Strategic Rationality)**.** *For any degenerate lotteries* (c, z, r), $(c, z', r) \in \Pi$,

$$(c, z, r) \succsim (c, z', r) \implies (c, z, r) \sim (c, z \cup z', r).$$

The lotteries (c, z, r) and (c, z', r) differ in their future decision problems in a specific way: They have different menus z and z', but the same default choice r. When the default is fixed, the preference for the menu z over z' implies the individual is indifferent between the better menu z and the joint menu $z \cup z'$. This behavior is consistent with the behavior of choosing the best option from the menu in the second period, except that the ranking may depend on the default. The Conditional Strategic Rationality axiom implicitly assumes that default bias is default-specific (instead of menu-specific). This axiom is weaker than the classic strategic rationality axiom as in Kreps (1979), as it requires the two decision problems to have the same default — it may not hold when the two decision problems have different defaults.

Axiom 6 (Timing Indifference)**.** *For any degenerate lotteries* (c, z, r), $(c, z', r') \in \Pi$ *and* $\lambda \in [0, 1]$, *if* $(c, z, r) \succ (c, \{r\}, r)$ *and* $(c, z', r') \succ (c, \{r'\}, r')$, *then*

$$\lambda(c, z, r) + (1 - \lambda)(c, z', r') \sim (c, \lambda z + (1 - \lambda)z', \lambda r + (1 - \lambda)r').$$

The left-hand side is a lottery that yields either (c, z, r) with probability λ or (c, z', r') with probabiltiy $(1 - \lambda)$, which differs in the decision problems for the second period and the uncertainty is resolved in the first period. By contrast, the right-hand side is a lottery that generates for sure the consumption level c, yet a random continuation decision problem that is the $\lambda : (1 - \lambda)$ mixture of the decision problems (z, r) and (z', r'), where the randomness is resolved in the second period. In the literature about menu choices (for example, Gul and Pesendorfer (2004) and Noor (2011)), it is assumed that the decision maker is indifferent between these two, interpreted as the indifference to the timing of the resolution of uncertainty. Our restriction on the decision maker's behavior is weaker, in that we only require the indifference when the decision maker plans to opt out of the default options.

Axiom 7 (Nontriviality)**.** *There exist degenerate lotteries* $(c^*, z, r), (c_*, z, r) \in \Pi$ *such that* $(c^*, z, r) \succ (c_*, z, r)$*.*

This is a common axiom in the literature that rules out the uninteresting situation where an individual is indifferent between any two lotteries.

Axiom 8 (Stationarity)**.** *For any lotteries* $p, q \in \Pi$*,*

$$p \succsim q \Longleftrightarrow (c, \{p\}, p) \succsim (c, \{q\}, q).$$

When a menu is singleton, there is no other choice but to take the default lottery. Since the current period consumption levels are the same, the preference $(c, \{p\}, p) \succsim (c, \{q\}, q)$ suggests that the lottery p is preferred to the lottery q in the next period, which is free of the impact of default bias. This axiom says that the *second period taste* over lotteries, induced by the comparisons of degenerate lotteries with singleton menus, is the same as the current preferences. In other words, preferences are stationary in time.

Axiom 9 (Default Bias)**.** *For any degenerate lottery* $(c, z, r) \in \Pi$ *and lottery* $p \in \Pi$*,*

$$(c, z \cup \{p\}, r) \succ (c, z, r) \implies (c, z \cup \{p\}, p) \sim (c, \{p\}, p) \succ (c, \{r\}, r).$$

This is the key axiom that leads to default bias. The ranking $(c, z \cup \{p\}, r) \succ (c, z, r)$ reveals that adding an option p to the menu z when r is the default makes the decision problem strictly more desirable ex ante. This strict ranking suggests that the individual plans to switch to the non-default option p in the menu $z \cup \{p\}$ in the next period when r is the default (otherwise it would not be a strict preference). This plan reveals that the lottery p is strictly preferred to the default r in the next period (otherwise she would not plan to switch), implying the ranking $(c, \{p\}, p) \succ (c, \{r\}, r)$. Moreover, if the individual plans to choose the non-default option p instead of everything else in the menu z when r is the default, one would expect that the individual would still choose p from the same menu when p becomes the default. In other words, when p is the default, removing everything else from the menu $z \cup \{p\}$ would not reduce its value. That implies the preference $(c, z \cup \{p\}, p) \sim (c, \{p\}, p)$.

Masatlioglu and Ok (2005) have a similar axiom written in terms of a choice correspondence. They argue that choosing p from the set z when it is not the default reveals that p is no worse than any feasible option, so p "would be even stronger relative to the alternatives" when it becomes the default, making it the *only* choice. We do not require that a lottery becomes strictly "stronger" when it becomes the default; we require that it cannot become strictly "weaker".

Axiom 10 (Dominance)**.** *For any lotteries* $p, q \in \Pi$ *with* $(c, \{p, r\}, r) \succ (c, \{r\}, r)$ *and* $(c, \{q, r\}, r) \succ (c, \{r\}, r)$,

$$(c, \{p\}, p) \succsim (c, \{q\}, q) \Longleftrightarrow (c, \{p, r\}, r) \succsim (c, \{q, r\}, r).$$

If both lotteries p and q dominate the fixed default lottery r, meaning that the individual is willing to switch from the default to any of the two lotteries, then we would expect the ranking over $(c, \{p, r\}, r)$ and $(c, \{q, r\}, r)$ to be the same as the taste over p and q in the second period.

Axiom 11 (Default Separability). *For any degenerate lotteries* (c, z, r), $(c, z', r') \in \Pi$ *with* $r, r' \in z \cap z'$, $(c, \tilde{z}, \tilde{r}) \succ (c, \{\tilde{r}\}, \tilde{r})$ *for* $\tilde{r} \in \{r, r'\}$ *and* $\tilde{z} \in \{z, z'\}$ *imply*

$$\frac{1}{2}(c, z, r) + \frac{1}{2}(c, z', r') \sim \frac{1}{2}(c, z, r') + \frac{1}{2}(c, z', r).$$

Axiom 11* (Strong Default Separability). *For any degenerate lotteries* (c, z, r), $(c, z', r') \in \Pi$ *with* $r, r' \in z \cap z'$,

$$\frac{1}{2}(c, z, r) + \frac{1}{2}(c, z', r') \sim \frac{1}{2}(c, z, r') + \frac{1}{2}(c, z', r).$$

These two axioms are related to the interaction between the menu and its default, and both assume separability to some degree. Notice that the preference $(c, z, r) \succ (c, \{r\}, r)$ reveals that the individual is willing to switch from the default r to a non-default option in the menu z (otherwise she would be indifferent between these two lotteries). The weaker version of the two axioms, Default Separability, says that the individual does not care about the correlation between a menu and its default in a random decision problem, as long as the marginal distribution of the menus and the defaults are the same and she is willing to switch in all realizations. The stronger version of the two, Strong Default Separability, does not requires willingness to switch. Both axioms have the similar interpretation to the one for Axiom 4, except for the domain restriction that $r, r' \in z \cap z'$. It will be shown in the next subsection that the weaker version corresponds to the general DB representation, while the stronger version corresponds to the DB representation with limited ability to switch (DB-LAS).

Axiom 12 (Default Consistency). *For any lotteries* $p, r, r' \in \Pi$ *with* $(c, \{p, r\}, r) \succ$

$(c, \{r\}, r)$ *and* $(c, \{p, r'\}, r') \succ (c, \{r'\}, r')$,

$$(c, \{r\}, r) \succsim (c, \{r'\}, r') \implies (c, \{p, r\}, r) \succsim (c, \{p, r'\}, r').$$

Axiom 12* (Default Irrelevance)**.** *For any lotteries* $p, r, r' \in \Pi$ *with* $(c, \{p, r\}, r) \succ (c, \{r\}, r)$ *and* $(c, \{p, r'\}, r') \succ (c, \{r'\}, r')$,

$$(c, \{p, r\}, r) \sim (c, \{p, r'\}, r').$$

These two axioms are related to the ex ante preferences when the ex post switching behaviors are the same. The premise $(c, \{p, r\}, r) \succ (c, \{r\}, r)$ and $(c, \{p, r'\}, r') \succ (c, \{r'\}, r')$ reveal that the individual is willing to switch to the non-default option p when r and r' are the default options. Default Consistency requires that if the individual prefers r to r' in the second period, she also prefers to have r as the default. As a special case, if the individual is indifferent between the default options r and r' (regardless of the second period taste), in which sense the default option is irrelevant (as long as the individual is willing to switch), we have the Default Irrelevance axiom. It will be shown in the next subsection that the weaker version corresponds to the general DB representation, while the stronger version corresponds to the DB representation with limited willingness to switch (DB-LWS).

2.3.3 Main Result

Theorem 2. *If the preference relation* $\succsim$ *satisfies the axioms A1–A12, it can be represented by the Default Bias Representation.*

Conversely, a function $V : \Pi \to \mathbb{R}$ *characterized by the tuple* $(u, \delta, \mu, \mathcal{F})$, *with the properties specified in Definition 1, represents a preference relation that satisfies* *axioms A1–A12.*

Moreover, suppose the preference $\succsim$ is represented by a Default Bias Representation with parameters $(u, \delta, \mu, \mathcal{F})$. Then $(u', \delta', \mu', \mathcal{F}')$ represents the same preference if and only if (1) there exist $a \in \mathbb{R}$ and $b \in \mathbb{R}_{++}$ so that $u = a + bu'$, (2) $\delta = \delta'$ and $\mu = \mu'$, and (3) $\mathcal{F} \approx \mathcal{F}'$.[6]

The theorem characterizes the Default Bias Representation with the axioms we discussed before. It also provides a uniqueness result: If two DB representations $(u, \delta, \mu, \mathcal{F})$ and $(u', \delta', \mu', \mathcal{F}')$ represent the same preference, they must have the same taste over consumption levels (u and u' are positive affine transformations of each other), the same discount factor, the same probability of making active decisions, and the "same" switching criteria $\mathcal{F}$ and $\mathcal{F}'$.

The DB representation has two special cases. If the limited ability to switch is shut down, the DB representation with limited willingness to switch (DB-LWS) follows. If the limited willingness to switch is shut down, the DB representation with limited ability to switch (DB-LAS) follows.

Corollary 2.1. *The preference relation $\succsim$ satisfies the axioms A1–A10 and A12*, if and only if it is represented by the DB-LWS representation,*[7] *which is the DB representation with $\mu = 1$.*

Corollary 2.2. *The preference relation $\succsim$ satisfies the axioms A1–A10, A11* and*

[6] This means informally that the set of switching criteria $\mathcal{F}$ and $\mathcal{F}'$ are the same.

Recall that we have argued that adding (or removing) convex combinations of the functions in $\mathcal{F}$ would not affect the preference. As a result, we will lose our uniqueness result. To address this issue, we can focus on the largest set of criteria functions, denoted by $\langle \mathcal{F} \rangle$, that is consistent with the preference by adding all the convex combinations of the functions in $\mathcal{F}$ as well as their positive affine transformations. Formally, define $\langle \mathcal{F} \rangle = cl(cone(\mathcal{F})) + \{\theta \mathbf{1}_D\}_{\theta \in \mathbb{R}}$, which is the closure of the affine convex cone of $\mathcal{F}$.

The theorem says that if two DB representations represent the same preference, we must have $\langle \mathcal{F} \rangle = \langle \mathcal{F}' \rangle$. To express the result in a more intuitive way, we write this as $\mathcal{F} \approx \mathcal{F}'$.

[7] Axiom A11 is implied by these axioms.

A12, if and only if it is represented by the DB-LAS representation,[8] *which is the DB representation with* $\mathcal{F} = \{V\}$.

We summarize the results in Table 2.3.3. The Default Bias Representation requires all the 12 axioms. The DB-LWS representation strengthens A12 (Default Consistency) to A12* (Default Irrelevance). The DB-LAS representation strengthens A11 (Default Separability) to A11* (Strong Default Separability)

Table 2.1: Summary of the Main Results.

Representations	Axioms
DB	Axioms A1 – A10, A11, A12
DB-LWS	Axioms A1 – A10, A11, A12*
DB-LAS	Axioms A1 – A10, A11*, A12

Proof Sketch

We sketch the proof for the first part of the theorem, with more details provided in Appendix 2.7. The proofs of the corollaries can also be found there.

Using the standard results in the literature, A1 – A3 yield an expected utility representation, A4 separates current consumption and the second period decision problem, and A5 reduces a menu to its submenus, one can get the following representation:[9]

$$V(p) = \int \left[u(c) + \max_{q \in z} \; w(\{q, r\}, r) \right] \mathrm{d}p(c, z, r).$$

The remainder of the proof is separated into two steps. Step 1 characterizes the ex post choice in the second period. In other words, it answers the question of what

[8] Axiom A9 can be slightly weakened in this case.

[9] Axiom 5, Conditional Strategic Rationality, is weaker than the usual strategic rationality. As a result, one can only reduce a menu to a two-element submenu rather than a singleton submenu.

the individual would want to choose from menu z when r is the default. This is related to the individual's willingness to switch. Step 2 characterizes the ex ante comparisons. Even though the individual wants to choose p from $\{p, r\}$ when r is the default, she may still prefer $(c, \{p\}, p)$ to $(c, \{p, r\}, r)$ since she has limited ability to switch. This is related to the individual's ability to switch.

Step 1:

Observe that the ex post switching behaviors of the individual can be identified by ex ante preferences: If adding the lottery p to the menu of $(c, \{r\}, r)$ increases its value, namely $(c, \{p, r\}, r) \succ (c, \{r\}, r)$, it reveals that the individual is willing to switch from the default r to the non-default option p (otherwise adding p will not increase its value). Note that it is possible that the individual is willing to switch from the default r to p, with p yielding the same value as r. But this would not result in a strict preference ex ante. To take this "weak willingness" into consideration, we can use the fact that the preference is continuous to define the weak willingness to switch by taking limits.

Definition 4 (Willingness to Switch)**.** For any p and $r \in \Pi$, define $p \succsim^* r$ if and only if there exist sequences $p_n \to p$ and $r_n \to r$ with $(c, \{p_n, r_n\}, r_n) \succ (c, \{r_n\}, r_n)$ for all n or $p = r$.

This induced binary relation $\succsim^*$ is transitive and satisfies the continuity and independence properties in the usual sense, if axioms A6 and A9 hold. However, it may not be complete. Invoke the result in Dubra et al. (2004) that a preorder that satisfies continuity and independence has an Expected Multi-Utility representation to get the following lemma.

Lemma 3 (Expected Multi-Utility)**.** *Assuming A1 – A6 and A9, there exists a closed and convex set of continuous linear functions $\mathcal{F} \subset \{F : \Pi \to \mathbb{R}\}$ so that for each*

$p, r \in \Pi$,

$$p \succsim^* r \Longleftrightarrow F(p) \geq F(r) \textit{ for all } F \in \mathcal{F}.$$

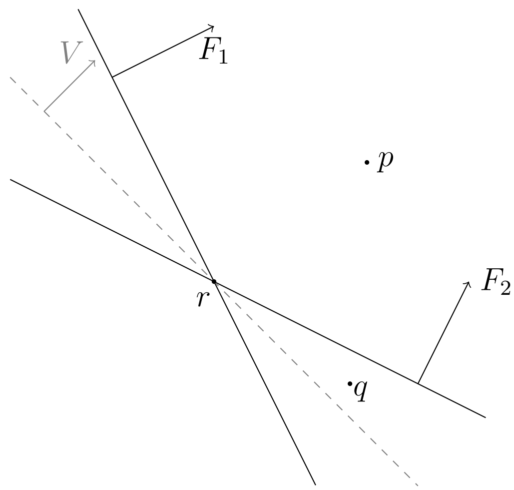
The lemma implies there is a set of expected utility functions, $\mathcal{F}$, and $p \succsim^* r$ if and only if the expected utility of p with respect to F is larger than the expected utility of r with respect to F, for each of the functions F in the set $\mathcal{F}$. One can further show that the Default Bias axiom (A9) implies that $p \succsim^* r$ implies $(c, \{p\}, p) \succsim (c, \{r\}, r)$, with the latter interpreted as the individual's taste in the second period. Therefore, the second period taste is a completion of $\succsim^*$.

Intuitively, there is a collection of different criteria $\mathcal{F}$ that is related to the evaluation of the lotteries. The individual is only willing to switch from r to p when p is better than r according to all criteria. Moreover, the individual's second period taste (which is the same as the first period one by the Stationarity axiom A8) can be interpreted as an aggregation of these criteria. Notice, however, that there could be a lottery p with $(c, \{p\}, p) \succsim (c, \{r\}, r)$ but *not* $p \succsim^* r$. The individual does not bother to switch from r to p, as long as there is a single criterion in which p is not superior to r, even though p is better than r *overall.*

This decision process is illustrated by Figure 2.1. The lines labeled F_1 and F_2 are the indifference curves of the two criteria in $\mathcal{F}$ and the dashed line labeled V is the taste in the second period. We have $p \succsim^* r$ because it is better than r according to both criteria, and it follows that it is also better according to V. On the other hand, $q \not\succsim^* r$, even though it is better according to V. This is because q is not better than r according to F_2. The figure also illustrates that $\succsim^*$ may not be complete: we cannot rank q and r according to $\succsim^*$. Masatlioglu and Ok (2005) have a similar interpretation in their paper.

Define $\mathcal{U}^*(r) = \{p \in \Pi : \ p \succsim^* r\}$, which is the upper contour set of r with respect

Figure 2.1: Illustration of switching behavior



to $\succsim^*$. The set $\mathcal{U}^*(r)$ consists of all the lotteries that the individual is willing to switch to when r is the default. One can show that the Conditional Strategic Rationality axiom (A5) implies that $(c, z, r) \sim (c, z \cap \mathcal{U}^*(r), r)$. That is, the individual ignores the options in the menu that she is not willing to switch to when r is the default.

Lastly, by assuming the Dominance axiom A10 and the Stationarity axiom A8, one can show that choice is made from the sub-menu, $z \cap \mathcal{U}^*(r)$, according to the second period taste, which is equivalent to the first period one represented by V.

Step 2:

While the ex post selection of the individual is characterized by Step 1, it remains to deal with the *ex ante* comparisons among the decision problems. Consider the case that $(c, \{p, r\}, r) \succ (c, \{r\}, r)$. This tells that the individual wants to switch from r to p when r is the default *ex post* in the second period. However, it does not tell whether $(c, \{p, r\}, r)$ is strictly preferred to or indifferent to $(c, \{p\}, p)$. In other words, characterizing the ex post ranking does not full pin down the ex ante ranking.

The Default Separability axiom, A11, implies that the individual evaluates the default separately from the menu, as long as she is willing to switch from the default. Despite the separability, she still prefers to have a better default, as is implied by the Default Consistency axiom A12. Intuitively, one can think of this as that the individual still "consumes" the defaults, even though she is willing to switch from them. Equivalently, this could be interpreted as the limited ability to switch from the defaults: With some probability, she is unable to switch to a desired non-default lottery in the menu, so that she consumes the default. This probability can be identified due to Axiom A12 using the uniqueness result of the expected utility representation.

2.4 Comparative Statics

In this section, we make behavioral comparisons of two individuals whose preferences are represented by Default Bias Representation. Recall that DB representation is characterized by the tuple $(u, \delta, \mu, \mathcal{F})$, in which the pair (u, δ) is related to the evaluation of the underlying consumption streams, $\mathcal{F}$ is related to whether she is willing to switch, and μ is the probability that the individual is able to switch to a non-default option.

To check whether two individuals have the same tastes over the consumption streams, one can check whether their preferences agree on a special subset of the domain, the set of all *temporal lotteries.* A temporal lottery yields a singleton menu that contains only the default option as the continuation decision problem in each period and for each realization. Default bias does not affect the rankings of temporal lotteries.

In what follows, we establish the behavioral criteria to compare the degree to which individuals are affected by default bias and connect them with the relevant parameters of the representation. We compare the two channels separately.

2.4.1 Comparing Willingness to Switch

Roughly speaking, if individual 1 is less willing to switch than individual 2, then whenever individual 1 is willing to switch from the default r to the non-default p, individual 2 cannot be unwilling to switch. Individual 1's willingness to switch is revealed by $(c, \{p, r\}, r) \succ_1 (c, \{r\}, r)$. Individual 2's unwillingness to switch is revealed by *both* $(c, \{p, r\}, r) \sim_2 (c, \{r\}, r)$ and $(c, \{p\}, p) \succ_2 (c, \{r\}, r)$. Note that we need the second condition because the first condition also holds when individual 2 prefers r to p, which has nothing to do with her willingness to switch. This observation leads to the behavioral comparison of willingness to switch defined below.

Definition 5 (Less Willing to Switch)**.** Individual 1 is *less willing to switch* than individual 2 if for any lotteries $p, r \in \Pi$ with $(c, \{p\}, p) \succ_i (c, \{r\}, r)$ for both $i = 1, 2$ and any consumption $c \in C$,

$$(c, \{p, r\}, r) \succ_1 (c, \{r\}, r) \implies (c, \{p, r\}, r) \succ_2 (c, \{r\}, r).$$

Again, the definition does not rule out the possibility that $(c, \{p, r\}, r) \succ_1 (c, \{r\}, r)$ yet $(c, \{p, r\}, r) \sim_2 (c, \{r\}, r)$. Instead, it says that whenever this happens, it cannot be the case that $(c, \{p\}, p) \succ_2 (c, \{r\}, r)$.

This behavioral comparison of willingness to switch can be connected to the parameter $\mathcal{F}$ that governs willingness to switch in the DB representation. Recall that the individual is willing to switch to a non-default lottery p when r is the default only when p dominates r according to all criteria $F \in \mathcal{F}$. Intuitively, she is less willing to switch when the set $\mathcal{F}$ is larger[10]. This intuition is not completely accurate. We could have $F_1(p) \geq F_1(r)$ for each $F_1 \in \mathcal{F}$ but $F_2(p) < F_2(r)$ for some $F_2 \in \mathcal{F}_2$, as long as $V_2(p) \leq V_2(r)$. To make the a violation of individual 1 being less willing

[10]Recall that we compare the largest such set of criteria $\langle \mathcal{F} \rangle$ to address the uniqueness issue.

to switch, we also need $V_2(p) > V_2(r)$. This exactly parallels the discussion in the beginning about why we need $(c, \{p\}, p) \succ_2 (c, \{r\}, r)$.

