The Focal Luce Model*

Matthew Kovach† Gerelt Tserenjigmid ‡

First Version: December, 2016; Current Version: September 8, 2019

Abstract

We develop a random choice model in which a decision maker divides the alternatives she faces into two groups, where one group is focal and thus she is more likely to choose alternatives in that group relative to alternatives in the non-focal group: *The Focal Luce Model* (FLM). The FLM generalizes Luce's model and naturally captures bounded rationality models while behaviorally distinguishing between what is focal and the magnitude of the bias due to focality. We show how to identify utilities, focal sets, and bias terms from choice frequencies and behaviorally found the FLM with two weakenings of *independence from irrelevant alternatives* (IIA) which account for the decision maker's revealed focus. We introduce a new condition on focal sets, *Conditionally Decreasing*, which captures an increasing difficulty of focusing on multiple alternative and encompasses several models of limited consideration. We apply our model to experimental data, illustrate the importance of accounting for focus in demand estimation, and propose a simple method to identify focal sets.

Keywords: Random Choice; Focal Luce Model; IIA; Consideration Set; Conditionally Decreasing.

JEL Classification Numbers: D01, D81, D9.

1 Introduction

Many papers in economics, marketing, and psychology find considerable evidence that *focal alternatives* (or alternatives receiving additional attention), enjoy "excess demand."[1] For example, with many stores selling upwards of 40,000 items it is impossible for a consumer to give reasonable attention to everything, hence capturing focus is very important. This drives brands pay for eye-level placement on shelves or large in-store displays. From the many studies on the role of focus in

*We thank Federico Echenique, Pietro Ortoleva, Fabio Maccheroni, Yusufcan Masatlioglu, Matthew Shum, Jay Lu, Khai Chiong, Sergio Montero, Victor Aguiar, and the audiences of D-TEA 2017, NASM 2017, SAET 2017, Midwest Economic Theory 2017, TUS 2017, RUD 2018, 2018 CIREQ Montreal Microeconomic Theory Conference, Barcelona GSE 2019, and the seminar participants at Virginia Tech, Johns Hopkins University, NYU Shanghai, Shanghai University of Finance and Economics, and University of Rochester. An early version of this paper was previously circulated as The Imbalanced Luce Model.

†Department of Economics, Virginia Tech. E-mail: mkovach@vt.edu

‡Department of Economics, Virginia Tech. E-mail: gerelt@vt.edu

[1] For a survey on advertising and attention, see Bagwell (2007).

choice[2], a few general patterns of emerge: (1) there is no clear correlation between which objects that are focal and their consumption utility; (2) manipulating focality has a causal effect on choice; and (3) the degree to which something is focal is directly related to its likelihood of being chosen.[3]

Therefore, we introduce the *Focal Luce Model* (FLM), a generalization of Luce's (1959) model (multinomial logit) that accounts for a decision maker's focus. In particular, our model captures the effect of focality while allowing for menu-dependent focal alternatives and menu-dependent bias of focality. The necessity of this separation between "what is focused on" and "the magnitude of the bias towards focal alternatives" is borne out by the third finding above: increased focality leads to increased choice likelihood. This is also why it is essential to develop a theory of focality within the a stochastic choice framework, as deterministic models of choice cannot capture this distinction. Further, by utilizing a stochastic choice framework our model can easily applied to experimental and market data.

After a brief discussion of stochastic choice and the Luce model, we formally define the FLM in section 2 In the FLM, each menu, A, is partitioned into *focal alternatives*, $F(A)$, and *non-focal alternatives*, $A \setminus F(A)$. The ratio of choice frequencies for alternatives *within a single submenu*, say $F(A)$, is equal to the ratio of their utilities just as in Luce's model. However, we depart from Luce with respect to how the decision maker chooses between alternatives in different groups. Alternatives in $F(A)$ are chosen relatively more frequently than alternatives in $A \setminus F(A)$ after accounting for their utilities. In particular, Luce's utility $u(a)$ for alternative a is biased by $1+\delta(A)$ when a is focal in menu A; i.e., $a \in F(A)$, hence we can behaviorally distinguish between *what is focal* ($F(A)$) and the *impact of focality* ($\delta(A)$).

The FLM is a minimal departure from the Luce model/multinomial logit that is nonetheless flexible enough to accommodate well-known behavioral phenomena such as the compromise effect, choice overload, and status quo bias. Behaviorally, $\delta(A)$ captures the magnitude of these biases in menu A (e.g., the extent to which a decision maker wishes to avoid extreme alternatives or the degree of status quo bias). Since the FLM allows for violations of *regularity* (e.g., choice overload), the FLM is not a special case of the random utility model. However, the decision making procedure of the FLM is similar to that of the random utility model. Our main point of departure is that we allow the decision maker's utility over alternatives to exhibit a small degree of menu dependence.

[2]Focality also seems to play a significant role in strategic environments. Consider a symmetric 2×2 coordination game. When strategies are labelled generically, s_1 and s_2, players often fail to coordinate. However, as noted by Schelling (1960), the labelling of the strategies matters greatly. By changing the name of the game to "Meeting in New York" and referring to strategy one "Grand Central Station," while strategy two is some street intersection, subjects overwhelmingly succeed in coordinating on Grand Central Station. Schelling argues that this is because the labeling has made this strategy focal, or more prominent than the other.