Theorem 4 (Part 1)**.** *Individual 1 is less willing to switch than individual 2 if and only if* $\langle \mathcal{F}_1 \cup \{V_2\} \rangle \supset \langle \mathcal{F}_2 \cup \{V_1\} \rangle$*, where* $\mathcal{F}_i$ *is the parameter in the DB representation* V_i *for individual* i*, with* $i = 1, 2$ *.*

Notice that the comparison of willingness to switch do not involve the rest of the parameters in the DB representation, namely u, δ, μ. Now we interpret the condition $\langle \mathcal{F}_1 \cup \{V_2\} \rangle \supset \langle \mathcal{F}_2 \rangle$, which is equivalent to the parametric condition in the theorem. We first restrict our attention to the smaller criteria set $\langle \mathcal{F}_1 \rangle \cup \{V_2\}$ for individual 1, since removing the convex combinations of $F \in \mathcal{F}_1$ and V_2 from the original set

It requires that whenever p dominates r according to the criteria in $\mathcal{F}_1$, *and* according to V_2, then p also dominates r according to the criteria in $\mathcal{F}_2$. To put it differently, the set of criteria $\mathcal{F}_2$ of individual 2 must be contained by the set generated by individual 1's criteria *and* V_2. The reason why we need the V_2 part is the same as why we need $(c, \{p\}, p) \succ_2 (c, \{r\}, r)$.

2.4.2 Comparing Ability to Switch

The default biased individual may not always be able to switch when she desires to do so. Effectively, the plan she makes to switch to the non-default lottery p when the default is r is equivalent to the singleton menu that contains the mixture of p and r. This observation can be used to compare the ability to switch for different individuals.

Definition 6 (Less Able to Switch)**.** Individual 1 is *less able to switch* than individual 2 if for any lotteries $p, r \in \Pi$ and consumption $c \in C$ with $(c, \{p, r\}, r) \succ_i (c, \{r\}, r)$

for both $i = 1, 2$, and any $\alpha \in [0, 1]$,

$$\begin{aligned} &(c, \{p, r\}, r) \succ_1 (c, \{\alpha p + (1-\alpha)r\}, \alpha p + (1-\alpha)r) \\ \Longrightarrow &(c, \{p, r\}, r) \succ_2 (c, \{\alpha p + (1-\alpha)r\}, \alpha p + (1-\alpha)r). \end{aligned}$$

Notice that we need the premise to guarantee that both individuals are willing to switch. This behavioral comparison is closely connected with the comparison of the parameter μ that governs the ability to switch, and the result follows naturally.

Theorem 4 (Part 2)**.** *Individual 1 is less able to switch than individual 2 if and only if $\mu_1 \leq \mu_2$, where μ_i is the parameter in DB representation, with $i = 1, 2$.*

2.5 Applications

Default bias arises through two channels, limited willingness and limited ability to switch. The first channel is closely related to status quo bias, whose applications have been relatively well-studied in the literature. Thus the focus of this section is discussing two stylized applications regarding the second channel only. That is, we shut down the first channel and focus only on default bias representation with limited ability to switch:

$$V(p) = \int \left[u(c) + \delta \left(\max_{q \in z} \ \mu V(q) + (1-\mu) V(r) \right) \right] \mathrm{d}p(c, z, r).$$

2.5.1 Asset Replacement

Consider an individual who holds an asset that generates payoffs in each period. In particular, if the asset is t periods old, it generates an expected payoff of $u_t \geq 0$, with $t = 1, 2, \cdots$. In each period the individual can either replace it with a brand new

(0-period-old) asset at cost $c > 0$ or keep using the old one, with the latter being the default choice. There is no installation time — the replacement asset can be used immediately. With probability $\mu \in (0, 1]$, the individual may be affected by limited ability to switch and has to keep the asset (independent of her preferences). This can be interpreted as the replacement asset being out of stock with probability $(1-\mu)$, or the individual's time being occupied by other activities that preclude replacing the asset.

The asset depreciates as it becomes older. Assume that the sequence of expected payoffs $\{u_t\}$ is decreasing in t and $\lim_{t\to\infty} u_t = 0$. This could be interpreted as the asset requires larger maintenance cost as it becomes older, which could be stochastic. However, we require there is no autocorrelation in these costs so that the payoffs to the asset holder can be completely summarized by the expected payoff sequence $\{u_t\}$. Let the discount factor be $\delta \in (0, 1)$. Assume that the individual can only hold a single asset and that the total payoff generated by the asset is lager than its cost, namely $\sum_{k=0}^{\infty} u_k \delta^k > c$, so that the individual wants to replace it before it generates negligible payoffs.

In the beginning of a period in which the individual holds a t-period-old asset (we refer to her as the t-period self) and she is able to make an active decision (the replacement asset is not out of stock), she chooses to keep the asset if its value is strictly larger than the value of a brand-new asset minus the replacement cost.[11] Notice that she is fully aware of the limited ability to switch. Let the value of a t-period-old asset, with replacement decisions taken into account, be denoted by V_t. Then the sequence $\{V_t\}$ satisfies the following expression:

$$V_t = u_t + \mu\delta \max\{V_0 - c, V_{t+1}\} + (1 - \mu)\delta V_{t+1}.$$

[11]We impose the tie-breaking rule that she replaces the asset when she is indifferent between the two options.

Observe that this is a DB-LAS representation.

Preemptive Replacement

In Appendix 2.9, we show that for any μ there is a threshold $T(\mu)$ such that the $t-$period self replaces the asset whenever she is able to do so if and only if $t \geq T$. We refer to $T(\mu)$ as the *optimal replacement time*. We define $T^* := T(1)$ as a benchmark: Without default bias, the individual replace the asset when it becomes T^* periods old.

Mathematically, the value of the asset $\{V_t\}$ is a decreasing sequence and there is a $T(\mu)$ such that

$$V_{T(\mu)-1} > V_0 - c \geq V_{T(\mu)}.$$

Note that V_t also depends on μ, although we have suppressed the dependency. We may also suppress the dependency of T on μ for simplicity.

With default bias ($\mu < 1$), the t-period self (with $t < T$) faces the following trade-off: While she still gains considerable payoffs from keeping old asset if she waits until T to replace, she could be stuck with the old asset for a sub-optimally longer periods because of default bias. This trade-off arises exactly because her earlier choice of keeping the asset determines the default option for the later period, which is the depreciated asset. The goal of this section is to investigate how $T(\mu)$ changes with μ, which is the probability that she is able to make active decisions.

As μ decreases, it is more likely that the individual cannot replace the asset when she wants to. Therefore, we expect that the value of the asset decreases. It is proved in Appendix 2.9 that V_{T-1}, $V_0 - c$, and V_T decrease as μ decreases. For a fixed optimal replacement time T[12], we are going to argue that V_{T-1} and V_T decrease faster than

[12]Note that $T(\mu)$ changes discretely in μ.

$V_0 - c$ when μ decreases. As a result, the individual would replace the asset earlier when she is less likely to make active actions.

To see why this is the case, consider a marginal decrease in μ such that $T(\mu)$ remains unchanged. If the individual plans to replace it in the next period, which is the case for the t-period self when $t = T-1$ or T, her valuation of the asset is sensitive to the change in μ, since it becomes more likely that she has to sub-optimally keep the depreciated asset instead of replacing it with the new one in the *next* period. On the other hand, the value of the brand new asset is less sensitive to the change in μ, this is because the event of being stuck in the sub-optimal depreciated asset happens T periods in the future. This intuition is captured by the following theorem, with its proof provided in Appendix 2.9.

Theorem 5. *The individual replaces the t-period-old asset whenever possible as long as $t \geq T(\mu)$. Moreover, she makes preemptive replacement decisions when she anticipates default bias. That is, $T(\mu)$ decreases as μ decreases.*

The theorem suggests that the individual wants to replace the asset when it is more than $T(\mu)$ periods old. However, due to limited ability to switch, there is an expected delay of $\frac{1-\mu}{\mu}$ periods. Therefore, the *expected replacement time* would be $S(\mu) := T(\mu) + \frac{1-\mu}{\mu}$. While the theorem suggests that $T(\mu) \leq T^*$, it is unclear whether $S(\mu)$ is earlier or later than T^*. In other words, while we know the default biased individual *plans* to replace the asset earlier, we do not know whether she *actually* replaces it earlier or later on average. The comparison of S and T^* depends on all the parameters and shall be studied case by case.

We provide intuition below regarding the effect of the payoff schedule $\{u_t\}$ on T when μ is fixed, followed by a simple simulation analysis that also demonstrates the comparison between S and T^*. The main purpose is to further elaborate the dynamic trade-off caused by default bias in the context of this application.

Recall that the individual replaces the asset preemptively because she anticipates that she may not be able to replace it at T^*. While preemptive replacement avoids holding a depreciated asset beyond T^* for too long, the individual has to give up the asset when it is still generating considerable payoffs. Therefore, if the payoffs right before T^* are sufficiently high relative to the payoffs right after T^*, the individual does not want to replace the asset too early. We would anticipate that T is not too small compared with T^* and S to be later than T^*. On the other hand, we would expect the opposite case if the payoffs right before T^* are not very high relative to the payoffs right after T^*.

To further demonstrate this trade-off, we provide a simple numerical analysis. Consider the following payoff schedule $\{u_t\}$:

$$u_t = \frac{1}{K} \begin{cases} 1 & \text{if } t \leq 24 \\ u & \text{if } 25 \leq t < 30 \\ 0 & \text{otherwise} \end{cases}$$

where $u \in [0, 1]$ and $K = \sum_{k=0}^{24} \delta^k \cdot 1 + \sum_{k=25}^{29} \delta^k \cdot u$. The parameter K is a normalizing factor so that keeping the asset forever yields total discounted payoff 1, which is larger than the replacement cost. We set $\delta = 0.9$ and $c = 0.75$.

Suppose the optimal replacement time is 30 in the absence of default bias. That is, $T^* = 30$. The parameter u captures the flow payoff she has to give up if she replaces the asset preemptively. To demonstrate the trade-off, we compute $T(\mu)$ and $S(\mu)$ for $\mu = 0.25, 0.5, 0.75, 1$ as we vary $u = 0.3, 0.2, \cdots, 1$. The results are summarized by the table below. More details of the numerical analysis can be found in Appendix 2.9.

Table 2.2: The optimal replacement time T and expected actual replacement time S under different parameters μ and u.

	T^*	$T(0.75)$	$S(0.75)$	$T(0.5)$	$S(0.5)$	$T(0.25)$	$S(0.25)$
$u = 0.3$	30	29	29.33	28	29	26	29
0.4	30	30	30.33	29	30	28	31
0.5	30	30	30.33	29	30	29	32
0.6	30	30	30.33	30	31	29	32
0.7	30	30	30.33	30	31	29	32
0.8	30	30	30.33	30	31	30	33
0.9	30	30	30.33	30	31	30	33
1	30	30	30.33	30	31	30	33

We know that for each fixed u, as the probability that the individual can make an active decision decreases, she has stronger incentives to replace the asset preemptively. In the table, in each row, T weakly decreases from the left to the right. This demonstrates Theorem 5.

Moreover, when we fix μ, as the individual has to give up larger payoffs if she replaces the asset preemptively, she has weaker incentives to do so. In the table, in each column, T weakly increases from the top to the bottom.

Finally, the expected actual replacement time S could be either earlier or later than T^*.

Discussion

We have assumed that the asset generates an expected payoff u_t when it is t periods old. It is important to notice that we implicitly requires the payoffs to be serial independent: The realization of payoff at period t does not provide any information about the distribution of u_{t+1}.

However, the theorem is not necessarily true when there is serial dependency in $\{u_t\}$. Consider the following alternative setting. Suppose that an asset can either be in a good state or a bad state, with payoffs $u > 0$ and 0 respectively. If a t-period-old asset was in the good state in the last period, it is still in the good state in the current period with probability π_t. Suppose that a bad asset cannot be fixed and therefore must be replaced.

The previous argument suggests whether an individual replaces the asset earlier when μ decreases depends on which of V_{T-1} and $V_0 - c$ is more sensitive to the change in μ. Previously V_0 is less sensitive to μ because it is only necessary to replace the asset when it gets old, which happens in the distant future. However, in the current setting, the asset could be replaced in every period since the bad state can realize. Furthermore, it is more urgent to replace a bad asset than an old asset in the good state. As a result, it is ambiguous which of the V_{T-1} and V_0 is more sensitive to the change in μ. Nonetheless, if it is very unlikely that a bad state can happen in the early age of the asset, a default biased individual still makes preemptive replacement decisions.

Implication for Dynamic Discrete Choices

Estimating dynamic stochastic games can be computationally costly when there is an exogenous move of nature, especially when the state space is large. This "curse of dimensionality" can be alleviated by focusing on continuous time games in which the move of each player, including nature, arrives stochastically at some rate (for example, Arcidiacono et al. (2016)). This is an analog of the DB-LAS representation (the DB representation with limited ability to switch only).

The asset replacement application discussed above suggests that addressing the computational problem by introducing the random arrivals of "decision opportunities" is not without cost, in the sense that it might cause estimation biases.

To see this, the asset replacement model suggests that the individual makes different decisions in dynamic problems when she anticipates default bias. Moreover, different combinations of the parameters of interest (which is u_t in the asset replacement model) and the probability of making active decisions (μ) may lead to the same observational behaviors. Therefore, if an econometrician only observes the equilibrium of a dynamic game, she may not be able to separately identify the parameters. Computational simplification may come at the cost of potential biases.

2.5.2 Exploiting Forgetful Consumers

Facing a new product or service, a consumer may want to give it a try and discover her valuation for it. Firms sometimes offer new-user benefits to attract potential consumers. Such promotions often come with automatic renewals: By default a new user stays as a subscriber and is automatically charged at the regular price after the promotional period, unless she explicitly cancels the subscription.

Here default bias can be interpreted as forgetting to cancel the subscription when a user does not like the service. Suppose a potential user is unaware of her forgetfulness. Then a firm can exploit her with auto-renewals, since she may continue to subscribe to the service for a while even though she does not like it. The situation in which a user correctly anticipates her forgetfulness is more interesting. When she anticipates that she may forget to unsubscribe, the firm needs to offer a larger promotion to attract her in the first place. It is not immediately clear whether the larger promotion can be covered by the potential exploitation of her forgetfulness under the *policy with auto-renewal*.

As an alternative policy, the firm could adopt the policy under which the subscription expires automatically after the promotional period for new users. In this case, by default a new user does not stay as a subscriber after the promotional period.

Those who do not like the service would leave automatically, and those who like the service could re-enroll and become a subscriber *forever*. Re-enrollment requires an active action, and default bias suggests that the users who would like to stay may forget to re-enroll.[13] We call this the *policy without auto-renewal* for new users.

In this application, we analyze whether a monopoly firm should use initial promotions with or without automatic renewals when facing potential users who are forgetful, modeled by the DB-LAS representation. Recall that the DB-LAS representation is the DB representation with limited ability to switch only: An individual may not be able to opt out of the default option with some probability.

Policy Analysis

Suppose a consumer either has high or low valuation $v \in \{v_H, v_L\}$ for a service per period, with $v_H > v_L$. Neither the user nor the firm knows the valuation for the service initially, and they have common prior belief that the probability of having high valuation is $q \in (0, 1)$. The user privately discovers her valuation once she uses the service. In each period, the user can either choose to stay enrolled or to cancel the subscription. The default option of this choice problem depends on firm's policy. Suppose with probability $1 - \mu$ she may forget to make an active choice and stick to the default option, where $\mu \in (0, 1]$. Both the user and the firm have the same discount factor $\delta \in (0, 1)$. The marginal cost for the firm to provide the service is $c > 0$. We define $\bar{v} := qv_H + (1 - q)v_L$, which is the expected valuation. We assume $\bar{v} - c + \frac{\delta}{1-\delta}(v_H - c) \geq 0$. That is, we require it to be profitable for the firm to serve a new user who does not forget.

Assume that the firm cannot commit to any future prices. In particular, this

[13] The firm could alternatively ask them to actively re-enroll in each period. However, anticipating default bias it does not make sense for the firm to offer this alternative.

rules out the possibility that the firm sets the future prices at the marginal cost and sets the initial price so that a potential user is indifferent between accepting the offer or not. Under this assumption of no commitment, it is an equilibrium for the firm to charge the static monopolist price and for the user to quit whenever the price is larger than her valuation in each period except for the initial one. To make the model interesting, assume that $q(v_H - c) > v_L - c$, so that the monopolist price is $p = v_H$.

Policy with Auto-Renewal

We first consider the policy with auto-renewal. Default bias is relevant only when a user turns out to have low valuation. The firm needs to compensate the new user for the possible loss from having low valuation and forgetting to quit, but it gains from her if she forgets.

Suppose the firm charges the promotional price p_0 in the first period. We have argued that it charges $p = p_H$ in all later periods. Then the expected value V_0 for a potential default biased user to subscribe is

$$V_0 = (-p_0 + \bar{v}) + q\frac{\delta}{1-\delta}(v_H - p) + (1-q)\frac{(1-\mu)\delta}{1-(1-\mu)\delta}(v_L - p),$$

where $p = v_H$. If the user never forgets ($\mu = 1$), her total expected payoff V_0 consists of the first period payoff $\bar{v} - p_0$ before she learns her valuation for the service, and the *experimental value* which is the sum of the discounted payoff $v_H - p$ from each of the later periods if she turns out to have high valuation. These are the first two terms in the expression above. Notice that the experimental value vanishes in this application since $p = v_H$.

If she may forget ($\mu < 1$) to quit, she would anticipate the loss from forgetting to unsubscribe if she turns out to have low valuation. If a low valuation user is not able to quit t periods after the initial trial, which happens with probability $(1-\mu)^t$, she suffers the discounted loss $\delta^t(v_L - p)$ in that period. Therefore, the total *anticipated loss* is $\sum_{t=1}^{\infty}(1-\mu)^t\delta^t(v_L - p)$, which yields the last term in the expression for V_0.

If the firm adopts the auto-renewal policy, it would set the highest possible initial price p_0 as long as a potential user is willing to subscribe initially, given that the price will be $p = v_H$ in the later periods. Therefore,

$$p_0 = \bar{v} + \frac{(1-\mu)\delta}{1-(1-\mu)\delta}(1-q)(v_L - v_H).$$

Without default bias (when $\mu = 1$), the initial price is set at $p_0 = \bar{v}$. Observe that p_0 is increasing in μ, meaning that as a user becomes more forgetful, the firm needs to set a lower price to compensate the growing anticipated loss from the bias.

The total expected profit for the firm under the auto-renewal policy is

$$\pi = (p_0 - c) + q\frac{\delta}{1-\delta}(p-c) + (1-q)\frac{(1-\mu)\delta}{1-(1-\mu)\delta}(p-c),$$

where $p = v_H$. If there is no default bias ($\mu = 1$), the firm earns $p_0 - c$ in the first period and the discounted sum of the flow profit $p - c$ in each of the later periods if the user has high valuation.

When there is default bias ($\mu < 1$), the firm can exploit the user if she has low valuation. This is captured by the last term, referred to as the *exploitation value.* The firm is able to earn the discounted flow profit $\delta^t(p-c)$ from a low valuation user if she stays as a subscriber t periods after the initial trial, which happens with probability $(1-\mu)^t$. Therefore, the total exploitation value is $\sum_{t=1}^{\infty}(1-\mu)^t\delta^t(p-c)$, which yields the last term in the expression for π.

We plug in the expression for the initial price p_0 and the price $p = v_H$ for later periods to get the following expression:

$$\pi = \bar{v} - c + q\frac{\delta}{1-\delta}(v_H - c) + (1-q)\frac{(1-\mu)\delta}{1-(1-\mu)\delta}(v_L - c),$$

where the last term vanishes if there is no default bias ($\mu = 1$). This last term is the *net benefit from default bias* (the exploitation value minus the compensation for the anticipated loss). As we have argued before, while the firm can exploit forgetful users, it has to compensate them if they correctly anticipates the forgetfulness. For each extra period as a subscriber, a low valuation user requires the firm to compensate her $p - v_L$, while the firm earns $p - c$. These two terms yield the net benefit of $v_L - c$, which gives the last term in the expression above.

If the per period net benefit from default bias $v_L - c$ is positive, meaning that the exploitation value is larger than the compensation for the anticipated loss, the more forgetful a user is (smaller μ), the longer she stays as a subscriber so that the firm enjoys more net benefit from the bias. Mathematically, the total expected profit π is decreasing in μ when $v_L \geq c$. We would have exactly the opposite result when $v_L - c$ is negative.

Policy without Auto-Renewal

Now we consider the policy without auto-renewal. This time, default bias does not affect a user's payoff: If she has low valuation, she quits automatically after the initial period; if she has high valuation, the prices for later periods are set at v_H so that she is indifferent between subscribing to the service or not. Therefore, it is straightforward to get the initial price $p_0 = \bar{v}$, which is the highest price under which a new user is willing to try the service.

On the other hand, default bias does affect the firm's expected profit, since the user with high valuation may forget to renew after the initial trial. Under the policy without auto-renewal, the total expected profit is

$$\tilde{\pi} = \bar{v} - c + q\delta \frac{\mu}{1-(1-\mu)\delta} \frac{1}{1-\delta}(v_H - c),$$

where $\bar{v} - c$ is the profit in the initial period and the last term is the profit from the user with high valuation. If a high valuation user is able to re-enroll t periods after the initial trial, which happens with probability $(1-(1-\mu)^t)$, the firm earns the discounted flow payoff $\delta^t(p-c)$ where $p = v_H$. Therefore, the total expected profit the firm earns from a high valuation user is $\sum_{t=1}^{\infty} \delta^t(1-(1-\mu)^t)(v_H - c)$, which is the last term in the expression.

Observe that multiplicative factor $\mu/(1-(1-\mu)\delta)$ in the expression is increasing in μ and becomes 1 when there is no default bias ($\mu = 1$). This factor arises since it might take a few periods for a user with high valuation to get re-enrolled due to the forgetfulness. Obviously under the policy without auto-renewal, the firm prefers less forgetful users.

Policy Comparison

The firm should adopt the policy with auto-renewal if $\pi \geq \tilde{\pi}$ and $\pi \geq 0$; it should adopt the policy without auto-renewal if $\tilde{\pi} \geq \pi$ and $\tilde{\pi} \geq 0$. Otherwise, it should not attempt to attract new users.

One can compute and simplify the difference in profits to get the following expression.

$$\pi - \tilde{\pi} = \frac{(1-\mu)\delta}{1-(1-\mu)\delta} \left[q(v_H - c) + (1-q)(v_L - c)\right].$$

This expression intuitive. By adopting the auto-renewal policy instead of the one without auto-renewal, the firm gains from a high valuation user $(v_H - c)$ in each period before she is able to re-enroll, and earns the net benefit from a low valuation user $(v_L - c)$ in each period (which is the exploitation value minus the compensation for anticipated loss). These values need to be weighted by the prior probabilities (q and $1-q$) and scaled by the expected length of the "forgetful periods" adjusted by the discount factor.

Observe first that the difference $\pi - \tilde{\pi}$ is positive when $v_L \geq c$. One can verify that in this case, $\pi \geq 0$ and π is decreasing in μ. That is, the more forgetful the users are, the more profit the firm earns. This is the case in which the exploitation value is larger than the compensation for the anticipated loss from default bias.

The difference $\pi - \tilde{\pi}$ is still positive if $\bar{v} = qv_H + (1-q)v_L \geq c$, even though $v_L < c$. One can verify that in this case $\pi \geq 0$[14] and π is increasing in μ. That is, the less forgetful the users are, the more profit the firm earns. This is the case in which the exploitation value is less than the compensation for the anticipated loss from default bias.

The firm prefers the policy without auto-renewal (to the alternative policy) if $\bar{v} < c$, since the difference $\pi - \tilde{\pi}$ becomes negative. Notice that $\tilde{\pi}$ is increasing in μ. When $\mu = 1$, $\tilde{\pi} = \bar{v} - c + q\frac{\delta}{1-\delta}(v_H - c)$, which is assumed to be positive. Therefore, it is always profitable to attract a new user under the policy without auto-renewal, as long as the probability that she remembers to re-enroll is high enough.

The discussion above is summarized by the following theorem:

Theorem 6. *The monopoly firm adopts the policy with auto-renewal policy if the expected valuation $\bar{v} \geq c$. In particular, it prefers more forgetful users when $v_L > c$, and prefers less forgetful users when $v_L < c$.*

The firm adopts the policy without auto-renewal if the expected valuation $\bar{v} < c$ and μ *is large enough so that the firm earns positive expected profit.*

Otherwise, the firm does not use promotion to attract new users.

[14]Since π is increasing in μ, π is bounded from below by $\underline{\pi} = \frac{1}{1-\delta}(\bar{v} - c) \geq 0$, which is the expected total profit when $\mu = 0$.

Discussion

The effect of default bias is still valid under different distribution of user types. In reality, those firms who offers free trials to consumers with auto-membership renewals are usually those with low marginal cost, such as Amazon Music, HBO, etc. That might explain why promotion with auto-renewal prevails in industries such as online streaming service.

While the assumption that the firm cannot commit to any future prices is reasonable and simplifies the analysis, we would have the same result in the following alternative setup: The firm has an existing pool of users and wants to attract new users. It cannot charge different prices to different users, except for the new-user benefit.

Finally, the above analysis relies on the assumption that μ is observable to the monopolist. The results are approximately the same when μ is private information, if there is little variation in μ in the population. However, there is an adverse selection problem when the variation in μ is large. For a given promotional price, the monopolist can only attract those who are less forgetful under the auto-renewal policy. This pool of users will not stay as subscribers for very long if they are low types. As a result, the firm's ability to exploit them is limited, which may not cover the initial compensation given to the new users due to the potential forgetfulness. The firm may design different contracts for different groups of potential users in terms of their forgetfulness.

2.6 Appendix: Proof of Theorem 1

The goal of this appendix is to construct the domain of preferences, Π, and show that there is a homeomorphism $f : \Pi \to \Delta\left(\{(c, z, \mu) \in C \times \mathcal{K}(\Pi) \times \Pi : \mu \in z\}\right)$.

2.6.1 Mathematical Facts

Let X be a metric space. The set $\mathcal{K}(X)$ is the collection of all nonempty compact subsets of X. Endowed with the Hausdorff metric, the set $\mathcal{K}(X)$ itself is a metric space. If X is compact, then $\mathcal{K}(X)$ is compact.

Let $\Delta(X)$ be the set of all measures on the Boral $\sigma-$algebra of X. Endow it with the weak topology. If X is compact, $\Delta(X)$ is also compact and metrizable. For any metric space X, let $\mathcal{B}(X)$ be the Borel $\sigma-$algebra of X.

For any sequence of metric spaces $\{X_t\}$, endow $\times_{t=1}^{\infty} X_t$ with the product topology. This topology is also metrizable, and if each X_t is compact, $\times_{t=1}^{\infty} X_t$ is also compact.

2.6.2 Domain Construction

Let C be a compact metric space of the possible consumption in each period. Let $\Pi_1 = \Delta(C)$. An element $\mu_1 \in \Pi_1$ is a one-period lottery on consumption. Inductively define

$$D_t = \{(c, z_t, \mu_t) \in C \times \mathcal{K}(\Pi_t) \times \Pi_t : \mu_t \in z_t\}$$

for $t \geq 1$, and

$$\Pi_t = \{\mu_t \in \Delta(C \times \mathcal{K}(\Pi_{t-1}) \times \Pi_{t-1}) : \mu_t(D_{t-1}) = 1\}$$

for $t \geq 2$.

A t-peiod lottery $\mu_t \in \Pi$ randomly yields a current consumption c, a continuation $(t-1)$-period menu z_{t-1}, and a default $(t-1)$-period lottery μ_{t-1}. With probability 1, the default lottery is contained in the continuation menu $\mu_{t-1} \in z_{t-1}$, which reflects the interpretation of a default. Note that the inductive construction guarantees that the continuation menu and lottery also have this default interpretation.

An element of $\times_{t=1}^{\infty}\Pi_t$ is a sequence of finite-period lotteries. Only the consistent sequences in Π^* could be the infinite-horizon lotteries with defaults. That is, the t-period lottery π_t in the sequence induces a $(t-1)$-lottery that is equal to $(t-1)$-period lottery π_{t-1} in the sequence. The formal definition is given later in section 2.1.

Let Π be the consistent subset of $\times_{t=1}^{\infty}\Pi_t$. This is the set of all lotteries considered in this paper, and the domain of the preferences in this model.

Consistent Sequences

Define the following functions:

- When $t = 1$, $G_1 : C \times \mathcal{K}(\Pi_1) \times \Pi_1 \to C$ is given by $G_1(c, z_1, \mu_1) = c$.
- When $t = 1$, $F_1 : \Delta(C \times \mathcal{K}(\Pi_1) \times \Pi_1) \to \Delta(C)$ is given by $F_1(\mu_2)(E) = \mu_2(G_1^{-1}(E))$ for every $E \in \mathcal{B}(C)$.
- When $t = 1$, $\bar{F}_1 : \mathcal{K}(\Delta(C \times \mathcal{K}(\Pi_1) \times \Pi_1)) \to \mathcal{K}(\Delta(C))$ is given by $\bar{F}_1(z_2) = \{F_1(\mu_2) : \mu_2 \in z_2\}$.
- For $t \geq 2$, $G_t : C \times \mathcal{K}(\Pi_t) \times \Pi_t \to C \times \mathcal{K}(\Pi_{t-1}) \times \Pi_{t-1}$ is given by $G_t(c, z_t, \mu_t) = (c, \bar{F}_{t-1}(z_t), F_{t-1}(\mu_t))$.
- For $t \geq 2$, $F_t : \Delta(C \times \mathcal{K}(\Pi_t) \times \Pi_t) \to \Delta(C \times \mathcal{K}(\Pi_{t-1}) \times \Pi_{t-1})$ is given by $F_t(\mu_{t+1})(E) = \mu_{t+1}(G_t^{-1}(E))$ for every $E \in \mathcal{B}(C \times \mathcal{K}(\Pi_{t-1}) \times \Pi_{t-1})$.
- For $t \geq 2$, $\bar{F}_t : \mathcal{K}(\Delta(C \times \mathcal{K}(\Pi_t) \times \Pi_t)) \to \mathcal{K}(\Delta(C \times \mathcal{K}(\Pi_{t-1}) \times \Pi_{t-1}))$ is given by $\bar{F}_t(z_{t+1}) = \{F_t(\mu_{t+1}) : \mu_{t+1} \in z_{t+1}\}$.

To interpret these functions, consider a realization (c, z_t, μ_t) of a lottery $\mu_{t+1} \in \Pi_{t+1}$. The function G_t maps it to the corresponding realization $(c, \bar{F}_{t-1}(z_t), F_{t-1}(\mu_t))$

that is one period shorter, by "chopping off" the last period. Given this transformation, any $(t+1)$-period lottery μ_{t+1} induces its t-period counterpart in a consistent way, and consequently any $(t+1)$-period menu induces its t-period counterpart, too.

Definition 2.6.1. The set of lotteries is $\Pi = \{\{\mu_t\} \in \times_{t=1}^{\infty} : \mu_t = F_t(\mu_{t+1}) \text{ for all } t \geq 1\}$.

For any lottery $\{\mu_t\} \in \Pi$, μ_t is interpreted as its t-period snapshot. The consistency requirement guarantees that all the snapshots are compatible with each other.

Compactness of Π

The compactness of the domain is itself useful, since given a continuous utility representation, which is the case in this paper, there exists a worst lottery as a result of compactness.

The following lemmas are going to be useful in the proof.

Lemma 2.6.1. *Let $Y \subset X$ be two metric spaces and suppose Y is closed. Then there is a homeomorphism $g : \{\mu \in \Delta(X) : \mu(Y) = 1\} \to \Delta(Y)$.*

Proof. Let g be given by $\nu = g(\mu)$ with $\nu(E) = \mu(E)$ for all $E \in \mathcal{B}(Y)$.

To see g is one-to-one, suppose $\mu, \mu' \in \{\mu \in \Delta(X) : \mu(Y) = 1\}$ and $\mu \neq \mu'$. There exists $E \in \mathcal{B}(X)$ such that $\mu(E) \neq \mu'(E)$. Since $\mu(Y) = 1$ and $\mu'(Y) = 1$, it follows $\tilde{E} := E \cap Y \neq \emptyset$, and $\mu(\tilde{E}) \neq \mu'(\tilde{E})$. Therefore, $g(\mu) \neq g(\mu')$.

To see g is onto, for any $\nu \in \Delta(Y)$, define $\mu(E) = \nu(E \cap Y)$ for all $E \in \mathcal{B}(X)$. It follows from definition that $\mu(Y) = 1$, and $g(\mu) = \nu$.

To see g is continuous, let $\{\mu_n\}$ be a sequence of measure that weakly converges to μ, and denote $\nu_n = g(\mu_n)$ and $\nu = g(\mu)$. Show that $\nu_n \to \nu$. For any real valued bounded and continuous function h defined on Y, let $\hat{h}$ be a continuous and bounded function on X with $\hat{h} = h$ on Y by Tietze extension theorem. Then

$$lim \int_Y h d\nu_n = lim \int_X \hat{h} d\mu_n = \int_X \hat{h} d\mu = \int_Y h d\nu.$$

To see g^{-1} is continuous, let $\{\nu_n\}$ be a sequence of measure that weakly converges to ν, and denote $\mu_n = g^{-1}(\nu_n)$ and $\mu = g^{-1}(\nu)$. Show that $\mu_n \to \mu$. For any real valued bounded and continuous function h defined on X,

$$lim \int_X h d\mu_n = lim \int_Y h d\nu_n = \int_Y h d\nu = \int_X h d\mu.$$

Conclude that g is a homeomorphism. □

Lemma 2.6.2. *Let X, Y be compact metric spaces, and $g : X \to Y$ be a continuous function. Then, (i) $\bar{g} : \mathcal{K}(X) \to \mathcal{K}(Y)$ defined by $\bar{g}(A) = \{g(a) : a \in A\}$ is also continuous, and (ii) If g is a bijection, it is a homeomorphism.*

Lemma 2.6.3. *Suppose Π_t is compact, then $D_t = \{(c, z_t, \mu_t) \in C \times \mathcal{K}(\Pi_t) \times \Pi_t : \mu_t \in z_t\}$ is also compact.*

Proof. Let $\{(c^n, z_t^n, \mu_t^n)\}$ be a sequence in D_t. Since $C \times \mathcal{K}(\Pi_t) \times \Pi_t$ is compact, there is a convergent subsequence with limit $(c, z_t, \mu_t) \in C \times \mathcal{K}(\Pi_t) \times \Pi_t$. It remains to prove that the limit $\mu_t \in z_t$.

Suppose $\mu_t \notin z_t$. Since z_t is compact, there is $\epsilon > 0$ so that $d(\mu_t, z_t) \geq 2\epsilon$. Since $\mu_t^n \to \mu_t$, there exists N so that $n \geq N$ implies $d(\mu_t, \mu_t^n) \leq \epsilon$. This implies $d(\mu_t^n, z_t) \geq \epsilon$ for all $n \geq N$. Note that $d_H(z_t, z_t^n) \geq sup_{\nu \in z_t^n} d(\nu, z_t) \geq d(\mu_t^n, z_t) \geq \epsilon$ for all $n \geq N$. Contradiction. □

Given the lemmas above, one can finally establish the compactness of the domain.

Lemma 2.6.4. *The set* Π *is compact.*

Proof. Note that G_1, F_1 are continuous. By lemma 2.6.2, $\bar{F}_1$ is continuous. Observe by induction that F_t and $\bar{F}_t$ are continuous for all $t \geq 1$. Suppose that each Π_t is compact. Then $\times_{t=1}^{\infty}\Pi_t$ is also compact. Continuity of those functions implies that Π is compact.

Since Π_1 is compact, it suffices to prove the following claim: If Π_t is compact, then Π_{t+1} is also compact. To show this, note that the compactness of Π_t and lemma 2.6.3 imply that D_t is compact. Therefore, $\Delta(D_t)$ is compact. Since D_t is a compact subset of $C \times \mathcal{K}(\Pi_t) \times \Pi_t$, lemma 2.6.1 implies that there is a homeomorphism between $\Delta(D_t)$ and $\Pi_{t+1} = \{\mu_{t+1} \in \Delta(C \times \mathcal{K}(\Pi_t) \times \Pi_t) : \mu_t(D_t) = 1\}$. The compactness of $\Delta(D_t)$ then suggests that Π_{t+1} is compact. □

2.6.3 Recursive Expression of Lotteries

From a recursive perspective, a lottery randomly yields a current consumption c, a continuation menu of lotteries and a default lottery. This suggest that a lottery $\{\mu_t\} \in \Pi$ can be seen as probability measure on $C \times \mathcal{K}(\Pi) \times \Pi$. Due to the default interpretation, the default lottery should be an element of the menu. The following theorem formalizes this intuition.

Theorem 2.6.5. *There exists a homeomorphism*

$$f : \Pi \to \{\mu \in \Delta(C \times \mathcal{K}(\Pi) \times \Pi) : \mu(D) = 1\},$$

where

$$D = \{(c, z, \mu) \in C \times \mathcal{K}(\Pi) \times \Pi : \mu \in z\}.$$

To get the intuition behind lemma 2.6.5, pick an infinite horizon decision problem

$z = \{z_t\} \in Z$. Consider a consistent sequence of finite-period choices $\{y_t\}$, with $y_t = (c_t, z_{t-1}, z'_{t-1}) \in z_t$. Consistency requires that

$$(c_t, z_{t-1}, z'_{t-1}) = y_t = F_t(y_{t+1}) = F_t((c_{t+1}, z_t, z'_t)) = (c_{t+1}, F_{t-1}(z_t), F_{t-1}(z'_t)),$$

which implies $c_t = c_1$ for all $t \geq 1$, $z_{t-1} = F_{t-1}(z_t)$ and $z'_{t-1} = F_{t-1}(z'_t)$ for all $t \geq 2$. This suggests that one can write z as a common first period consumption c_1 and two consistent sequences of continuation problems $\{z_t\}$ and $\{z'_t\}$ interpreted as continuation problems.

Proof of the Theorem

This proof largely replicates the one given in Gul and Pesendorfer (2004). Similar notations are used here for convenient comparison.

Definition 2.6.2. Let $Y_1 = \Delta(C)$ and $Y_t = \Delta(C \times (\times_{k=1}^{t} \mathcal{K}(\Pi_k)) \times (\times_{k=1}^{t} \Pi_k))$ for $t \geq 2$. A sequence $\{\hat{\mu}_t\} \in \times_{t=1}^{\infty} Y_t$ is called Kolmogorov consistent if the marginal distribution

$$marg_{C \times (\times_{k=1}^{t-1} \mathcal{K}(\Pi_k)) \times (\times_{k=1}^{t-1} \Pi_k)} \hat{\mu}_t = \hat{\mu}_{t-1}.$$

Let Y^{KC} be the Kolmogorov consistent subset of $\times_{t=1}^{\infty} Y_t$.

Lemma 2.6.6. *For every $\{\hat{\mu}_t\} \in Y^{KC}$, there exists a unique probability measure $\mu \in \Delta(C \times (\times_{k=1}^{\infty} \mathcal{K}(\Pi_k)) \times (\times_{k=1}^{\infty} \Pi_k))$ such that $marg_{C \times (\times_{k=1}^{t} \mathcal{K}(\Pi_k)) \times (\times_{k=1}^{t} \Pi_k)} \mu = \hat{\mu}_t$. Let ψ be this mapping. Then ψ is a homeomorphism.*

Note that the first statement in lemma 2.6.6 is the Kolmogorov existence theorem. This is the analog of lemma 3 in Gul and Pesendorfer (2004) and one can also refer to Brandenburger and Dekel (1993) for the proof of the second statement. Hence the proof is omitted here.

Definition 2.6.3. Let Y^c be the set of Kolmogorov consistent sequences that also satisfies the consistency requirement and default interpretation. Formally,

$$Y^c = \{\{\hat{\mu}_t\} \in Y^{KC} : \hat{\mu}_t(\hat{D}_t) = 1\},$$

where

$$\hat{D}_t = \{(c, z_1, \cdots, z_t, \mu_1, \cdots, \mu_t) \in C \times (\times_{k=1}^t \mathcal{K}(\Pi_k)) \times (\times_{k=1}^t \Pi_k) : \\ z_k = \bar{F}_k(z_{k+1}),\, \mu_k = F_k(\mu_{k+1}),\; \mu_k \in z_k \text{ for all } k = 1, \cdots, t\}.$$

Lemma 2.6.7. *There is a homeomorphism $\phi : \Pi \to Y^c$.*

Proof. Let m_0 and m_1 be identity functions on C and D_1. For $t \geq 2$, let $m_t : D_t \to \hat{D}_t$ be given by $(\hat{c}, \hat{z}_1, \cdots, \hat{z}_t, \hat{\mu}_1, \cdots, \hat{\mu}_t)$, where $\hat{c} = c$, $\hat{z}_t = z_t$, $\hat{\mu}_t = \mu_t$, $\hat{z}_{t-k} = \bar{F}_{t-k}(\hat{z}_{t-k+1})$, and $\hat{\mu}_{t-k} = F_{t-k}(\hat{\mu}_{t-k+1})$.

Let π_0 and π_1 be identity functions on C and $\hat{D}_1$. For $t \geq 2$, let $\pi_t : \hat{D}_t \to D_t$ be given by $\pi_t(c, z_1, \cdots, z_t, \mu_1, \cdots, \mu_t) = (c, z_t, \mu_t)$.

Observe that m_t and π_t are continuous for all t. Note also $m_t(D_t) = \hat{D}_t$ and $\pi_t(\hat{D}_t) = D_t$ due to the consistency requirement. Hence m_t and π_t are the inverse function of each other.

Now pick any $\{\mu_t\} \in \Pi$. By definition, $\mu_t \in \Pi_t$. By lemma 2.6.1, μ_t can be seen as a probability measure in $\Delta(D_t)$. Define $\hat{\mu}_t$ by $\hat{\mu}_t(E) = \mu_t(m_{t-1}^{-1}(E))$ for all $E \in \hat{D}_{t-1}$. The measure $\hat{\mu}_t$ can be seen as an element of Y_t. Consider the sequence $\{\hat{\mu}_t\}$. Consistency of $\{\mu_t\}$ guarantees that $\{\hat{\mu}_t\} \in Y^{KC}$ and satisfies the consistency part of $\hat{D}_{t-1}$ with probability 1. Moreover, $\mu_t(D_t) = 1$ implies that the default part of $\hat{D}_{t-1}$ is also satisfied with probability 1. Therefore $\{\hat{\mu}_t\} \in Y^c$. Let ϕ be this mapping. Observe that π is one-to-one.

For each $\{\hat{\mu}_t\} \in Y^c$, define $\mu_1 = \hat{\mu}_1$ and $\mu_t(E) = \hat{\mu}_t(\pi_{t-1}^{-1}(E))$ for every $E \in \mathcal{B}(D_t)$. Observe that $\phi(\{\mu_t\}) = \{\hat{\mu}_t\}$ so that ϕ is onto.

Finally, let h and $\hat{h}$ be real-valued continuous and bounded functions on D_t and $\hat{D}_t$. Observe that $\int \hat{h} d\hat{\mu}_t = \int \hat{h} \circ m_{t-1} d\mu_t$ and $\int h d\mu_t = \int h \circ \pi_{t-1} d\hat{\mu}_t$, with μ_{t-1} and π_{t-1} being also continuous. Therefore the mapping ϕ and ϕ^{-1} are both continuous.

Conclude that ϕ is a homeomorphism. □

Lemma 2.6.8. *$\psi(Y^c) = \{\mu \in \Delta(C \times (\times_{k=1}^{\infty}\mathcal{K}(\Pi_k)) \times (\times_{k=1}^{\infty}\Pi_k)) : \mu(D) = 1\}$, where $\hat{D} = \{(c, \{z_t\}, \{\mu_t\}) \in C \times (\times_{k=1}^{\infty}\mathcal{K}(\Pi_k)) \times (\times_{k=1}^{\infty}\Pi_k) : z_t = \bar{F}_t(z_{t+1}),\ \mu_t = F_t(\mu_{t+1}),\ and\ \mu_t \in z_t,\ for\ all\ t \geq 1\}$.*

Proof. Define $\Gamma_t = \{(c, \{z_\tau\}, \{\mu_\tau\}) \in C \times (\times_{k=1}^{\infty}\mathcal{K}(\Pi_k)) \times (\times_{k=1}^{\infty}\Pi_k) : z_\tau = \bar{F}_\tau(z_{\tau+1}),\ \mu_\tau = F_\tau(\mu_{\tau+1}),\ \text{and}\ \mu_\tau \in z_\tau,\ \text{for all}\ \tau \leq t\}$.

For any $\{\hat{\mu}_t\} \in Y^c$, let $\mu = \psi(\{\hat{\mu}_t\})$. Since $\mu(\gamma_t) = \hat{\mu}_t(\Gamma_t) = 1$ for all t. Then $\mu(D) = \mu(\cap\Gamma_t) = lim \mu(\Gamma_t) = 1$. If $\mu(D) = 1$, then $\mu(\Gamma_t) = 1$ for all t. Hence $\{\hat{\mu}_t\} \in Y^c$. □

Lemma 2.6.9. *Define $D = \{(c, z, \{\mu_t\}) \in C \times \mathcal{K}(\Pi) \times \Pi : \{\mu_t\} \in z\}$. There is a homeomorphism between $\hat{D}$ and D, and therefore a homeomorphism $\zeta : \Delta(\hat{D}) \to \Delta(D)$.*

Proof. For any $\{z_t\} \in \{\{z'_t\} \in \times_{t=1}^{\infty}\mathcal{K}(\Pi) : z_t = \bar{F}_t(z_{t+1})\}$, define $\xi(\{z_t\}) = \{\{\mu_t\} \in Y^c : \mu_t \in z_t \text{ for all } t \geq 1\}$. Lemma 6 in Gul and Pesendorfer (2004) shows that $\xi : \{\{z'_t\} \in \times_{t=1}^{\infty}\mathcal{K}(\Pi) : z_t = \bar{F}_t(z_{t+1})\} \to \mathcal{K}(\Pi)$ is a homeomorphism.

Pick any $(c, \{z_t\}, \{\mu_t\}) \in \hat{D}$. Define $\tilde{\zeta}(c, \{z_t\}, \{\mu_t\}) = (c, \xi(\{z_t\}), \{\mu_t\})$. By definition, $\tilde{\zeta}(c, \{z_t\}, \{\mu_t\}) \in D$. Note also that $\tilde{\zeta}$ is one-to-one. For any $(c, z, \{\mu_t\}) \in D$, verify that $\tilde{\zeta}(c, \xi^{-1}(z), \{\mu_t\}) = (c, z, \{\mu_t\})$ with $(c, \xi^{-1}(z), \{\mu_t\}) \in \hat{D}$. Finally, observe that $\tilde{\zeta}$ and $\tilde{\zeta}^{-1}$ are continuous. Therefore, $\tilde{\zeta} : \hat{D} \to D$ is a homeomorphism.

As a result, there is a homeomorphism $\zeta : \Delta(\hat{D}) \to \Delta(D)$. $\square$

Proof of the theorem. Lemma 2.6.7, 2.6.8 and 2.6.1 suggest that there is a homeomorphism $g \circ \psi \circ \phi : \Pi \to \Delta(\hat{D})$. Lemma 2.6.1 also suggests that $g^{-1}(\Delta(D)) = \Delta\{(c, z, \mu) \in C \times \mathcal{K}(\Pi) \times \Pi : \mu \in z\}$. Therefore, by lemma 2.6.9, the mapping $f = g^{-1} \circ \zeta \circ g \circ \psi \circ \phi$ is the desired homeomorphism. $\square$

2.7 Appendix: Proof of the Main Result

The main result is proved by taking several steps. In step 0, the goal is to use axioms A1 – A5 to get a general representation of the form

$$V(p) = \int \left[u(c) + \max_{q \in z} \; w(\{q, r\}, r) \right] \mathrm{d}p(c, z, r),$$

without touching the default bias.

In addition to A1 – A5, if the preference also satisfies A6 and A9, then Lemma 3 follows. This characterizes the willingness to switch and get the representation below. These are shown in Step1.

$$V(p) = \int \left[u(c) + \max_{q \in z : F(q) \geq F(r) \forall F \in \mathcal{F}} \; w(\{q, r\}, r) \right] \mathrm{d}p(c, z, r).$$

By adding the rest of the axioms, Step 3 characterizes the limited ability to switch and finishes the proof of Theorem 2.

The last subsection proves the two corollaries by strengthening A11 and A12, respectively.

2.7.1 Step 0

Axioms A1 – A5 are assumed in this step.

Recall that the set of all lotteries Π can be identified by the set $\Delta(D)$, where $D = \{(c, z, r) \in C \times \mathcal{K}(\Pi) \times \Pi : r \in z\}$ is the set of all possible realizations of the lotteries. Since $\Delta(D)$ is a mixture space, Lemma 2.7.1 is simply the expected utility theorem.

Lemma 2.7.1. *Suppose $\succsim$ satisfies Weak Order, Continuity and Independence (A1 – A3). There exists a continuous function $v : D \to \mathbb{R}$, such that the function $V : \Pi \to \mathbb{R}$ defined by*

$$V(p) = \int v(c, z, r)\, \mathrm{d}p(c, z, r)$$

represents the preference $\succsim$ on Π. Moreover, if $\tilde{V}(p) = \int \tilde{v}(c, z, r)\, \mathrm{d}p(c, z, r)$ also represents the preference $\succsim$, then there exist $a \in \mathbb{R}$ and $b \in \mathbb{R}_{++}$ such that $v = a + b\tilde{v}$.

If the preference also satisfies Intertemporal Separability (A4), the next result shows that today's consumption c and tomorrow's decision problem (z, r) are separately evaluated.

Lemma 2.7.2. *Suppose that the preference $\succsim$ that is represented as in Lemma 2.7.1 also satisfies Intertemporal Separability (A4). There exists continuous functions $u : C \to \mathbb{R}$ and $w : \{(z, r) \in \mathcal{K}(\Pi) \times \Pi : r \in z\} \to \mathbb{R}$ such that the function v in the representation can be written as*

$$v(c, z, r) = u(c) + w(z, r),$$

for all $(c, z, r) \in D$. Moreover, if $\tilde{V}(p) = \int [\tilde{u}(c) + \tilde{w}(z, r)]\, \mathrm{d}p(c, z, r)$ represents the same preference, then the functions $(\tilde{u}, \tilde{w})$ must be a common affine transformation of (u, w).

Proof. Fix a degenerate lottery $(c', z', r') \in \Pi$. For any degenerate lottery $(c, z, r) \in \Pi$, Lemma 2.7.1 and A4 imply that

$$\frac{1}{2}v(c, z, r) + \frac{1}{2}v(c', z', r') = \frac{1}{2}v(c, z', r') + \frac{1}{2}v(c', z, r).$$

Equivalently,

$$v(c, z, r) = v(c, z', r') + [v(c', z, r) - v(c', z', r')] =: u(c) + w(z, r).$$

Continuity and uniqueness of u and w follow from the continuity and uniqueness of v. □

By Conditional Strategic Rationality (A5), one can further reduce the decision problem (z, r) to the decision problem with a sub-menu of size no more than than 2 and the same default r.

Lemma 2.7.3. *Suppose that the preference $\succsim$ that is represented as in Lemma 2.7.2 also satisfies Conditional Strategic Rationality (A5). The function w has the following property:*

$$w(z, r) = \max_{q \in z} \; w(\{q, r\}, r).^{15}$$

Proof. Fix a default $r \in \Pi$. It suffices to show that for any $(c, z, r) \in D$, there exists some $q^* \in z$ such that $(c, z, r) \sim (c, \{q^*, r\}, r) \iff (c, \{q^*, r\}, r) \succsim (c, \{q, r\}, r)$ for all $q \in z$.

To show that claim, observe that $A5$ implies that for (c, z_1, r) and $(c, z_2, r) \in D$, $(c, z_1, r) \succsim (c, z_2, r)$ if $z_2 \subset z_1$ and $(c, z_1 \cup z_2, r) \sim (c, z_i, r)$ for some $i = 1$ or 2.

Suppose $(c, \{q^*, r\}, r) \succsim (c, \{q, r\}, r)$ for all $q \in z$. By the second observation,

[15] With a slight abuse of notation, the set $\{q, r\}$ reduces to $\{r\}$ when $q = r$.

$(c, \{q^*, r\}, r) \succsim \bigcup_{q \in z}(c, \{q, r\}, r) = (c, z, r)$. The first observation implies that $(c, z, r) \succsim (c, \{q^*, r\}, r)$.

Suppose $(c, z, r) \sim (c, \{q^*, r\}, r)$. If there is a $q' \in z$ with $(c, \{q', r\}, r) \succ (c, \{q^*, r\}, r)$, then by the first observation $(c, z, r) \succsim (c, \{q', r\}, r) \succ (c, \{q^*, r\}, r)$. Contradiction.

To see the existence of such q^*, if $(c, z, r) \succ (c, \{q, r\}, r)$ for all $q \in z$, the second observation implies $(c, z, r) \succ \bigcup_{q \in z}(c, \{q, r\}, r) = (c, z, r)$. Contradiction. □

2.7.2 Step 1

Axioms A1 – A6 and A9 are assumed in this step.

Recall that we have defined the binary relation $p \succsim^* r$ if and only if there exist sequences $p_n \to p$ and $q_n \to q$ with $(c, \{p_n, q_n\}, q_n) \succ (c, \{q_n\}, q_n)$ for all n or $p = q$. We argued that this binary relation represents the willingness to switch. We first show that if the preference $\succsim$ also satisfies A6 and A9, then the induces binary relation $\succsim^*$ satisfies the following properties. After that we will use the additional axioms A8 and A10 to fully characterize the willingness to switch.

Lemma 2.7.4. *Assuming A1 – A6 and A9, the binary relation* $\succsim^*$ *satisfies the properties below (P1 – P3).*

Property 1 (Preorder*). *The binary relation* $\succsim^*$ *defined on* Π *is a preorder: It is reflexive and transitive.*

Property 2 (Continuity*). *For any convergent sequences* $\{p_n\} \to p$ *and* $\{q_n\} \to q$ *in* Π,

$$p_n \succsim^* q_n \text{ for all } n \implies p \succsim^* q.$$

Property 3 (Independence*). *For any* $p, q, r \in \Pi$ *and* $\lambda \in (0, 1)$,

$$p \succsim^* q \implies \lambda p + (1 - \lambda) r \succsim^* \lambda q + (1 - \lambda) r.$$

Proof. We are going to show that Continuity* is satisfied by construction, and Independence* is satisfied by using axioms A3 and A6. The reflexivity of the binary relation is guaranteed by its definition. Hence we only need to show it is transitive to claim it is a preorder, which requires axioms A5 and A9.

Continuity*. Suppose $p_n \to p$, $q_n \to q$ and $p_n \succsim^* q_n$ for all n. By definition, for each n, there exists sequences $\{p_n^k\}_k$ and $\{q_n^k\}_k$ with $p_n^k \to p_n$ and $q_n^k \to q_n$ as $k \to \infty$, and $(c, \{p_n^k, q_n^k\}, q_n^k) \succ (c, \{q_n^k\}, q_n^k)$ for each k. Consider the sequences $\{p_n^n\}_n$ and $\{q_n^n\}_n$. Observe that $p_n^n \to p$ and $q_n^n \to q$ as $n \to \infty$, $(c, \{p_n^n, q_n^n\}, q_n^n) \succ (c, \{q_n^n\}, q_n^n)$ for each n. By definition $p \succsim^* q$.

Independence*. Suppose $(c, \{p, q\}, q) \succ (c, \{q\}, q)$. For any $\lambda \in (0, 1)$ and $r \in \Pi$, A3 implies

$$\lambda(c, \{p, q\}, q) + (1 - \lambda)(c, \{r\}, r) \succ (c, \{q\}, q) + (1 - \lambda)(c, \{r\}, r),$$

which by A7 implies

$$(c, \{\lambda p + (1 - \lambda)r, \lambda q + (1 - \lambda)r\}, \lambda q + (1 - \lambda)r) \succ (c, \{\lambda q + (1 - \lambda)r\}, \lambda q + (1 - \lambda)r).$$

With this result, since $p \succsim^* q$, there are sequences $p_n \to p$ and $q_n \to q$ with $(c, \{p_n, q_n\}, q_n) \succ (c, \{q_n\}, q_n)$. It follows that for each n,

$$(c, \{\lambda p_n + (1-\lambda)r_n, \lambda q_n + (1-\lambda)r_n\}, \lambda q_n + (1-\lambda)r_n) \succ (c, \{\lambda q_n + (1-\lambda)r_n\}, \lambda q_n + (1-\lambda)r_n),$$

which shows independence*

Preorder*. Reflexivity is guaranteed by definition. It remains to show transitivity. We start from the strict part of $\succ^*$ and show $p \succ^* q$ and $q \succ^* r$ imply $p \succ^* r$. That is, $(c, \{p, q\}, q) \succ (c, \{q\}, q)$ and $(c, \{q, r\}, r) \succ (c, \{r\}, r)$ imply $(c, \{p, r\}, r) \succ (c, \{r\}, r)$. Observe that the first expression and A5 implies $(c, \{p, q, r\}, q) \succ (c, \{q\}, q)$.

Claim that $(c, \{p, r\}, r) \succsim (c, \{q, r\}, r)$. Suppose on the contrary that $(c, \{q, r\}, r) \succ (c, \{p, r\}, r)$. By A5, it follows that $(c, \{p, q, r\}, r) \succ (c, \{p, r\}, r)$. Apply A9 to get $(c, \{q\}, q) \sim (c, \{p, q, r\}, q)$, which contradicts the previous observation.

The claim, together with the second expression $(c, \{q, r\}, r) \succ (c, \{r\}, r)$, implies that $(c, \{p, r\}, r) \succ (c, \{r\}, r)$.

Now that we have shown that $\succ^*$ satisfies transitivity. One can take convergent sequences as in the proof for Independence* to conclude that $\succsim^*$ also satisfies transitivity. □

Lemma 2.7.5 (Expected Multi-Utility Theorem (Dubra et al., 2004))**.** *Let X be a compact metric space and let $\succsim^*$ be a binary relation on $\Delta(X)$. It satisfies P1 – P3 if and only if there exits a closed and convex set $\mathcal{F} \subset C(X)$ such that for each $p, q \in \Delta(X)$, $p \succsim^* q$ if and only if*

$$\int_X f \, \mathrm{d}p \geq \int_X f \, \mathrm{d}q \text{ for all } f \in \mathcal{F}.$$

Moreover, if $\mathcal{F}$ and $\mathcal{F}'$ both represent $\succsim^*$, then $\langle \mathcal{F} \rangle = \langle \mathcal{F}' \rangle$, where $\langle \mathcal{F} \rangle := cl(cone(\mathcal{F}) + \{\theta \mathbf{1}_X\}_{\theta \in \mathbb{R}})$ is the largest set of utility functions in $\mathcal{C}(X)$ that represents the same binary relation.

Proof of Lemma 3. Since $\Pi = \Delta(D)$, where D is a compact metric space. By Lemma 2.7.4 and 2.7.5, one can represent $\succsim^*$ as described in Lemma 3. □

As is argued in the main text, we can interpret $(c, \{p\}, p) \succsim (c, \{q\}, q)$ as the individual's preference over p and q in the second period. It is a weak order. The next lemma shows that the second period preference is a completion of the binary relation $\succsim^*$.

Lemma 2.7.6. $p \succsim^* q \implies (c, \{p\}, p) \succsim (c, \{q\}, q)$.

Proof. If $p \succsim^* q$, there exist $p_n \to p$, $q_n \to q$ and $(c, \{p_n, q_n\}, q_n) \succ (c, \{q_n\}, q_n)$ for each n. By A9, it follows that $(c, \{p_n\}, p_n) \succ (c, \{q_n\}, q_n)$ for each n. Then in the limit, $(c, \{p\}, p) \succsim (c, \{q\}, q)$. □

Recall that $\mathcal{U}^*(r) = \{p \in \Pi : \ p \succsim^* r\}$ is the set of lotteries that the individual is willing to switch to when r is the default. The next lemma implies that the individual ignores those lotteries in the menu that she is not willing to switch to.

Lemma 2.7.7. $(c, z, r) \sim (c, z \cap \mathcal{U}^*(r), r)$.

Proof. Suppose instead that $(c, z, r) \succ (c, z \cap \mathcal{U}^*(r), r)$. By A5, there must be $q \in z \backslash \mathcal{U}^*(r)$ with $(c, \{q, r\}, r) \sim (c, z, r) \succ (c, z \cap \mathcal{U}^*(r), r) \succsim (c, \{r\}, r)$, which implies $q \succsim^* r$. Contradiction. □

2.7.3 Step 2

All of the axioms (A1 – A12) are assumed in this step.

Lemma 2.7.8. *For any* $p, q, s \in \mathcal{U}^*(r)$ *and* $\lambda \in [0, 1]$,

$$(c, \{p, r\}, r) \succsim (c, \{q, r\}, r) \implies (c, \{\lambda p + (1-\lambda)s, r\}, r) \succsim (c, \{\lambda q + (1-\lambda)s, r\}, r).$$

Proof. By A3, $(c, \{p, r\}, r) \succsim (c, \{q, r\}, r)$ implies

$$LHS := \lambda(c, \{p, r\}, r) + (1-\lambda)(c, \{s, r\}, r) \succsim \lambda(c, \{q, r\}, r) + (1-\lambda)(c, \{s, r\}, r) := RHS.$$

It then follows from A6 that

$$LHS \sim (c, \{\lambda p + (1-\lambda)s, \lambda p + (1-\lambda)r, \lambda r + (1-\lambda)s, r\}, r). \tag{2.7.1}$$

Since $s \in \mathcal{U}^*(r)$, Lemma 2.7.6 implies that $(c, \{s\}, s) \succsim (c, \{r\}, r)$. By mixing with $(c, \{p\}, p)$, it follows

$$(c, \{\lambda p + (1-\lambda)s\}, \lambda p + (1-\lambda)s) \succsim (c, \{\lambda p + (1-\lambda)r\}, \lambda p + (1-\lambda)r). \tag{2.7.2}$$

By property Independence* of $\succsim^*$, $p, s, r \in \mathcal{U}^*(r)$ implies that $\lambda p + (1-\lambda)s$, $\lambda p + (1-\lambda)r \in \mathcal{U}^*(r)$. A9 implies

$$(c, \{\lambda p + (1-\lambda)s, r\}, r) \sim (c, \{\lambda p + (1-\lambda)s\}, \lambda p + (1-\lambda)s),$$

$$(c, \{\lambda p + (1-\lambda)r, r\}, r) \sim (c, \{\lambda p + (1-\lambda)r\}, \lambda p + (1-\lambda)r),$$

which, combined with Expression 2.7.2 and A5, suggests that one can drop the lottery $\lambda p + (1-\lambda)r$ from the menu of the right-hand-side of Expression 2.7.1. One can similarly drop the other lottery $\lambda r + (1-\lambda)s$, so that

$$LHS \sim (c, \{\lambda p + (1-\lambda)s, r\}, r).$$

Go through the same steps to get

$$RHS \sim (c, \{\lambda q + (1-\lambda)s, r\}, r),$$

which concludes the proof. □

Lemma 2.7.9. *For any $r \in \Pi$, there exist $\alpha(r) \in \mathbb{R}$ and $\beta(r) \in \mathbb{R}_{++}$ such that for* *any* $p \in \mathcal{U}^*(r)$,

$$w(\{p, r\}, r) = \alpha(r) + \beta(r)V(p).$$

Proof. Begin by the observation that the ranking over the set $\{(c, \{p, r\}, r) : p \in \Pi\}$ is represented by the continuous function $w(\{p, r\}, r)$. Lemma 2.7.8 shows that the ranking is affine. Therefore, this ranking can be represented by an expected utility function.

For any $p, q \in \mathcal{U}^*(r)$, Stationarity (A8) implies $p \succsim q \Longleftrightarrow (c, \{p\}, p) \succsim (c, \{q\}, q)$. Using Dominance (A10) and the fact that $p, q \in \mathcal{U}^*(r)$[16], we have

$$p \succsim q \Longleftrightarrow (c, \{p, r\}, r) \succsim (c, \{q, r\}, r).$$

Since both V and $w(\{\cdot, r\}, r)$ are expected utility representations on $\mathcal{U}^*(r)$, they must be affine transformations of each other. $\square$

Combining the results, if the preference $\succsim$ satisfies A1 – A6 and A8 – A10 , it can be represented by

$$V(p) = \int \left[u(c) + \max_{q \in z:\, q \succsim^* r} (\alpha(r) + \beta(r)V(q)) \right] \mathrm{d}p(c, z, r), \tag{2.7.3}$$

where $\succsim^*$ is represented as in Lemma 3.

Lemma 2.7.10. *For any* $p \in \mathcal{U}^*(r)$, *there exists* $\mu \in (0, 1]$ *such that*

$$w(\{p, r\}, r) = \mu V(p) + (1 - \mu)V(r).$$

[16]One may need to take a limiting sequence here.

Proof. As a result of Default Separability (A11), for any $p, q \in \mathcal{U}^*(r_1) \cap \mathcal{U}^*(r_2)$,

$$\frac{1}{2}(c, \{p, r_1\}, r_1) + \frac{1}{2}(c, \{q, r_2\}, r_2) \sim \frac{1}{2}(c, \{p, r_2\}, r_2) + \frac{1}{2}(c, \{q, r_1\}, r_1).$$

That is,

$$\beta(r_1)V(p) + \beta(r_2)V(q) = \beta(r_2)V(p) + \beta(r_1)V(q).$$

Therefore, $\beta(r) = b$ for all $r \in \Pi$.

It remains to pin down the value $\alpha(r)$. For any $r, r' \in Pi$, one can pick a lottery p such that $(c, \{p, r\}, r) \succ (c, \{r\}, r)$ and $(c, \{p, r'\}, r') \succ (c, \{r'\}, r')$. Then

$$\begin{aligned} V(r) \geq V(r') &\implies r \succsim r' \\ &\implies (c, \{r\}, r) \succsim (c, \{r'\}, r') \\ &\implies (c, \{p, r, r'\}, r) \succsim (c, \{p, r, r'\}, r') \\ &\implies (c, \{p, r\}, r) \succsim (c, \{p, r'\}, r') \\ &\implies \alpha(r) \succsim \alpha(r'), \end{aligned}$$

where the first line invokes the representation, the second line uses Stationarity (A8), the third line uses Default Consistency (A12), the fourth line uses Dominance (A10), and the last line uses the result from Lemma 2.7.9. By Timing Indifference (A6), we also know that for any $\lambda \in [0, 1]$,

$$\lambda(c, \{p, r, r'\}, r) + (1 - \lambda)(c, \{p, r, r'\}, r) \sim (c, \{p, r, r'\}, \lambda r + (1 - \lambda)r'),$$

which implies that $\alpha(\lambda r + (1 - \lambda)r') = \lambda\alpha(r) + (1 - \lambda)\alpha(r')$. That is, α is a linear functional. Since both V and α are linear functionals defined on the convex subset Π of a vector space, it follows (from, for example, Corollary B.3. in Ghirardato et al.

(2004)) that

$$\alpha(r) = a' + b'V(r),$$

where $a \in \mathbb{R}$ and $b' \geq 0$. One can normalize the coefficients to get the desired expression. □

Proof of Theorem 2. Combine the results we have proved above and normalize the coefficients to get the following representation,

$$V(p) = \int \left[u(c) + \delta \left(\max_{q \in z:\ F(q) \geq F(r)\ \forall F \in \mathcal{F}} \mu V(q) + (1-\mu) V(r) \right) \right] \mathrm{d}p(c, z, r).$$

We finally use Nontriviality (A7) to conclude that $\delta \in (0, 1)$. This is again a standard result. See, for example, Sarver (2017). □

2.7.4 Special Cases

One of the special case, DB-LWS, is straightforward to get.

Proof of Corollary 2.1, DB-LWS. Observe first that A12* implies A12. Therefore, we still have the DB-representation. Recall that Lemma 2.7.10 says that $w(\{p, r\}, r) = \mu V(p) + (1-\mu)V(r)$, for some $\mu \in (0, 1]$. Default Irrelevance (A12*) implies, for any q, r, r' such that $(c, \{q, r\}, r) \succ (c, \{r\}, r)$ and $(c, \{q, r'\}, r') \succ (c, \{r'\}, r')$, $(c, \{p, r\}, r) \sim (c, \{p, r'\}, r')$ regardless of the preference over r and r'. This implies $\mu = 1$. That is the DB-LWS representation. □

Below we show the other special case, DB-LAS.

Proof of Corollary 2.2, DB-LAS. Again, since A11* implies A11, the main representation is still valid. We only need to show that $\mathcal{F} = \{V\}$ to get DB-LAS. To show this, it suffices to prove that $\succsim^*$ is complete.

Suppose not. There are lotteries r and p with $\neg p \succsim^* r$ and $\neg r \succsim^* p$. By Lemma 3, there exists $F, G \in \mathcal{F}$ with $F(r) < F(p)$ and $G(r) > G(p)$. By continuity, there exists a lottery q, with $\neg q \succsim^* r$ and $\neg r \succsim^* q$ but $p \succ^* q$.

Observe that $\neg p \succsim^* r$ implies $(c, \{p, r\}, r) \sim (c, \{r\}, r)$. Define $r \succsim' p$ if and only if $(c, \{p, r\}, r) \sim (c, \{r\}, r)$. By construction of p, q, r, one has $r \succsim' p$, $p \succsim' r$, $r \succsim' q$, $q \succsim' r$, $p \succsim' q$, $\neg q \succsim' p$.

Notice that these relations suggest that $\succsim'$ cannot be transitive: If $\succsim'$ is transitive, $q \succsim' r$ and $r \succsim' p$ imply $q \succsim' p$, which contradicts $\neg q \succsim' p$. Therefore, to show completeness of $\succsim^*$, it suffices to show that $\succsim'$ is indeed transitive!

To prove transitivity, suppose $p \succsim' q$, $q \succsim' r$, but $r \succ' p$. The first and third conditions imply that $(c, \{p, r\}, p) \succ (c, \{p\}, p) \sim (c, \{p, q\}, p)$. Conditional Strategic Rationality A5 implies $(c, \{p, q, r\}, p) \sim (c, \{p, r\}, p) \succ (c, \{p, q\}, p)$. By Strong Default Separability A11*, it follows that $(c, \{p, q, r\}, q) \succ (c, \{p, q\}, q)$. Conditional Strategic Rationality again implies $(c, \{q, r\}, q) \succ (c, \{q\}, q)$, which contradicts $q \succsim' r$. This establishes the transitivity of $\succsim'$. Conclude that $\succsim^*$ is complete.

Once completeness of $\succsim^*$ is proven, Lemma 3 implies that $\mathcal{F}$ is singleton. It follows trivially from the proof of the main representation that $\mathcal{F} = \{V\}$.

□

2.8 Appendix: Proof of Comparative Statics

We only prove the first part of the Theorem 4, since the second part is rather straightforward.

Proof of Part 1. Suppose on the contrary that there exists a function $F^* \in \langle \mathcal{F}_2 \cup \{V_1\} \rangle$ such that $F^* \notin \langle \mathcal{F}_1 \cup \{V_2\} \rangle$. By definition, this implies that there exist lotteries

$p, r \in \Pi$ such that $F(p) \geq F(r)$ for all $F \in \langle \mathcal{F}_1 \cup \{V_2\} \rangle$ and $F^*(p) < F^*(r)$. We can without loss assume that $F^* \in \langle \mathcal{F}_2 \rangle$.[17] As a result, we have $(c, \{p, r\}, r) \sim_2 (c, \{r\}, r)$.

We can also without loss assume that $V_2(p) > V_2(r)$.[18] Since $F(p) \geq F(r)$ for all $F \in \langle \mathcal{F}_1 \cup \{V_2\} \rangle$, individual 1 has the preference $(c, \{p, r\}, r) \succ_1 (c, \{r\}, r)$.[19] This violates the behavioral condition that individual 1 is less willing to switch than individual 2.

Now suppose $\langle \mathcal{F}_2 \cup \{V_1\} \rangle \supset \langle \mathcal{F}_2 \cup \{V_1\} \rangle$. We show that $p \succ_i r$ for $i = 1, 2$ and $(c, \{p, r\}, r) \succ_1 (c, \{r\}, r)$ imply $(c, \{p, r\}, r) \succ_2 (c, \{r\}, r)$. Since $(c, \{p, r\}, r) \succ_1 (c, \{r\}, r)$ and $V_2(r) > V_2(r)$, it implies that $F(p) \geq F(r)$ for all $F \in \langle \mathcal{F}_1 \cup \{V_2\} \rangle$. Since $\langle \mathcal{F}_2 \rangle$ is a subset, we must have $F(p) \geq F(r)$ for all $F \in \langle \mathcal{F}_2 \rangle$. This, together with $V_2(p) > V_2(r)$, concludes the proof.

$\square$

2.9 Appendix: Proof of the Asset Replacement Model

2.9.1 Analytical Analysis

Before we look at the asset replacement problem, we show the following result, which will be useful later.

---

[17] If $F^* \notin \langle \mathcal{F}_2 \rangle$, it must be a linear combination of some functions in $\langle \mathcal{F}_2 \rangle$ and V_1. But we know that $V_1(p) \geq V_1(r)$, which implies that there must be a function in $\langle \mathcal{F}_2 \rangle$ with the value at p less than the value at r.

[18] If this is not the case, one can pick a lottery q with $V_2(q) > V_2(p)$. The any mixture of p and q is strictly preferred to r by individual 2, and by taking the limit, one can get a mixture whose value according to F^* is less than the value of r according to F^*, since F^* is an affine function.

[19] One might need to take the limiting sequence $(p_n, r_n) \to (p, r)$ with $(c, \{p_n, r_n\}, r_n) \succ_1 (c, \{r_n\}, r_n)$

Lemma 2.9.1. *Given the DB-LAS representation*

$$V(p;\mu) = \int u(c) + \mu\delta \max_{q\in z} V(q) + (1-\mu)\delta V(p)\, \mathrm{d}p(c,z,q),$$

the value of the lottery p increases as μ increases.

Proof. We take the derivative with respect to μ.

$$\frac{d}{d\mu}V(p) = \frac{d}{d\mu}\int u(c) + \mu\delta \max_{q\in z} V(q) + (1-\mu)\delta V(p)\, \mathrm{d}p(c,z,q)$$

Note that this is an integral of a continuous function a compact set. Then it is dominated by some integrable function. Dominated convergence theorem allows to change the order of differentiation and integral. Therefore, to show it is positive, it suffices to show that the derivative of the integrand is positive for each realization:

$$\begin{aligned}
&\frac{d}{d\mu}\left(u(c) + \mu\delta \max_{q\in z} V(q) + (1-\mu)\delta V(p)\right) \\
=&\delta\left(\max_{q\in z} V(q) - V(p) + \mu\frac{d}{d\mu}\max_{q\in z} V(q) + (1-\mu)\frac{d}{d\mu}V(p)\right) \\
\geq&\delta\left(\mu\frac{d}{d\mu}V(q^*) + (1-\mu)\frac{d}{d\mu}V(p)\right),
\end{aligned}$$

where in the second line the first two term is positive and one can invoke the envelope theorem to get the last line (q^* is the most preferred element in z). One can iteratively plug in the expression for $\frac{d}{d\mu}V(q^*)$ and $\frac{d}{d\mu}V(p)$ to conclude. □

Now we go back to the asset replacement model. Define the following value function:

$$V_t = u_t + \mu\delta \max\{V_0 - c, V_{t+1}\} + (1-\mu)\delta V_{t+1}$$

where V_t is the value of a t-period-old asset in equilibrium. When the individual owns a t-period-old asset and has the opportunity to make an active decision, she chooses to keep the asset if $V_t > V_0 - c$ and to replace the asset otherwise.

We treat this as a game played by multi-selves, in which the t-period self is the holder of the t-period-old asset. The first step is to characterize the equilibrium of the game, and the second step is to study how the optimal replacement time is affected by μ.

The following lemma implies that the τ-period self knows that a future self will eventually replace the asset.

Lemma 2.9.2. *For any $\tau \geq 0$, there is a $t > \tau$ such that $V_t < V_0 - c$*

Proof. Note that $V_t \geq u_t + \delta V_{t+1}$ holds for all $t \geq 0$, which implies that $V_0 \geq \sum_{k=0}^{t-1} u_k \delta^k + \delta^t V_t$ holds for all $t \geq 0$. Suppose $V_t \geq V_0 - c$ for all $t > \tau$. Then the following holds for all $t \geq \tau$:

$$V_t \geq V_0 - c \geq \sum_{k=0}^{t-1} u_k \delta^k + \delta^t V_t - c.$$

Note that the assumption also implies that for all $t \geq \tau$,

$$V_t = \sum_{k=t}^{\infty} u_k \delta^k \leq u_t \sum_{k=t}^{\infty} \delta^k = \frac{u_t}{1-\delta}.$$

Therefore, for all $t \geq \tau$,

$$(1-\delta^t)\frac{u_t}{1-\delta} \geq \sum_{k=0}^{t-1} u_k \delta^k - c.$$

Note that the right-hand-side is positive in the limit by assumption, while the left-hand-side converges to 0. Contradiction. □

The following lemma suggests that the value of the asset is lower in the next period, if the individual wants to keep the asset in the current period.

Lemma 2.9.3. *If $V_t > V_0 - c$, then $V_{t+1} < V_t$.*

Proof. For any $t \geq 0$,

$$V_t - V_{t+1} = [u_t - u_{t+1}] + (1-\mu)\delta[V_{t+1} - V_{t+2}] + \mu\delta[\max\{V_0 - c, V_{t+1}\} - \max\{V_0 - c, V_{t+2}\}].$$

By the last lemma, let T be the smallest $t' > t$ such that $V_t \leq V_0 - c$. Observe that $V_{T-1} \geq V_0 - c > V_T$ by definition.

Case 1: If $T = t+1$, trivially it follows that $V_t > V_0 - c \geq V_{t+1}$ by definition.

Case 2: If $T = t+2$, then $V_t, V_{t+1} > V_0 - c \geq V_{t+2}$. It follows that $V_t - V_{t+1} = [u_t - u_{t+1}] + (1-\mu)\delta[V_{t+1} - V_{t+2}] + \mu\delta[V_{t+1} - (V_0 - c)]$ which is positive.

Case 3: If $T > t+2$, $V_t - V_{t+1} = [u_t - u_{t+1}] + \delta[V_{t+1} - V_{t+2}] \geq \delta[V_{t+1} - V_{t+2}]$. By induction, it follows that $V_t - V_{t+1} \geq \delta^{T-t-2}[V_{T-2} - V_{T-1}]$, which is again positive by case 2. □

Lemma 2.9.4. *If $V_t \leq V_0 - c$, then $V_{t+1} \leq V_0 - c$.*

Proof. Suppose on the contrary that $V_{t+1} > V_0 - c$. Then $V_0 - c \geq V_t = u_t + \delta V_{t+1} \geq u_t + \delta(V_0 - c)$. This implies $V_0 - c > \frac{1}{1-\delta}u_t$.

Now we consider V_{t+1}. Let T be the smallest $t' > t$ such that $V_{t'} \leq V_0 - c$. By Lemma 2.9.2, a future self will eventually replace the asset. Thus T is well-defined. Then we have

$$V_{t+1} = u_{t+1} + \cdots + \delta^{T-t-2}u_{T-1} + \delta^{T-t-1}(1-\mu)V_T + \delta^{T-t-1}\mu(V_0 - c).$$

This implies $V_0 - c < V_{t+1} \leq u_{t+1}\frac{1-\delta^{T-t-1}}{1-\delta} + \delta^{T-t-1}(V_0 - c)$. As a result $V_0 - c < \frac{1}{1-\delta}u_{t+1}$. This, together with the previous inequality, contradicts the condition that $u_t \geq u_{t+1}$. □

Lemma 2.9.5. *If $V_t \leq V_0 - c$, then $V_{t+1} < V_t$.*

Proof. Since $V_t \leq V_0 - c$, by Lemma 2.9.4, we have $V_{t+k} \leq V_0 - c$ for all $k = 1, 2, \cdots$. Therefore, $V_t - V_{t+1} = (u_t - u_{t+1}) + (1-\mu)\delta(V_{t+1} - V_{t+2}) = \sum_{k=0}^{\infty}(\delta(1-\mu))^k(u_{t+k} - u_{t+k+1}) > 0$ □

Lemma 2.9.3 and 2.9.5 imply that in equilibrium $\{V_t\}$ is a decreasing sequence. By Lemma 2.9.2, we know there is an integer $T \geq 1$ such that $V_{T-1} > V_0 - c \geq V_T$. This proves the existence of the equilibrium.

Now we study how V_{T-1}, $V_0 - c$ and V_T change as μ changes. For a fixed T, Lemma 2.9.1 implies that all three values decrease as μ decreases. This suggests whether the individual replace the asset earlier or later depends on which is more sensitive to the change in μ.

Proof of the Theorem. Fix any $T \geq 2$. Let $S(T)$ be the set such that $\mu \in S(T)$ implies $T(\mu) = T$. Since $V_t = u_t + \delta V_{t+1}$ for $t < T - 1$, then

$$V_0 - c = \sum_{k=0}^{T-2} u_k \delta^k + \delta^{T-1} V_{T-1} - c.$$

It follows that

$$\frac{d}{d\mu}(V_{T-1} - (V_0 - c)) = (1 - \delta^{T-1})\frac{d}{d\mu}V_{T-1} \geq 0.$$

This suggests that $S(T)$ is an interval, and $\mu < \mu'$ implies $T(\mu) \leq T(\mu')$. □

2.9.2 Numerical Analysis

In this section, we describe how to numerically find the optimal replacement time T given the parameters (u_t, c, δ, μ). We use the guess and verify strategy: we first guess a time T and compute V_{T-1}, $V_0 - c$ and V_T to verify whether this is an equilibrium. For a given T, we know that

$$\begin{aligned} V_T =& u_T + \mu\delta(V_0 - c) + (1-\mu)\delta V_{T+1} \\ =& \sum_{k=0} (\delta(1-\mu))^k u_{T+k} + \frac{\mu\delta}{1-(1-\mu)\delta}(V_0 - c). \\ V_0 =& u_0 + \delta V_1 = \sum_{k=0}^{T-1} \delta^k u_k + \delta^T V_T. \end{aligned}$$

One can solve for V_0 and V_T from the two equations. Then it is straightforward to get

$$V_{T-1} = u_{T-1} + \mu\delta(V_0 - c) + (1-\mu)\delta V_T.$$

If $V_{T-1} \leq V_0 - c$, it suggests the guess T is too large. If $V_T > V_0 - c$, it suggests the guess T is too small. These three equations allow us to find the optimal replacement time T.

Chapter 3

Attribute-Dependent Random Utility Maximization

A decision maker may be able to evaluate a product according to various attributes, but is uncertain about the right trade-offs to form an overall evaluation. In order to compare products, she needs to learn about the right trade-offs across attributes from the market that consists firms who design product lines according to average consumers. In particular, she believes that an attribute is more important when there are more alternatives that are salient in that attribute. Due to this learning process, the comparison between any pair of products may depend on other products that are present in the market. We propose an Attribute-Dependent Random Utility Maximization model that takes this learning process into consideration and characterize the representation with behavioral conditions. The representation explains violations of independence of irrelevant alternatives and violations of monotonicity, which cannot be explained by the classic models.

3.1 Introduction

A consumer may not be able to decide which is the best product to purchase among the available ones, even when she knows the full specifications of all those products. This is more likely to happen when she is not familiar with the product, so that she finds it difficult to understand the relative importance of these technical details and how to make a holistic evaluation based on them. Learning about the specifications facilitates her purchase decision.

For example, let's consider a decision maker who needs to purchase a smart phone. While she understand the performance of the phones in terms of their displays, cameras and chips, she does not understand what is the relative importance of these three attributes of the phones. However, she may believe that a phone with an excellent display is important, relative to other two attributes, if she finds that lots of phones in the market feature cutting-edge display technology.

This belief makes sense. It is costly for a firm to improve the quality of the displays of their phones and it might be the case that it becomes increasingly costly as the quality improves. Investing in display technologies despite the costs implies the importance of displays from the market perspective. While the consumer might not be familiar with the smart phone market, she is still able to infer from the market composition if she believes that her preference about smart phones is not different from the preferences of other consumers in the market. As a result, she would rationally conclude that the display technology is an important feature of smart phones. This idea is motivated by Kamenica (2008), who models the equilibrium of the market when uninformed consumers infer about their preferences from the market composition of products.

Notice that in this motivating example the consumer has a set of attributes to consider in her evaluation of alternatives, which may or may not be observable to an analyst. She also understands what is the most salient feature of a smart phone, perhaps from reading the advertisement about it. We assume that she is able to compare phones in terms of their display, camera or chips, according to the specifications of phones. However, she is unable to directly make trade-offs across these attributes and forms a holistic evaluation. Instead, she learns the relative importance of these attributes from the market composition to evaluate the phones.

To be more concrete, imagine that phone x is salient in its display and phone y is salient in its camera. When there are only these two phones in the market, our consumer believes that displays and cameras are as important as each other and chips are not important at all even though she reads about chips from the technical specifications of the phones. She evaluates phone x at

$$u(x, \{x, y\}) = \frac{1}{2}v(x, \text{display}) + \frac{1}{2}v(x, \text{camera}) + \frac{0}{2}v(x, \text{chip}).$$

That is, the value of phone x when in the market $\{x, y\}$ is the weighted average of its qualities in the three attributes. In particular, the decision weights are $\frac{1}{2}$, $\frac{1}{2}$, and $\frac{0}{2}$, which are the proportions of the products in the market that are salient in each of the attributes. In comparison, suppose now that the market has both x and y and another phone z which is also salient in its display. Then phone x is evaluated at

$$u(x, \{x, y, z\}) = \frac{2}{3}v(x, \text{display}) + \frac{1}{3}v(x, \text{camera}) + \frac{0}{3}v(x, \text{chip}),$$

where the decision weights again represents the market composition of phones in terms of their salient attributes.

We assume that the consumer purchases phone x from the market $\{x, y\}$ with probability

$$\rho(x, \{x, y\}) = \frac{\exp(u(x, \{x, y\}))}{\exp(u(x, \{x, y\})) + \exp(u(y, \{x, y\}))},$$

which is the "Luce value", $\exp(u(x, \{x, y\}))$, of phone x divided by the "total Luce value" of the market $\{x, y\}$, $\exp(u(x, \{x, y\})) + \exp(u(y, \{x, y\}))$. While we will describe the mathematical similarity between our model and Luce's random choice model (Luce (1959)) in more details in the next section, notice that the choices of our consumer are modeled as stochastic ones in order to capture the residual information that she learns from other independent sources, including reviews, past sales and word of mouth.

Despite the mathematical similarity, this model is not a Luce model. Luce's model exhibits independence of irrelevant alternatives. That is, the choice probability ratio between any two alternatives is independent of the presence of any other alternatives in the choice set. In this model, however, the choice probability between alternatives x and y in choice set A is

$$\frac{\rho(x,A)}{\rho(y,A)} = \exp(u(x,A) - u(y,A)),$$

which depends on the set A. This is because in our model the value of an alternative depends not only on its own physical characteristics but also on the composition of the choice set. The dependence on the irrelevant alternatives is the result of the fact that the consumer learns about how to evaluate the alternatives from the market. Our model allows menu effects.

Luce's model is a random utility maximization model, which assumes that a decision maker chooses the alternative that yields the largest utility from the choice set, with the utility of each alternative being stochastic. As a result, the probability of choosing alternative x from choice set A is the probability that the utility of x is larger than the utility of any of the rest alternatives in the choice set A. This choice probability decreases as the choice set becomes larger, which is often referred to as the monotonicity property. While our model also resembles the random utility models, it violates monontonicity. We use the example introduced above to demonstrate the violation of monotonicity. Suppose

$$v(x, \text{display}) = v(y, \text{camera}) = 1, \quad v(y, \text{display}) = v(x, \text{camera}) = 0.$$

One can verify that $u(x, \{x, y\}) = u(y, \{x, y\})$ and $\rho(x, \{x, y\}) = \frac{1}{2}$. Now suppose the phone z, which is salient in its display has extremely low values in both display and camera so that it is almost unlikely to be chosen.[1] This time, the consumer evaluates the phones x and y differently, with $u(x, \{x, y, z\}) = \frac{2}{3}$ and $u(y, \{x, y, z\}) = \frac{1}{3}$. As a consequence, $\rho(x, \{x, y, z\}) > \frac{1}{2} = \rho(x, \{x, y\})$, violating monotonicity.

We model the evaluation of an alternative as the weighted average of its values in multiple attributes. Other authors also use the multi-dimensional evaluation approach to model random choices. Gul et al. (2014) model a decision maker who first chooses an attribute according to the Luce rule and then chooses an alternative according to the Luce rule again based on the values in that selected attribute. If one think of the Luce rule as a random utility maximization model, the timing of their model can be described in two steps. The decision maker selects an attribute based on the attribute-specific information in the first step. After committing to an attribute, she then chooses an alternative based on the alternative-specific information. On the other hand, the decision maker in our model receives both information before making her choice. In particular, she does not need to commit to an attribute before making her choices. Kovach and Tserenjigmid (2019) also have a two-step Luce model, in which a decision maker commits to a subset of alternatives within the choice set before making the final choice. While both models explain violations of independence of irrelevant alternatives, they do not explain violations of monotonicity. (See Debreu (1960) for an example of the violation of the first property and

[1] This is just assumed for simplicity. Extreme values are not necessary to generate violations of monotonicity.

Natenzon (2019) for an example of the violation of the second property.)

Our model is similar to both Gul et al. (2014) and Kovach and Tserenjigmid (2019) in that it explains choice reversals (in the stochastic sense): It is possible that alternative x is more likely to be chosen than y from one choice set, but less likely to be chosen in another choice set. Choice reversals are often observed in the literature (eg., Lichtenstein and Slovic (1971)) and perhaps the attraction effect is the most popular example of choice reversal which is widely studied in the applied economics and marketing literature (eg., Huber et al. (1982) and Huber and Puto (1983)). It is commonly assumed that the analyst observes the attributes (for example, price and quality) of the alternatives in many of these studies. However, we do not assume that we can observe the attributes. Instead, we are able to uniquely pin down the attributes from the choice behaviors of the decision maker.

The remainder of the paper are organized as follow: We formally setup, introduce and interpret our model, which is called the Attribute-Dependent Random Utility Maximization (AD-RUM) representation, in Section 2. We discuss the uniqueness of AD-RUM representation in Section 3. After that, we discuss the behavioral implications of AD-RUM representation. It turns out that under a mild condition those behavioral implications are sufficient for the existence of the representation when the salient mapping is observable. Moreover, even when the salient mapping is unobservable, we are able to construct it from the choice data, and hence recover the representation. These main results of the paper are discussed in Section 4.

3.2 Model and Representation

Let X be the finite set of all possible alternatives, which could be interpreted as all possible products that can be produced in the market. A choice set, referred to as a menu in the model, $A \subseteq X$, is a subset of feasible products that a decision maker can choose from, which could be interpreted as the available products.

3.2.1 Choice Data

We begin with defining the choice probability function, which specifies the probability that a decision maker chooses each alternative from a menu.

Definition 3.2.1 (Choice Probability Function)**.** We say ρ is a choice probability function if ρ maps the alternative and menu pair (x, A) to $[0, 1]$ that satisfies

$$\sum_{x \in A} \rho(x, A) = 1 \text{ and } \rho(x, A) = 0 \text{ if and only if } x \notin A$$

for all $A \subseteq X$.

The choice probability function requires that any product x can be chosen by the decision maker with strictly positive probability whenever it is available in the market, and unavailable products are never chosen. A choice probability function induces a log-likelihood ratio function that specifies the logarithm of the choice probability ratio between two alternatives in a menu.

Definition 3.2.2 (Log-Likelihood Ratio Function)**.** We say l is a log-likelihood ratio function if there is a choice probability function ρ such that for all $x, y \in A$,

$$l_A(x, y) = \log\left(\frac{\rho(x, A)}{\rho(y, A)}\right).$$

A log-likelihood ratio function is well-defined since we require $\rho(x, A)$ to be strictly positive for any $x \in A$. It is only defined for menus with two or more alternatives in it. Note also that by definition, we have $l_A(x, y) = l_A(x, z) - l_A(y, z)$ whenever $x, y, z \in A$.

The definitions above suggest that for any choice probability function ρ, we have a unique and well-defined log-likelihood ratio function l^ρ. The converse is also true: For any log-likelihood ratio function l, one can define a choice probability function ρ^l by

$$\rho^l(x, A) = \frac{1}{\sum_{y \in A} \exp\left(l_A(y, x)\right)},$$

for all $x \in A$ with A containing at least 2 alternatives, and $\rho^l(x, \{x\}) = 1$ and $\rho^l(x, A) = 0$ if $x \notin A$. One can verify that if a log-likelihood ratio function l is induced by a choice probability function ρ, then ρ^l constructed from l is exactly ρ.

In this paper, we assume that we observe the choice probability function ρ that summarizes the decision maker's choice behaviors. While the primitive of the model is ρ, we may use the induced log-likelihood ratio function l^ρ as the choice data and impose restrictions based on it, due to the equivalence between them established above.

3.2.2 AD-RUM Representation

As we have motivated, we model a decision maker who is uncertain about how to aggregate her evaluations for an alternative along multiple attributes and learns it based on the information carried by the menu. This choice behavior is modeled by the following representation.

Definition 3.2.3 (AD-RUM)**.** An Attribute-Dependent RUM (AD-RUM) Representation, (v, I, s), consists of a finite index set $I = \{1, 2, \cdots, K\}$, a salient attribute mapping $s : X \to I$ and an attribute-dependent value function $v : X \times I \to \mathbb{R}$, such that alternative x contained in menu A is evaluated at

$$u(x, A) = \sum_i v(x, i) q(i, A),$$

with

$$q(i, A) = \frac{n_A^i}{n_A},$$

where $n_A = |A|$ and $n_A^i = |\{x \in A : s(x) = i\}|$. The probability of choosing alternative x from menu A is

$$\rho(x, A) = \frac{\exp(u(x, A))}{\sum_{y \in A} \exp(u(y, A))} \tag{3.2.1}$$

We say that the choice probability function ρ has an AD-RUM representation if there exists a tuple (v, I, s) such that Equation 3.2.1 is satisfied.

In this model, our decision maker is able to evaluate alternative $x \in X$ along any attribute $i \in I$, which is summarized by the attribute-dependent value function v. However, she is uncertain about the right trade-offs to make between different attributes, captured by the decision weights assigned to attributes. She learns them from the composition of the menu she faces and assigns the decision weight $q(i, A)$ to attribute i, which is the proportion of alternatives in menu A that are salient in alternative i. By aggregating the attribute-dependent values with respect to the decision weights, the decision maker forms the overall evaluation of the alternative, denoted by $u(x, A)$. Finally, choices are made based on these values according to Equation 3.2.1.

The evaluation of alternative x based on menu A can also be equivalently written as follows:

$$
\begin{aligned}
u(x, A) = & \sum_i v(x, i) \frac{n_A^i}{n_A} \\
= & \frac{1}{n_A} \sum_i \left(\sum_{y \in A} 1_{s(z')=i} \right) v(x, i) \\
= & \frac{1}{n_A} \sum_{z' \in A} v(x, s(z'))
\end{aligned}
$$

This alternative expression suggests that each alternative z in the menu is providing the decision maker with a piece of information about the value of x. While distinct alternatives that are salient in the same attribute are considered as different information sources, they convey the same information. The value of x is just the simple average of information.

Given this alternative expression, we have

$$
l_A(x, y) = \frac{1}{n_A} \sum_{z \in A} \left[v(x, s(z)) - v(y, s(z)) \right].
$$

This expression turns out to be useful in later sections. In the remainder of this section, we discuss the connection of this model with existing literature and its micro-foundation.

3.2.3 Micro-Foundation

Random Utility Maximization

Equation 3.2.1 above in the representation is closely related to Luce's choice model (Luce (1959)): If we think of the term $\exp(u(x, A))$ as the Luce value, the equation suggests that the probability of choosing alternative x from menu A is its Luce value divided by the total Luce value of the menu, which is exactly the Luce rule. Just as Luce's model can be interpreted as a random utility model, we interpret our model in the same way.

We interpret this choice procedure as if a decision maker is making choices under two independent sources of uncertainties. First, she is uncertain about the the overall values of alternatives in a menu and has to make inference based on the information carried by the menu. Second, she receives an additional piece of information about the values of alternatives, which is independent of her inference on the menu. As we have motivated in the introduction section, the first source of uncertainty represents the decision maker's learning from the market about the trade-offs among attributes and the second source of uncertainty could be interpreted as reading expert reviews or hearing friends' recommendations.

We assume that the decision maker is rational in the sense that she chooses the alternative that yields the highest utility in the menu, given all her information. That is, she chooses x from menu A if and only if

$$u(x, A) + \epsilon_x > u(y, A) + \epsilon_y,$$

for all $y \in A$ and $y \neq x$, where the idiosyncratic utility shock ϵ_x represents the second source of uncertainties. We make the parametric assumption that $\{\epsilon_x\}$ is distributed i.i.d. extreme value, which is commonly assumed in applied economic literature (Train (2009)). As a result, the probability that $x \in A$ is chosen by the decision maker has the closed form expression specified by Equation 3.2.1.

The choice rule specified by Equation 3.2.1 has many names in the literature. It is commonly called the discrete choice model in applied economics and referred to as the multinomial logistic model in statistics. In decision theory, one can think of it as the extension of the classic random utility maximization (RUM) model, except that the decision maker now has an additional inference step in her decisions. Precisely due to this inference step, her behaviors violate both the independence of irrelevant alternatives property and the monotonicity property, which are implied by the Luce model and many RUM models. We will focus on this in a later section of the paper.

Bayesian Inference on Decision Weights

In our model, the decision maker learns how to evaluate a product from the market composition. This idea has been modeled by Kamenica (2008). In his paper, firms design their product lines and take the preferences of informed consumers, who understand their own preferences, into consideration. An uninformed consumer who does not know her preference believes that she is no different than other consumers, and thus she learns about her preference from the market.

Borrowing this insight into our model, a rational decision maker should believe that an attribute is more important when she finds more products in the market that are salient in this attribute. This implies that her decision weight $q(i, A)$ assigned to attribute i under menu A is monotone in the number of alternatives in the menu that are salient in this attribute.

We push this monotonicity implication further and make the parametric assumption that $q(i, A) = n_A^i / n_A$, which is the proportion of the alternatives in the menu that are salient in an attribute. This assumption is made for tractability. While it could be justified by the Bayes rule applied to a specific pair of prior belief and data-generating process, which we discuss in detail in Appendix 3.6, we admit that this imposes a strong restriction on the decision maker's inference procedure.

3.3 Uniqueness

One might recall that in a discrete choice model, which our model resembles mathematically, it is impossible to identify the exact scales of utility but the differences in utility. Our model also has this feature: One can shift the utility of all alternatives along an attribute with the same constant without affecting the choice behaviors induced by the representation. The following proposition formalizes this argument.

Observation 1 (Uniform Shift of Utility)**.** If a choice probability function ρ is represented by the AD-RUM representation (v, I, s), it is also represented by the AD-RUM representation $(\hat{v}, I, s)$ with

$$\hat{v}(x, i) = v(x, i) + \alpha_i,$$

with $\alpha_i \in \mathbb{R}$ being any constant, for all $x \in X$ and $i \in I$.

It is straightforward to see this: For any $x, y \in A$, $l_A(x, y) = \sum_{i \in I} \frac{n_A^i}{n_A}[v(x, i) - v(y, i)]$ since ρ (and equivalently l) is represented by (v, I, s). By construction of $\hat{v}$, $l_A(x, y) = \sum_{i \in I} \frac{n_A^i}{n_A}[\hat{v}(x, i) - \hat{v}(y, i)]$ for all $x, y \in A$.

In addition to the uniform shift of utility, there is a second obvious way of constructing a new AD-RUM representation from one that represents the choice data.

Observation 2 (Duplicated Attributes)**.** If a choice probability function ρ is represented by the AD-RUM representation (v, I, s), it is also represented by the AD-RUM representation $(\hat{v}, I \cup \{K+1\}, \hat{s})$ such that there exists a non-empty set of alternatives $\emptyset \neq A \subsetneq \{x \in X : s(x) = i\}$ that are salient in i for some $i \in I$ such that for any

$x \in X$,

$$\hat{v}(x,j) = \begin{cases} v(x,i) & \text{if } j = K+1 \\ v(x,j) & \text{otherwise,} \end{cases}$$

and

$$\hat{s}(x) = \begin{cases} K+1 & \text{if } x \in A \\ s(x) & \text{otherwise.} \end{cases}$$

This construction just creates a new attribute $K+1$ that "duplicates" an existing one i. For a proper subset of alternatives that are salient in attribute i, denoted by A above, they are salient in the new attribute in the new representation. The new attribute-dependent value function is the same as the original one for any attribute $i \in I$. The values for the new attribute $K+1$ again "copies" those for attribute i.

In light of the above two observations and in order to rule out a possibly uninteresting case of our model, in which none of the alternatives share the same salient attribute, we make the following definition.

Definition 3.3.1 (Regular AD-RUM)**.** We say an AD-RUM representation (v, I, s) is regular, if $K < |X|$ and there exists no α such that

$$v(x,i) = v(x,j) - \alpha,$$

holds for all $x \in X$ for any distinct $i, j \in I$.

A regular AD-RUM representation satisfies two conditions. First, it has strictly less attributes than the number of all possible alternatives. This implies that at least there are two alternatives that share the same salient attribute. Second, there are no duplicated attributes in the representation. To facilitate the discussion of the uniqueness result, we make the following definitions. Define the set $X_i = \{x \in X : s(x) = i\}$ be the set of all alternatives that are salient in attribute i. Moreover, define $\mathcal{X}(I, s) = \{X_1, X_2, \cdots, X_K\}$ as the partition of X induced by (I, s). We state the uniqueness proposition below.

Proposition 1 (Uniqueness of AD-RUM)**.** Suppose (v, I, s) and $(\hat{v}, \hat{I}, \hat{s})$ are two regular AD-RUM representations of the same choice probability function ρ, then $\mathcal{X}(I, s) = \mathcal{X}(\hat{I}, \hat{s})$ and $v(x, i) - \hat{v}(x, i) = \alpha_i$ for some $\alpha_i \in \mathbb{R}$ for all $x \in X$ and $i \in I$, after possibly permuting $(\hat{I}, \hat{s})$.

The proposition says that whenever two AD-RUM representations represent the same data, they must induce the same partition of X with each subset consists of all alternatives that are salient in the same attribute and the only way the attribute-dependent value function can differ is by uniform shifts of utility. The complete proof is in Appendix 3.7.

We finish this section by noting that for any regular AD-RUM representation, it is possible to construct an irregular AD-RUM representation that represents the same choice data. In the irregular AD-RUM representation, none of the alternatives shares the same salient attribute. That is, we have $K = |X|$. More details are discussed in Appendix 3.7.

3.4 Characterization

3.4.1 Behavioral Implications

We first discuss the behavioral implications of the AD-RUM representation. Recall that the AD-RUM representation (v, I, s) implies that

$$\begin{aligned} l_A(x,y) =& u(x,A) - u(y,A) \\ =& \sum_{i \in I} \frac{n_A^i}{n_A} \left[v(x,i) - v(y,i) \right]. \end{aligned}$$

That is, the log-likelihood ratio between choosing x and y from menu A is the weighted average of the utility differences between x and y, with the weight for an attribute being the proportion of alternatives in the menu that are salient in that attribute.

In other words, in addition to being an option for the decision maker to choose from, every alternative in the menu plays the additional role of providing the decision maker with market information that facilitates her choice. As long as the decision maker learns the same decision weights from different information sources, her choice behaviors should remain the same. Our first set of properties exactly pertains to this observation.

If the choice probability function ρ has an AD-RUM representation (v, I, s), the choice data must exhibit the following properties.

Property 4 (Independence of Replacements (IR))**.** *For any $z, z' \in A \subseteq X$ and $s(x) = s(y)$ with $x, y \notin A$,*

$$l_{A \cup \{x\}}(z, z') = l_{A \cup \{y\}}(z, z').$$

In the expression above, the left-hand-size is the log-likelihood ratio between z

and z' in menu $A \cup \{x\}$ and the right-hand-side is the ratio in the modified menu that replaces alternative x with y. Since x and y are salient in the same attribute, the modified menu $A \cup \{y\}$ has the same composition as the original one in terms of its proportion of alternatives that are salient in each of the attributes – One just replaces an alternative with another that brings the same information as the removed one. The decision maker learns the same decision weights from both of the menus and hence makes the same choice.

Property 5 (Independence of Duplications (ID))**.** *For any $x, x', y, y' \in X$ with $s(x) = s(x')$ and $s(y) = s(y')$,*

$$l_{\{x,y\}}(x, y) = l_{\{x,y,x',y'\}}(x, y).$$

Since alternative x (alternative y) and alternative x' (alternative y') are salient in the same attribute, they provide the decision maker with the same information about the values of x and y at comparison. Therefore, the menu $\{x, y, x', y'\}$ conveys the same information as $\{x, y\}$ except that it simply "duplicates" the "evidence". As a result, the decision maker forms the same decision weights and hence makes the same choice.

On a different note, as we argued in the beginning of the section, every alternative in a menu plays an informational role in forming the decision weights. One can think of each alternative as a piece of evidence. In this sense, two different menus carry two different information contents, even when they lead to the same choice behavior. More specifically, consider $x, y \in A \cap B$. While the decision maker could have $\frac{n_A^i}{n_A} = \frac{n_B^i}{n_B}$ and hence $l_A(x, y) = l_B(x, y)$, the effects of adding a new alternative z into A and B can be different. Suppose A is a small menu and B is a large menu, adding z into A might have a strong effect on the choice between x and y, while adding z into B might hardly have any impact. This is because when the decision maker has already got many pieces of evidence from B, an additional piece of evidence has limited influence.

From this perspective, we interpret $n_A l_A(x, y)$ as the magnitude to which alternative x is better than alternative y based on all the information content carried by menu A. The AD-RUM representation implies the following property.

Property 6 (Bayesian Information Aggregation (BIA))**.** *For any $A, B \subseteq X$ with $x, y \in A \cap B$,*

$$n_{A\cup B} l_{A\cup B}(x, y) = n_A l_A(x, y) + n_B l_B(x, y) - n_{A\cap B} l_{A\cap B}(x, y).$$

This says that in terms of the magnitude to which x is better than y, the information content carried by the menu $A \cup B$ consists of the information content from both menus A and B after eliminating the double-counted content from $A \cap B$. The information aggregation is Bayesian in the sense that the magnitudes are additive.

While these three properties are necessary conditions of the AD-RUM representation, we will show that the converse is also true under a very mild condition.

3.4.2 Main Result with Known Attributes

We begin to state our main result by assuming that we observe the set of attributes that the decision maker may consider and the salient attribute of each alternative. That is, we observe (I, s) in addition to the choice data ρ. We are going to show that Independence of Replacements (IR), Independence of Duplicates (ID) and Bayesian Information Aggregation (BIA) imply the AD-RUM representation under the following mild condition.

Assumption 1 (Richness)**.** *There is rich variation in alternatives: For each $i =$*

$1, 2, \cdots, K$, the set $X_i = \{x \in X : s(x) = i\}$ has at least two elements. Moreover, $K \geq 2$.

We require that each attribute has at least two alternatives that are salient in it. Moreover, we also require that there are at least two attributes. We begin with the following lemma.

Lemma 3.4.1 (Weak IIA)**.** *Suppose (I, s) is observable and assume Richness.*

If the choice probability function satisfies IR, ID and BIA, then $s(x_1) = s(x_2) = s(x_3)$ implies

$$l_{\{x_1,x_2\}}(x_1, x_2) = l_{\{x_1,x_2,x_3\}}(x_1, x_2).$$

That is, ID, IR and BIA imply that when three alternatives have the same salient attribute, the comparison between two of them is independent of the third. This is named as Weak IIA because of its close relation to the Independence of Irrelevant Alternative (IIA) property of the classic Luce's choice model.

Luce model posits that $l_{\{x_1,x_2\}}(x_1, x_2) = l_{\{x_1,x_2,x_3\}}(x_1, x_2)$ always hold, which is regardless of the possible connection between the so-called irrelevant alternative x_3 and the two alternatives x_1 and x_3 at comparison. (For example, x_3 might be a complement or substitute good of x_1.) Weak IIA in our model is weaker: it requires that IIA holds when the alternatives share the same salient attribute. For a decision maker who behaves as if she is using our AD-RUM model, the menus $\{x_1, x_2\}$ and $\{x_1, x_2, x_3\}$ both yield the full weight on the salient attribute of these alternatives. We prove in Appendix 3.8 that IR, ID and BIA imply Weak IIA under Richness.

Our main result says that the AD-RUM representation is characterized by these three properties.

Proposition 2 (AD-RUM with Known Attributes)**.** Suppose (I, s) is observable and assume Richness.

The choice probability function ρ is represented by the AD-RUM representation (v, I, s) if and only if WIIA, IR, ID and BIA hold.

We sketch the proof of the proposition to provide the main idea. While we only require each attribute to have two distinct alternatives that are salient in it, we assume in the main text that there are three distinct alternatives for simplicity. One can find the complete proof without this strengthened assumption in Appendix 3.8.

We have a constructive proof: We begin by constructing the utility differences between any two alternatives for each attribute and then show that the choice data can be written as the weighted average of those utility differences with the right weights. We focus on the utility differences because it is impossible to identify the exact scales of the utility. (As a reminder, the AD-RUM representation is unique up to uniform utility shifts.)

Recall that we interpret $n_A l_A(x, y)$ as the magnitude to which x is better than y based on all pieces of evidence in menu A. This interpretation suggest that the difference between $3l_{x,y,z}(x, y)$ and $2l_{x,y}(x, y)$ represents the information carried by z about the comparison between x and y. This is exactly the choice data we use to construct utility differences.

For any two alternatives x, y and any attribute $i \in I$, we pick some $z \in X$ with $s(z) = i$ and define

$$d^i(x, y) := 3l_{\{x,y,z\}}(x, y) - 2l_{\{x,y\}}(x, y),$$

as the utility difference between x and y along attribute i. The utility differences are well-defined because such z always exists by our strengthened Richness assumption and the value is independent of the exact choice of z by IR.

Our goal is to prove that the choice data, summarized by $l_A(x, y)$ for all $x, y \in$

$A \subseteq X$, can be written as the aggregation of all the utility differences between x and y according to all pieces of "evidence" carried by the menu:

$$l_A(x, y) = \frac{1}{n_A} \sum_{z \in A} \left[v(x, s(z)) - v(y, s(z)) \right].$$

Recall that we have argued that this is the equivalent expression of our representation. We prove $l_A(x, y) = \frac{1}{n_A} \sum_{z \in A} \left[d^{s(z)}(x, y) \right]$ in the first stage and show the existence of a function $v : X \times I \to \mathbb{R}$ such that $v(x, i) - v(y, i) = d^i(x, y)$ in the second stage. We sketch the first stage below since it demonstrates how the behavioral conditions are linked with our representation. All omitted details are in the appendix.

We begin with binary menus. For a binary menu $\{x, y\}$ with $s(x) = s(y) = i$,

$$d^i(x, y) = 3l_{\{x,y,z\}}(x, y) - 2l_{\{x,y\}}(x, y) = l_{\{x,y\}}(x, y),$$

by Weak IIA. Therefore,

$$\begin{aligned} \frac{1}{2} \left[d^i(x, y) + d^j(x, y) \right] =& \frac{1}{2} \left[l_{\{x,y\}}(x, y) + l_{\{x,y\}}(x, y) \right] \\ =& l_{\{x,y\}}(x, y). \end{aligned}$$

For a binary menu $\{x, y\}$ with $s(x) = i \neq j = s(y)$, we have

$$\begin{aligned} \frac{1}{2} \left[d^i(x, y) + d^j(x, y) \right] =& \frac{1}{2} \left[\left(3l_{\{x,y,z_i\}}(x, y) - 2l_{\{x,y\}}(x, y) \right) + \left(3l_{\{x,y,z_j\}}(x, y) - 2l_{\{x,y\}}(x, y) \right) \right] \\ =& \frac{1}{2} \left[\left(4l_{\{x,y,z_i,z_j\}}(x, y) + 2l_{\{x,y\}}(x, y) \right) - 4l_{\{x,y\}}(x, y) \right] \\ =& l_{\{x,y\}}(x, y), \end{aligned}$$

In the derivation, we apply the definition of utility differences in the first step. By

BIA, we have

$$4l_{\{x,y,z_i,z_j\}}(x,y) = 3l_{\{x,y,z_i\}}(x,y) + 3l_{\{x,y,z_j\}}(x,y) - 2l_{\{x,y\}}(x,y).$$

This is used in the second step of the derivation. In the last set we invoke ID that

$$l_{\{x,y,z_i,z_j\}}(x,y) = l_{\{x,y\}}(x,y).$$

To extend this argument from binary menus to any menus, we prove by induction and invoke BIA in a similar manner as above. Then we have proven that

$$l_A(x,y) = \frac{1}{n_A} \sum_{z \in A} \left[d^{s(z)}(x,y) \right],$$

for all $x, y \in A \subset X$.

3.4.3 Main Result with Unknown Attributes

Even when we do not observe the subjective salient mapping of the decision maker, our last proposition suggests that if there exists a partition of X,

$$X = X_1 \cup X_2 \cup \cdots \cup X_K,$$

such that each X_i has at least two alternatives and $K \geq 2$ and such that IR and ID hold when we define $s(x) = i$ if $x \in X_i$, then BIA is equivalent to the AD-RUM representation. Our focus in this subsection is to construct the partition from the choice data.

Observe that if two alternatives provide the decision maker with the same information about how to evaluate alternatives, they must have the same impacts on the log-likelihood ratio for any decision problems. We attempt to rely on the converse of this statement to cluster the alternatives with the same informative content.

Definition 3.4.1. Define $x \sim_{\text{IR}} y$ if

$$l_{A\cup\{x\}}(z,z') = l_{A\cup\{y\}}(z,z'),$$

holds for any $z, z' \in A$ with $x, y \notin A$.

The definition requires that replacing x with y does not affect the log-likelihood ratio in any decision problems *that do not involve x and y themselves.* We denote this binary relation by $\sim_{\text{IR}}$ since it is the same type of behavior as the one identified in the property of Independence of Replacements. At this stage, it is attempting to argue that we can rely on $\sim_{\text{IR}}$ to partition the set of alternatives X into multiple subsets and interpret each subset as the alternatives that share the same salient attribute. Unfortunately, this is not correct for two reasons. We demonstrate it with an AD-RUM representation summarized by the table below.

Table 3.1: Binary relation $\sim_{\text{IR}}$ is not sufficient to identify attributes.

	att 1	att 2	att 3
x_1	$\boxed{v_1}$	v_1	v_1
x_2	$\boxed{v_2}$	a	a
y	b	$\boxed{v_3}$	d
z	c	c	$\boxed{v_4}$

In this table, the first column consists of the alternatives and the first row consists of the attributes. Every cell in the table indicates the value of an alternative according to an attribute and the box around the value indicates salience.

The first reason that $\sim_{\text{IR}}$ is not sufficient to cluster the alternatives is that it is not even a transitive binary relation. Notice that we would need an equivalence relation to partition a set, which requires transitivity. To see the intransitivity of $\sim_{\text{IR}}$, observe that the choice behavior induced by this representation would imply that $x_1 \sim_{\text{IR}} x_2$ and $x_2 \sim_{\text{IR}} y$. However, it also implies that $x_1 \not\sim_{\text{IR}} y$: We have

$$l_{\{x_2,z,x_1\}}(x_2, z) = \frac{1}{3}\left(2[v_2 - c] + [a - v_3]\right)$$

and

$$l_{\{x_2,z,y\}}(x_2, z) = \frac{1}{3}\left([v_2 - c] + 2[a - v_3]\right),$$

which are not equal in general, unless $v_2 - c = a - v_3$.

In this example, while we know from the table that $s(x_2) \neq s(y)$, the induced behavior suggests $x_2 \sim_{\text{IR}} y$ because they reveals the same values of alternatives *excluding themselves*. The problem arises exactly due to this exclusion: The behavior data we use to conclude $x_2 \sim_{\text{IR}} y$ does not involve x_2 and y themselves. In this example, a could be distinct from v_2 and b could be distinct from v_3. This example suggests that we would need more conditions than the ones used to define $\sim_{\text{IR}}$ to conclude whether two alternatives provide the same information.

Before we move to the stronger condition, we argue that there is a second reason why $\sim_{\text{IR}}$ is not sufficient to cluster the alternatives, in additional to intransitivity. Recall our argument for Independence of Duplicates: the decision maker forms the same decision weights if we duplicate the informational content carried by the menu. The above representation suggests $x_1 \sim_{\text{IR}} x_2$ and $y \sim_{\text{IR}} z$. As such, we would expect that

$$l_{\{x_1,y\}}(x_1, y) = \frac{1}{2}\left([v_1 - b] + [v_1 - v_3]\right)$$

and

$$l_{\{x_1,y,x_2,z\}}(x_1, y) = \frac{1}{4}\left(2[v_1 - b] + [v_1 - v_3] + [v_1 - d]\right)$$

are equal. This is not true in general, unless $v_3 - d$.

Again, this problem occurs exactly because $y \sim_{\mathrm{IR}} z$ does not tell us whether the value of y according to the salient attribute of y coincides with the value of y according to the salient attribute of z. On the other hand, this example also suggests we whether $l_{\{x_1,y\}}(x_1, y)$ is equal to $l_{\{x_1,y,x_2,z\}}(x_1, y)$ or not could be used to check whether $v_3 = d$. This turns out to be the missing piece that leads to the stronger definition for partitioning X.

Definition 3.4.2. For any four distinct alternatives satisfying $x \sim_{\mathrm{IR}} x'$ and $y \sim_{\mathrm{IR}} y'$, we define $(x, y) \sim_{\mathrm{ID}} (x', y')$ when the following holds:

$$l_{\{x_1,y_1\}}(x_1, y_1) = l_{\{x_1,y_1,x_2,y_2\}}(x_1, y_1),$$

for any distinct $x_1, x_2 \in \{x, x'\}$ and distinct $y_1, y_2 \in \{y, y'\}$.

We interpret $(x, y) \sim_{\mathrm{ID}} (x', y')$ as that the information carried by $\{x, y\}$ can be duplicated by that carried by $\{x', y'\}$. The definition requires that x and x' reveal the same value regarding the rest of the alternatives in X, and same for y and y'. In addition, it also requires that the log-likelihood ratio between any pair of alternatives, one from the x-group and one from the y-group, is independent of duplications, which is why we use the notation $\sim_{\mathrm{ID}}$.

As we motivated by the second example above, the behaviors required by $\sim_{\mathrm{ID}}$ can be used to detect if two alternatives reveal the same value regarding *themselves*. Combining these two requirements, $(x, x') \sim_{\mathrm{ID}} (y, y')$ suggests that x and x' reveals the same information about *all* alternatives in X, including themselves. With this interpretation, it leads to the following definition that combines both notions of $\sim_{\mathrm{IR}}$ and $\sim_{ID}$.

Definition 3.4.3. Define $x \sim x'$ if the following two conditions hold:

1. $x \sim_{\text{IR}} x'$, and

2. $(x, y) \sim_{\text{ID}} (x', y')$ for all $(y, z) \sim_{\text{ID}} (y', z')$.

Note first that the first condition, $x \sim_{\text{IR}} x'$, is actually implied by the second condition. We keep it merely for completeness. The second condition is more complicated to understand. To say that x and x' conveys the same information about the values of all alternatives, we need more than a pair y and y' such that the information carried by $\{x, y\}$ can be duplicated by $\{x', y'\}$. Instead, it requires $(x, y) \sim_{\text{ID}} (x', y')$ to hold for all pairs such as y and y'. The following example demonstrates why the stronger requirement is necessary.

Table 3.2: A stronger notion of $\sim$ is necessary.

	att 1	att 2	att 3	att 4	att 5	att 6	att 7	att 8
x_1	$\boxed{1}$	1	0	0	0	0	0	0
x_2	1	$\boxed{1}$	0	0	0	0	0	0
y_1	0	0	$\boxed{1}$	1	0	0	0	0
y_2	0	0	1	$\boxed{1}$	0	0	0	0
z_1	0	0	0	0	$\boxed{1}$	2	0	0
z_2	0	0	0	0	2	$\boxed{1}$	0	0
w_1	0	0	0	0	0	0	$\boxed{1}$	2
w_2	0	0	0	0	0	0	2	$\boxed{1}$

Consider the choice behavior induced by the AD-RUM representation summarized by the above table. Suppose we weaken the second condition in the last definition as $(x, y) \sim_{\text{ID}} (x', y')$, we would conclude that the two alternatives x_1 and x_2 have the same salient attribute. This is because we have $(x_1, y_1) \sim_{\text{ID}} (x_2, y_2)$. The same holds for the y's, z's and w's. However, while

$$l_{\{x_1, z_1\}}(x_1, z_1) = \frac{1}{2}\left[(1 - 0) + (0 - 1)\right] = 0,$$

$$l_{\{x_1,z_1,x_2,z_2\}}(x_1, z_1) = \frac{1}{4}\left[(1-0)+(0-1)+(1-0)+(0-2)\right] \neq 0.$$

The partition according to the weakened definition does not satisfies ID. In fact, we can tell from the table directly that z_1 and z_2 do not reveal the same information about themselves.

Before we show the identification result, we make the following richness condition.

Assumption 2 (Subjective Richness)**.** *There is enough variation in the alternatives such that for all $x \in X$ there exists $x' \in X$ such that $x \sim x'$ and there exists $x \not\sim y$.*

Proposition 3 (Identification of Attributes)**.** Assume Subjective Richness. If the choice probability function ρ satisfies BIA, the binary relation $\sim$ is an equivalence relation on X. Therefore, X could be partitioned into K disjoint subsets

$$X = X_1 \cup X_2 \cup \cdots \cup X_K.$$

Moreover, properties IR and ID hold if we define $s : X \to \{1, \cdots, K\}$ by $s(x) = i$ if $x \in X_i$.

We interpret $x \sim x'$ as that x and x' provide the decision maker with the same information about the values of all alternatives, including themselves. The proposition says that this interpretation is valid in the sense that IR and ID hold. The proposition also suggests that even when we do not observe the subjective salient mapping, as long as the choice data satisfies BIA, we can still construct it from the choice data.

To conclude, observe that when there is sufficient variation among the alternatives, BIA alone is equivalent to the AD-RUM representation.

Corollary 3.4.1.1 (AD-RUM with Unknown Attributes)**.** *Assume Subjective Richness. The choice probability function ρ is represented by the AD-RUM representation*

(v, I, s) *if and only if BIA hold.*

3.5 Conclusion

To conclude, we have modeled a decision maker who needs to learn about how to make trade-offs among attributes when evaluating alternatives and making choices. Based on the choice behaviors of the decision maker, we can uniquely classify the alternatives using choice data, and interpret them as the subjective salient mapping that is only observable to the decision maker. Under mild conditions, the choice data can be explained as if it is generated by the decision maker according to the AD-RUM representation. Our representation, despite its similarity to Luce's choice model and random utility maximization models, explains violations of independence of irrelevant alternatives and monotonicity property by assuming that the decision maker needs to learn from the market.

3.6 Appendix: Micro-foundation for the Inference Process

We assume in the model that the decision maker assigns decision weight $q(i, A) = n_A^i / n_A$ to attribute i, which is the proportion of alternatives in the menu A that are salient in attribute i. While this assumption is made mainly for tractability, we discuss the micro-foundation for this inference process.

Suppose there are K attributes, where K is finite. The relative weights assigned to those K attributes can be summarized as a K-dimensional vector, which is an element of the $K-1$ dimensional standard simplex

$$\Delta_{K-1} = \{\mathbf{q} \in \mathbb{R}^K : \sum_j q_j = 1 \text{ and } q_j \geq 0,\ \forall j = 1, 2, \cdots K\}.$$

Notice that requiring that $\sum_j q_j = 1$ is merely a normalization. There is a "correct" way of making trade-offs among attributes, denoted by $\mathbf{q}^* \in \Delta_{K-1}$. One can think of this as the preferences of informed consumers, as is motivated in the main text.

Let $\mathbf{q}^*$ be the "true weight" that our decision maker aims to learn from the market, which is a fixed vector that takes value from the whole $K-1$ dimensional standard simplex. The decision maker has a prior belief about the true weight, modeled as a random vector $\tilde{\mathbf{q}}$.[2] Specifically, we assume that

$$\tilde{\mathbf{q}} \sim \text{Dir}(K, \mathbf{a}).$$

That is, the prior belief is a Dirichlet distribution of order K with parameter $\mathbf{a} \in \mathbb{R}^K_+$. the vector $\mathbf{a}$ is called the hyper-parameter in Bayesian statistics since it is exogenously specified before the learning takes place.

Now that we have specified the decision maker's prior belief about the true weights, we start to specify the decision maker's belief about how the menu is composed of the alternatives with various salient attributes. Since she believes that her preference is no different from other consumers in the market, and the menu's composition takes the average preference into account, she may have the following belief: When the weight assigned to attribute i is q_i, she believes that when she randomly draws an alternative x from a menu, the alternative should be salient in attribute i

[2] To interpret this, let Q be the CDF of $\tilde{\mathbf{q}}$. For any Borel set $E \subseteq \Delta_{K-1}$, the decision maker believes that the true weight is in this set with probability $Q(E)$.

with probability q_i. Mathematically, we have actually assumed that

$$\mathbf{n_A} \sim \mathrm{Mul}(K, \tilde{\mathbf{q}}),$$

where $\mathbf{n_A} = (n_A^1, \cdots, n_A^K)$ represents the composition of menu A in terms of the salient attributes of its elements. This assumption says that $\mathbf{n_A}$ is drawn from a multinomial distribution of order K with parameter $\tilde{\mathbf{q}}$. Note that the parameter itself is random.

Given this information structure, one can show that when a menu A yields $\mathbf{n_A}$, the posterior belief over the states is

$$\tilde{\mathbf{q}}^* \sim \mathrm{Dir}(K, \mathbf{a} + \mathbf{n_A}),$$

which is another Dirichlet distribution of order K with parameter $\mathbf{a}+\mathbf{n_A}$. Therefore, the expected value of alternative $x \in X$ in menu A is given by

$$\begin{aligned} u(x, A) &= \int_{\tilde{q} \in \Delta_{K-1}} \left(\sum_j v(x, j) \tilde{q}(j, A) \right) dQ(\tilde{q}) \\ &= \sum_j v(x, j) \frac{a^j + n_A^j}{\sum_k (a^k + n_A^k)}, \end{aligned}$$

where we switched the order of the summations and used the fact that $\mathbb{E}(\tilde{\mathbf{q}}_{\mathbf{j}}^*) = \frac{a^j + n_A^j}{\sum_{k=1}^K (a^k + n_A^k)}$ in the second line. We have this nice closed-form expression for the posterior exactly because the Dirichlet distribution is conjugate to the multinomial distribution[3]. Yet, assuming other distributions should not affect the behavioral implications of the representation at least qualitatively.

We finally assume that the decision maker as an extremely weak prior. Mathe-

[3] One can refer to Hoff (2009) for the technical details.

matically, this is assuming that a^j is arbitrarily close to 0. This could be interpreted as that the decision maker has never bought or heard of the product before, and hence has completely know idea about the importance of the attributes. With this assumption, it follows that

$$u(x, A) = \sum_j v(x, j) \frac{n_A^j}{n_A}.$$

3.7 Appendix: Proof of Uniqueness

3.7.1 Uniqueness of Regular AD-RUM

We prove the uniqueness result of the AD-RUM representation here. Suppose (v, I, s) and $(\hat{v}, \hat{I}, \hat{s})$ both represent the same data ρ.

If $X(I, s)$ does not coincide with $X(\hat{I}, \hat{s})$, there must exist $x, y \in X$ with $s(x) = x(y)$ according to the first representation but $\hat{x} \neq \hat{y}$ in the second representation. Since the representation is regular, it follows that there exist $z, z' \in X$ with

$$\hat{v}(z, \hat{s}(x)) - \hat{v}(z, \hat{s}(y)) \neq \hat{v}(z', \hat{s}(x)) - \hat{v}(z', \hat{s}(y)),$$

otherwise attribute $\hat{s}(x)$ is a duplicate of attribute $\hat{s}(y)$, contradicting the fact that the representation is regular.

The first representation then induces that

$$\begin{aligned}
&l_{\{z,z',x\}}(z,z') \\
=&\frac{1}{3}\left([v(z,s(z))-v(z',s(z))]+[v(z,s(z'))-v(z',s(z'))]+[v(z,s(x))-v(z',s(x))]\right) \\
=&\frac{1}{3}\left([v(z,s(z))-v(z',s(z))]+[v(z,s(z'))-v(z',s(z'))]+[v(z,s(y))-v(z',s(y))]\right) \\
=&l_{\{z,z',y\}}(z,z'),
\end{aligned}$$

because $s(x)=s(y)$. However, the second representation induces that

$$\begin{aligned}
&\hat{l}_{\{z,z',x\}}(z,z') \\
=&\frac{1}{3}\left([\hat{v}(z,s(z))-\hat{v}(z',s(z))]+[\hat{v}(z,s(z'))-\hat{v}(z',s(z'))]+[\hat{v}(z,s(x))-\hat{v}(z',s(x))]\right) \\
\neq&\frac{1}{3}\left([\hat{v}(z,s(z))-\hat{v}(z',s(z))]+[\hat{v}(z,s(z'))-\hat{v}(z',s(z'))]+[\hat{v}(z,s(y))-\hat{v}(z',s(y))]\right) \\
=&l_{\{z,z',y\}}(z,z').
\end{aligned}$$

This shows that the two representations induce two different behaviors.

Now without loss we can assume that we have (v,I,s) and $(\hat{v},I,s)$ representing the same data. Suppose there exist i and z,z' such that $v(z,i)-\hat{v}(z,i)\neq v(z',i)-\hat{v}(z',i)$. One can construct the contradiction by picking $s(x)=i$ and verify $l_{\{z,z',x\}}(z,z')\neq\hat{l}_{\{z,z',x\}}(z,z')$. This completes the proof.

3.7.2 Irregular AD-RUM

We argue in the main text that a regular AD-RUM representation may have an equivalent irregular AD-RUM representation. The following example demonstrates this argument.

The two tables summarize two AD-RUM representations. The left one is regular

Table 3.3: Regular and Irregular AD-RUM representations

	att 1	att 2
x_1	$\boxed{v_1}$	a
x_2	$\boxed{v_2}$	b
y_1	c	$\boxed{v_3}$
y_2	d	$\boxed{v_4}$

	att 1	att 1'	att 2	att 2'
x_1	$\boxed{v_1+\theta}$	v_1	a	a
x_2	v_2	$\boxed{v_2+\theta}$	b	b
y_1	c	c	$\boxed{v_3+\theta}$	v_3
y_2	d	c	v_4	$\boxed{v_4+\theta}$

while the right one is irregular with each alternative salient in its own attribute. A box around an attribute-dependent value indicates salience. One can verify both representations induce the same choice behaviors.

The idea is that one can construct a maximal set of attributes so that each alternative is salient in its own attribute. The constructed attribute-dependent value function inherits the original values according to the original salient attribute, except that the value of each alternative along its own attribute can be shifted by a common constant θ. When $\theta = 0$, the new representation just duplicates the attributes.

The case where $\theta \neq 0$ is interesting. Note that choice behaviors are revealed by the log-likelihood ratios, which always involve two alternatives at comparison. Adding the same θ's to the value of alternative at their salient features only does not affect the induced behaviors because they are exactly cancelled out.

The example is summarized by the following observation.

Observation 3 (Maximal Attributes)**.** If a choice probability function ρ is represented by the AD-RUM representation (v, I, s), it is also represented by the AD-RUM representation $(\hat{v}_\theta, \hat{I}, \hat{s})$ with

$$\hat{I} = X, \qquad \hat{s}(x) = x,$$

$$\hat{v}(x,z) = \begin{cases} v(x,s(z)) + \theta & \text{if } z = x \\ v(x,s(z)) & \text{otherwise,} \end{cases}$$

for any constant $\theta \in \mathbb{R}$.

We show that both representations induce the same behaviors. The constructed irregular representation implies that

$$\begin{aligned}
n_A \hat{l}_A(x,y) =& \sum_{z \in A} [\hat{v}(x,\hat{s}(z)) - \hat{v}(y,\hat{s}(z))] \\
=& \sum_{z \in A: z \neq x, z \neq y} [\hat{v}(x,\hat{s}(z)) - \hat{v}(y,\hat{s}(z))] + [\hat{v}(x,\hat{s}(x)) - \hat{v}(y,\hat{s}(x))] + [\hat{v}(x,\hat{s}(y)) - \hat{v}(y,\hat{s}(y))] \\
=& \sum_{z \in A: z \neq x, z \neq y} [v(x,s(z)) - v(y,s(z))] \\
& \qquad + [v(x,s(x)) + \theta - v(y,s(x))] + [v(x,s(y)) - v(y,s(y)) - \theta] \\
=& \sum_{z \in A} [v(x,s(z)) - v(y,s(z))] \\
=& n_A l_A(x,y),
\end{aligned}$$

which establishes the result.

3.8 Appendix: Proof of Main Result

3.8.1 With Known Attribute

We only show that the behavioral properties collectively imply the existence of a AR-DUM representation that represent the choice probability function. The converse statement is much easier and hence omitted.

Recall that the exact scale of the utility in the representation is not identifiable, we focus on the differences in utility. We denote the difference in utility between x and y in attribute i as $d^i(x, y)$ and we construct these differences from the choice data. For any two x and y, we define

$$d^i(x, y) := \begin{cases} 3l_{\{x,y,z\}}(x, y) - 2l_{\{x,y\}}(x, y) \text{ for some } s(z) = i & \text{if such } z \text{ exists,} \\ l_{\{x,y\}}(x, y) & \text{otherwise.} \end{cases}$$

Lemma 3.8.1. *The difference in utility between x and y in attribute i, $d^i(x, y)$, is well-defined.*

Proof. We have assumed that there are at least two alternatives that are salient in each of the attributes. Therefore, we are always able to find the alternative z with $s(z) = i$ when $s(x) \neq i$ or $s(y) \neq i$. When there are multiple alternatives satisfying $s(z) = s(z') = i$, we have $l_{\{x,y,z\}}(x, y) = l_{\{x,y,z'\}}(x, y)$ by IR. Hence the utility difference is well-defined in this case.

Suppose $s(x) = s(y) = i$. There may or may not be a third alternative z that is also salient in attribute i. When such $s(z) = i$ exists, we still define the difference as $3l_{\{x,y,z\}}(x, y) - 2l_{\{x,y\}}(x, y)$, which is well-defined according to the same argument. The only case in which we define $d^i(x, y)$ as $l_{\{x,y\}}(x, y)$ is when $s(x) = s(y) = i$ and they are the only two alternatives that are salient in i. □

Lemma 3.8.2. *For any distinct $x, y, z \in X$ and $i \in I$ such that there exist another alternative w with $s(w) = i$,*

$$d^i(x, y) = d^i(x, z) - d^i(y, z).$$

Proof. Let $s(w) = i$. By BIA,

$$4l_{\{x,y,z,w\}}(x,y) = 3l_{\{x,y,z\}}(x,y) + 3l_{\{x,y,w\}}(x,y) - 2l_{\{x,y\}}(x,y).$$

This implies that

$$d^i(x,z) = 3l_{\{x,z,w\}}(x,z) - 2l_{\{x,z\}}(x,z) = 4l_{\{x,y,z,w\}}(x,z) - 3l_{\{x,y,z\}}(x,z),$$

$$d^i(y,z) = 3l_{\{y,z,w\}}(y,z) - 2l_{\{y,z\}}(y,z) = 4l_{\{x,y,z,w\}}(y,z) - 3l_{\{x,y,z\}}(y,z),$$

$$d^i(x,y) = 3l_{\{x,y,w\}}(x,y) - 2l_{\{x,y\}}(x,y) = 4l_{\{x,y,z,w\}}(x,y) - 3l_{\{x,y,z\}}(x,y).$$

Therefore we have

$$\begin{aligned} d^i(x,z) - d^i(y,z) =& 4[l_{\{x,y,z,w\}}(x,z) - l_{\{x,y,z,w\}}(y,z)] - 3[l_{\{x,y,z\}}(x,z) - l_{\{x,y,z\}}(y,z)] \\ =& 4l_{\{x,y,z,w\}}(x,y) - 3l_{\{x,y,z\}}(x,y) \\ =& d^i(x,y), \end{aligned}$$

where in the first and the last line we apply the definition and in the middle line we use the fact that $l_A(x,y) = l_A(x,z) - l_A(y,z)$. □

Now we are ready to prove Lemma 3.4.1

Proof of WIIA. Let $s(x_1) = s(x_2) = s(x_3) = 1$ and $s(y_1) = s(y_2) = 2$.

Observe that

$$\begin{aligned} 4l_{\{x_1,x_2,y_1,y_2\}}(x_1,x_2) =& 3l_{\{x_1,x_2,y_1\}}(x_1,x_2) + 3l_{\{x_1,x_2,y_2\}}(x_1,x_2) - 2l_{\{x_1,x_2\}}(x_1,x_2) \\ =& 6l_{\{x_1,x_2,y_1\}}(x_1,x_2) - 2l_{\{x_1,x_2\}}(x_1,x_2), \end{aligned}$$

where we apply BIA in the first line and IR between y_1 and y_2 in the second line. This implies

$$\begin{aligned}
l_{\{x_1,x_2\}}(x_1,x_2) =& 3l_{\{x_1,x_2,y_1\}}(x_1,x_2) - 2l_{\{x_1,x_2,y_1,y_2\}}(x_1,x_2) \\
=& \left[3l_{\{x_1,x_2,y_1\}}(x_1,y_1) - 3l_{\{x_1,x_2,y_1\}}(x_2,y_1)\right] \\
&\quad - \left[2l_{\{x_1,x_2,y_1,y_2\}}(x_1,y_1) - 2l_{\{x_1,x_2,y_1,y_2\}}(x_2,y_1)\right] \\
=& \left[3l_{\{x_1,x_2,y_1\}}(x_1,y_1) - 2l_{\{x_1,y_1\}}(x_1,y_1)\right] \\
&\quad - \left[3l_{\{x_1,x_2,y_1\}}(x_2,y_1) - 2l_{\{x_2,y_1\}}(x_2,y_1)\right] \\
=& d^1(x_1,y_1) - d^1(x_2,y_1) \\
=& d^1(x_1,x_2) \\
=& 3l_{\{x_1,x_2,x_3\}}(x_1,x_2) - 2l_{\{x_1,x_2\}}(x_1,x_2),
\end{aligned}$$

where we apply the definition of log-likelihood ratio in the second line, ID in the third line, and the lemma 3.8.2 in the fifth line. This completes the proof. □

We prove Proposition 2 below.

Proof of Proposition 2. In what follows, we show that for all menus $A \subseteq X$, the "predicted" log-likelihood

$$\hat{l}_A(x,y) = \frac{1}{n_A}\sum_{z\in A}[v(x,s(z)) - v(y,s(z))] = \frac{1}{n_A}\sum_{z\in A} d^{s(z)}(x,y),$$

computed according to the "hypothesized" representation and the differences in utility constructed above is equal to the actual log-likelihood. We prove this in two steps:

- Establish this statement for binary menus in the first step.

- Extend it to general menus in the second step.

Then in the third and final step, we need to construct the attribute-dependent value function $v : X \times I \to \mathbb{R}$ from the utility differences, which completes the proof.

Step 1: Binary Menus

For any $x, y \in X$, we want to show that $\hat{l}_{\{x,y\}}(x, y) = l_{\{x,y\}}(x, y)$.

Suppose $s(x) = s(y) = i$ for some $i \in I$. If x and y are the only two alternatives that are salient in i,

$$\hat{l}_{\{x,y\}}(x, y) = \frac{1}{2}[d^{s(x)}(x, y) + d^{s(y)}(x, y)] = \frac{1}{2}[l_{\{x,y\}}(x, y) + l_{\{x,y\}}(x, y)] = l_{\{x,y\}}(x, y).$$

If there exists a third alternative z with $s(z) = i$,

$$\begin{aligned}
\hat{l}_{\{x,y\}}(x, y) =& \frac{1}{2}[d^{s(x)}(x, y) + d^{s(y)}(x, y)] \\
=& \frac{1}{2}[3l_{\{x,y,z\}}(x, y) - 2l_{\{x,y\}}(x, y) + 3l_{\{x,y,z\}}(x, y) - 2l_{\{x,y\}}(x, y)] \\
=& 3l_{\{x,y,z\}}(x, y) - 2l_{\{x,y\}}(x, y) \\
=& l_{\{x,y\}}(x, y),
\end{aligned}$$

where WIIA is applied in the last line.

Suppose $s(x) = i$ and $s(y) = j$ for some distinct $i, j \in I$.

$$\begin{aligned}
\hat{l}_{\{x,y\}}(x, y) =& \frac{1}{2}\left[d^i(x, y) + d^j(x, y)\right] \\
=& \frac{1}{2}\left[\left(3l_{\{x,y,z_i\}}(x, y) - 2l_{\{x,y\}}(x, y)\right) + \left(3l_{\{x,y,z_j\}}(x, y) - 2l_{\{x,y\}}(x, y)\right)\right] \\
=& \frac{1}{2}\left[\left(4l_{\{x,y,z_i,z_j\}}(x, y) + 2l_{\{x,y\}}(x, y)\right) - 4l_{\{x,y\}}(x, y)\right] \\
=& l_{\{x,y\}}(x, y).
\end{aligned}$$

The first and second lines follow by definition. In the third line, we invoke BIA that

$$4l_{\{x,y,z_i,z_j\}}(x,y) = 3l_{\{x,y,z_i\}}(x,y) + 3l_{\{x,y,z_j\}}(x,y) - 2l_{\{x,y\}}(x,y).$$

In the last line, we invoke ID that

$$l_{\{x,y,z_i,z_j\}}(x,y) = l_{\{x,y\}}(x,y).$$

This establishes the representation for binary menus.

Step 2: General Menus

Since we have already shown that the log-likelihood ratio can be written as the weighted average of the utility differences for any binary menus. In order to establish this result for general menus, we prove by induction and show the inductive step below.

Let A be a menu of size $n_A > 2$ that contains x, y, and suppose that we have already shown that $\hat{l}_B(x', y') = l_B(x', y')$ for any menu $x', y' \in B$ with $n_B = n_A - 1$. Pick some $w \in A$ distinct from x and y.

$$\begin{aligned}
n_A \hat{l}_A(x,y) =& \sum_{z \in A} d^{s(z)}(x,y) \\
=& \sum_{z \in A: z \neq w} d^{s(z)}(x,y) + d^{s(w)}(x,y) \\
=& n_{A\setminus\{w\}} \hat{l}_{A\setminus\{w\}}(x,y) + 3l_{\{x,y,w\}}(x,y) - 2l_{\{x,y\}}(x,y) \\
=& n_{A\setminus\{w\}} l_{A\setminus\{w\}}(x,y) + 3l_{\{x,y,w\}}(x,y) - 2l_{\{x,y\}}(x,y) \\
=& n_A l_A(x,y).
\end{aligned}$$

We used the definition in the first and the third line and the induction assumption in the fourth line. The last line follows from BIA.

Notice that we implicitly assumed that $d^{s(w)}(x,y) = 3l_{\{x,y,w\}}(x,y) - 2l_{\{x,y\}}(x,y)$ instead of $d^{s(w)}(x,y) = l_{\{x,y\}}(x,y)$. This is valid: The latter case occurs only when $s(x) = s(y) = i$, $s(w) = i$, and there is no alternative other than x and y that is salient in i, which is impossible.

Step 3: Construct $v : X \times I \to \mathbb{R}$

We show that there exists an attribute-dependent value function $v : X \times I \to \mathbb{R}$ such that $d^i(x,y) = v(x,i) - v(y,i)$ for all $x, y \in X$ and i. There are two cases to consider.

We begin with the simpler case where there exists one attribute with at least three alternatives that are salient in it. Let this attribute be 1 and fix an alternative z^* with $s(z^*) = 1$. Let $v(z^*, i) = 0$ for any $i \in I$ and $v(x,i) = v(z^*,i) + d^i(x,z^*) = d^i(x,z^*)$ for any x and i. It remains to show that the constructed function v satisfies $v(x,i) - v(y,i) = d^i(x,y)$. This is true because $v(x,i) - v(y,i) = d^i(x,z^*) - d^i(y,z^*) = d^i(x,y)$ by Lemma 3.8.2.[4]

We are left with the more complicated case where each attribute has exactly two alternatives that are salient in it. Let x_i and x_i' be the two and only two alternatives that are salient in attribute i. We define $v(x_i', i) = 0$ for each i and $v(y,i) = v(x_i',i) + d^i(y,x_i') = d^i(y,x_i')$ for any y and i. It remains to show that $v(y,i) - v(z,i) = d^i(y,z)$.

[4] To be perfectly accurate, the only case in which Lemma 3.8.2 is not applicable is when x, y, z^* are the only three alternatives that are salient in 1. However, in this case,

$$\begin{aligned} d^i(x,z^*) - d^i(y,z^*) &= \left[3l_{\{x,y,z^*\}}(x,z^*) - 2l_{\{x,z^*,\}}(x,z^*)\right] - \left[3l_{\{x,y,z^*\}}(y,z^*) - 2l_{\{x,z^*,\}}(y,z^*)\right] \\ &= l_{\{x,y,z^*\}}(x,z^*) - l_{\{x,y,z^*\}}(y,z^*) \\ &= l_{\{x,y,z^*\}}(x,y) \\ &= 3l_{\{x,y,z^*\}}(x,y) - 2l_{\{x,z^*,\}}(x,y) \\ &= d^i(x,y). \end{aligned}$$

When none of y and z are salient in i, this is again true by Lemma 3.8.2. Suppose $y = x_i$ and $x = x_j$.

$$\begin{aligned}
v(x_i, i) - v(x_j, i) =& d^i(x_i, x_i') - d^i(x_j, x_i') \\
=& l_{\{x_i, x_i'\}}(x_i, x_i') - 3l_{\{x_i, x_i', x_j\}}(x_i, x_j) + 2l_{\{x_i, x_j\}}(x_i, x_j) \\
=& 3l_{\{x_i, x_i', x_j\}}(x_i, x_i') - 2l_{\{x_i, x_i', x_j, x_j'\}}(x_i, x_i') \\
& \quad - 3l_{\{x_i, x_i', x_j\}}(x_i, x_j) + 2l_{\{x_i, x_j\}}(x_i, x_j) \\
=& 3l_{\{x_i, x_i', x_j\}}(x_i, x_i') - 2l_{\{x_i, x_i', x_j, x_j'\}}(x_i, x_i') \\
& \quad - 3l_{\{x_i, x_i', x_j\}}(x_i, x_j) + 2l_{\{x_i, x_i', x_j, x_j'\}}(x_i, x_j) \\
=& 3l_{\{x_i, x_i', x_j\}}(x_i, x_j) - 2l_{\{x_i, x_i', x_j, x_j'\}}(x_i, x_j) \\
=& 3l_{\{x_i, x_i', x_j\}}(x_i, x_j) - 2l_{\{x_i, x_j\}}(x_i, x_j) \\
=& d^i(x_i, x_j).
\end{aligned}$$

In the derivation, we apply the observation in the proof for WIIA in the third line, ID in the fourth line, definition of log-likelihood ratio in the fifth line, and ID again in the sixth line.

This establishes the existence of attribute-dependent value function and therefore completes the proof of the proposition. □

3.8.2 With Unknown Attributes

We show that $\sim$ is an equivalence equation and the subjective versions of IR and ID hold given our subjective partition.

To see $\sim$ is transitive, we need to show $x \sim x'$ and $x' \sim x''$ imply $x \sim x''$. That is, $x \sim_{\text{IR}} x''$ and for all $(y, y') \sim_{\text{ID}} (z, z')$ we have $(x, x'') \sim_{\text{ID}} (y, y')$.

For any pair (y, y') satisfying $(y, y') \sim_{\mathrm{ID}} (z, z')$, we have

$$l_{\{x',y\}}(x', y) = l_{\{x',y,x,y'\}}(x', y),$$

since $x \sim x'$, and

$$l_{\{x',y\}}(x', y) = l_{\{x',y,x'',y'\}}(x', y),$$

since $x' \sim x''$. That implies

$$l_{\{x',y,x,y'\}}(x', y) = l_{\{x',y,x'',y'\}}(x', y).$$

By BIA, we know the following two expression:

$$5l_{\{x',y,x,x'',y'\}}(x', y) = 4l_{\{x',y,x,y'\}}(x', y) + 3l_{\{x',y,x''\}}(x', y) - 2l_{\{x',y\}}(x', y),$$

and

$$5l_{\{x',y,x,x'',y'\}}(x', y) = 4l_{\{x',y,x'',y'\}}(x', y) + 3l_{\{x',y,x\}}(x', y) - 2l_{\{x',y\}}(x', y).$$

Combining the last three equations to conclude $l_{\{x',y,x''\}}(x', y) = l_{\{x',y,x\}}(x', y)$.

Note that for any $w, w' \in X \backslash \{x, x', x''\}$, we have

$$l_{\{w,w',x\}}(w, w') = l_{\{w,w',x'\}}(w, w') = l_{\{w,w',x''\}}(w, w'),$$

because $x \sim_{\mathrm{IR}} x'$ and $x' \sim_{\mathrm{IR}} x''$. We have also shown that $l_{\{x',y,x\}}(x', y) = l_{\{x',y,x''\}}(x', y)$ for any y satisfying $(y, y') \sim_{\mathrm{ID}} (z, z')$. By Richness, this actually holds for any $y \in X \backslash \{x, x', x''\}$. Therefore, we have shown that $l_{\{w,w',x\}}(w, w') = l_{\{w,w',x''\}}(w, w')$

for any $w, w' \in X \backslash \{x, x''\}$. By BIA, this can be extended to

$$l_{A\cup\{x\}}(w, w') = l_{A\cup\{x''\}}(w, w'),$$

for all $w, w' \in A \subseteq X \backslash \{x, x''\}$. Thus we have established $x \sim_{\text{IR}} x''$.

For any pair (y, y') satisfying $(y, y') \sim_{\text{ID}} (z, z')$, we now show $l_{\{x,y\}}(x, y) = l_{\{x,y,x'',y'\}}(x, y)$. Since $(x, x') \sim_{\text{ID}} (y, y')$, we have $l_{\{x,y\}}(x, y) = l_{\{x,y,x',y'\}}(x, y)$. Apply $x' \sim_{\text{IR}} x''$ to get the desired result. We can similarly established the rest of the four conditions and conclude $(x, x'') \sim_{\text{ID}} (y, y')$.

To conclude we have established that $x \sim x''$. With transitivity proved, we have proved that $\sim$ is an equivalence relation, since symmetry trivially holds by definition. As a result, we can partition the set of alternatives X into multiple subsets according to $\sim$:

$$X = X_1 \cup X_2 \cup \cdots \cup X_K.$$

It remains to show that properties IR and ID hold if we define $s : X \to \{1, \cdots, K\}$ by $s(x) = i$ if $x \in X_i$. This follows immediately from the definition of $\sim$.

Chapter 4

Conclusions

Empirical and experimental findings suggest that the behavior of decision makers may not always seem rational. In particular, there could be preference reversals revealed by choice data, which suggest that a decision maker cannot be modeled as if she is maximizing a preference relation. The research projects in the dissertation focus on multi-dimensional decision problems and attempt to rationalize preference reversals in the sense that they attribute reversals to the nature of the decision problems instead of the irrationality of decision makers.

A decision maker may have limited ability to compare between two alternatives when an alternative is better than the other alternative in one dimension but worse in another dimension. When one of the alternative is the default option, the decision maker feels a sense of loss if she switches to a non-default alternative unless the latter dominates the default one in all dimensions. The extreme loss aversion leads to default bias. We study the implication of default bias and the novel trade-off between current consumption and future flexibility to switch in the context of dynamic decision problems, and propose the Default Bias Representation. We are able to characterize the representation with a collection of behavioral conditions and show the equivalence between them. In addition, we show that a decision maker with default bias may act preemptively in an asset replacement problem and monopolist may design contract accordingly to exploit the bias, which demonstrate the usefulness of our model.

To overcome the limited ability to compare in multi-dimensional comparisons, the decision maker may learn about the right trade-offs to make across dimensions from other more informed people. This learning process leads to seemingly irrational behavior that violates independence of irrelevant alternatives and monotonicity in choice probability, two appealing properties implied by classic models. We explain these violations by arguing that those irrelevant alternatives carry relevant information regarding the comparison between alternatives, and such information could be so strong that monotonicity in choice probabilities could be violated. This model, called the Attribute-Dependent Random Utility Maximization model, can be characterized by three simple behavioral conditions under a very mild assumption.

Bibliography

Aliprantis C, Border K (2006) *Infinite Dimensional Analysis* (Berlin, Germany: Springer-Verlag), 3 edition.

Ambrus A, Ishii Y (2015) On asynchronicity of moves and coordination, working paper.

Arcidiacono P, Bayer P, Blevins JR, Ellickson PB (2016) Estimation of dynamic discrete choice models in continuous time with an application to retail competition. *The Review of Economic Studies* 83:889–931.

Bewley TF (2002) Knightian decision theory. part i. *Decisions in economics and finance* 25:79–110.

Cailloux O, Destercke S (2017) Reasons and means to model preferences as incomplete. *International Conference on Scalable Uncertainty Management*, 17–30 (Springer).

Calvo GA (1983) Staggered prices in a utility-maximizing framework. *Journal of monetary Economics* 12:383–398.

Carroll GD, Choi JJ, Laibson D, Madrian BC, Metrick A (2009) Optimal defaults and active decisions. *Quarterly Journal of Economics* 124:1639–1674.

Debreu G (1960) Review of rd luce, individual choice behavior: A theoretical analysis. *American Economic Review* 50:186–188.

Dekel E, Lipman BL, Rustichini A (2001) Representing preferences with a unique subjective state space. *Econometrica* 69:891–934.

Dekel E, Lipman BL, Rustichini A, Sarver T (2007) Representing preferences with a unique subjective state space: A corrigendum. *Econometrica* 75:591–600.

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

Epstein LG, Zin SE (1989) Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica* 57:937–969.

Ghirardato P, Maccheroni F, Marinacci M (2004) Differentiating ambiguity and ambiguity attitude. *Journal of Economic Theory* 118:133–173.

Gul F, Natenzon P, Pesendorfer W (2014) Random choice as behavioral optimization. *Econometrica* 82:1873–1912.

Gul F, Pesendorfer W (2004) Self-control and the theory of consumption. *Econometrica* 72:119–158.

Hoff PD (2009) *A first course in Bayesian statistical methods* (Springer).

Huber J, Payne JW, Puto C (1982) Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of consumer research* 9:90–98.

Huber J, Puto C (1983) Market boundaries and product choice: Illustrating attraction and substitution effects. *Journal of Consumer Research* 10:31–44.

Johnson EJ, Goldstein D (2003) Do defaults save lives? *Science* 302:1338–1339.

Kahneman D, Knetsch JL, Thaler RH (1991) Anomalies: The endowment effect, loss aversion, and status quo bias. *The Journal of Economic Perspectives* 5:193–206.

Kamenica E (2008) Contextual inference in markets: On the informational content of product lines. *American Economic Review* 98:2127–2149.

Kovach M, Tserenjigmid G (2019) Behavioral foundations of nested stochastic choice and nested logit, working paper.

Kreps DM (1979) A representation theorem for 'preference for flexibility'. *Econometrica* 47:565–577.

Kreps DM, Porteus EL (1978) Temporal resolution of uncertainty and dynamic choice theory. *Econometrica* 46:185–200.

Lichtenstein S, Slovic P (1971) Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology* 89:46–55.

Loomes G, Sugden R (1983) A rationale for preference reversal. *American Economic Review* 73:428–432.

Luce RD (1959) *Individual choice behavior: A theoretical analysis* (Wiley).

Manzini P, Mariotti M (2014) Stochastic choice and consideration sets. *Econometrica* 82:1153–1176.

Masatlioglu Y, Nakajima D (2013) Choice by iterative search. *Theoretical Economics* 8:701–728.

Masatlioglu Y, Nakajima D, Ozbay EY (2012) Revealed attention. *American Economic Review* 102:2183–2205.

Masatlioglu Y, Ok EA (2005) Rational choice with status quo bias. *Journal of Economic Theory* 121:1–29.

Natenzon P (2019) Random choice and learning. *Journal of Political Economy* 127:419–457.

Noor J (2011) Temptation and revealed preference. *Econometrica* 79:601–644.

Rust J (1987) Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica* 999–1033.

Samuelson W, Zeckhauser R (1988) Status quo bias in decision making. *Journal of risk and uncertainty* 1:7–59.

Sarver T (2017) Mixture-averse preferences, working paper.

Seidl C (2002) Preference reversal. *Journal of Economic Surveys* 16:621–655.

Slovic P, Lichtenstein S (1983) Preference reversals: A broader perspective. *American Economic Review* 73:596–605.

Train KE (2009) *Discrete choice methods with simulation* (Cambridge university press).

Tversky A, Slovic P, Kahneman D (1990) The causes of preference reversal. *American Economic Review* 204–217.

Tversky A, Thaler RH (1990) Anomalies: Preference reversals. *Journal of Economic Perspectives* 4:201–211.