[3]Reutskaja et al. (2011) demonstrates both points (1) and (3) in an eye-tracking experiment. Gossner et al. (2019) develops a theory in which manipulating attention towards an alternative leads to an increase in demand for that alternative regardless of its utility value, also supporting point (1). Armel et al. (2008) found that manipulating the length of time subjects could look at desirable alternatives influenced their choice frequency. Subjects were between 6 to 11% more likely to choose the alternative that was displayed for more time, providing evidence for points (2) and (3).

In the FLM, the (random) utility of a given alternative x in menu A is

$$v(x, A) = \bar{u}(x) + \bar{\delta}(A)\,\mathbb{1}\{x \in F(A)\} + \epsilon_x. \tag{1}$$

In the above equation, $\bar{u}(x)$ is standard and represents the fixed utility of x. The additional term, $\bar{\delta}(A)$, is the (menu-dependent) benefit of being a focal alternative.[4] Since $\bar{\delta}(A)$ impacts each alternative in the focal set equally, it may be interpreted as a common (fixed) utility shock that exclusively affects alternatives in $F(A)$, such as might occur after reading a news story that red wine is good for your health.[5] The decision maker then chooses the alternative with the highest utility according to random utility v. Moreover, as we will demonstrate, the FLM parameters can be estimated as easily as the Luce model (multinomial logit).

Our focal set is conceptually similar to the idea of a consideration set, however we do not assume that the decision maker completely fails to notice the non-focal alternatives. In most consideration sets models attention is binary[6] and "unattended" items can never be chosen. This is bit too strong of an assumption when modeling choice from smaller sets and it is difficult to fit experimental and market data with such a discrete model. In contrast, the FLM allows for different magnitudes of "attention," and thus distinguishes between focus and the impact of focality. This is important when taking the model to data, such as experimental data like Armel et al. (2008).

In some settings we may have reason to believe we know which alternatives are focal, such as in the two examples discussed earlier. However, in many choice settings the focal set $F(A)$ is not known *a priori* and must be inferred from choice data. Therefore, in section 3 we show how to identify focal sets, even when not all menus are observable. In order to identify focal sets in the FLM, we introduce a notion *revealed equally focal* (in menu A). We say two alternatives are revealed equally focal if their relative choice frequencies are the same in two different menus. With this notion, we can then identify cases where not all alternatives are equally focal. The focal set, $F(A)$, must consist of those alternatives which are chosen relatively more frequently than others. Hence, even without a specific theory of what makes something is focal, we can identify focal sets and the impact of focus, $\delta(A)$, just from choice frequencies. This result is especially useful when trying to distinguish between various theories of attention. We also introduce the *Binary Focal Luce Model* (BFLM), which assumes that for binary menus both alternatives are equally focal. This assumption allows for much simpler identification of focal sets; we can fully identify the focal set of any menu if we know sufficiently many binary choice frequencies.

We provide behavioral foundations for the FLM and the BFLM in section 4 and show that, in both models, the decision maker's utility index, focal set, and impact of focality can be identified

[4]Finally, ϵ_x is independently distributed according to the standard extreme value type I distribution as in Luce model.

[5]An alternative interpretation is that there are two (unobserved) types of agents, where one type maximizes over $F(A)$ and the other maximizes over A. In this case $\bar{\delta}(A)$ is related to the proportion of the population that only considers alternatives in $F(A)$.

[6]See for instance Masatlioglu et al. (2011), Cherepanov et al. (2013), and Lleras et al. (2017). We further discus our relation to consideration sets in section 5.

uniquely from her choice behavior. For simplicity, we will discuss the characterization BFLM first. The first axiom is a weakening of Luce's *independence from irrelevant alternatives* (IIA) axiom that accounts for differential focus. IIA requires that for all menus, A, and any two alternatives, a and b, the relative probability of choosing a over b from A is independent of other alternatives in A. Our axiom, *Focal IIA*, relaxes IIA to account for the decision maker's revealed focus and hence only requires IIA to hold between equally focal alternatives. Our second axiom, Luce's product rule (Luce (1959)), is a transitivity property that applies to choice from binary menus and is necessary to construct u. Theorem 1 shows that the above two axioms are necessary and sufficient. Theorem 2 characterizes the FLM by generalizing both Focal IIA and Luce's product rule in accordance with our revealed equally focal relation. Lastly, we characterize Increasing BFLM, which imposes a monotonicity property on δ: $A \supset A'$ implies $\delta(A) \geq \delta(A')$. This captures an increasing tendency to ignore non-focal alternatives as the menus gets larger (consistent with choice overload or increasing status quo bias) without restricting which alternatives are focal.

In addition to parsimony, we believe that a single focal set is quite natural. Firstly, it has an intuitive psychological interpretation based on dual self models, instinctive and contemplative types (Rubinstein, 2016), or Kahneman's model of system I and system II thinking. Secondly, in many examples the natural division of alternatives is binary. For example, the decision maker may focus on Pareto efficient, risk free, or unambiguous alternatives. Thirdly, the FLM with a single focal set naturally relates to limited consideration, and thus provides a structural framework for estimating consideration sets empirically.

In section 5 we impose additional structure on the model parameters, F and δ, by restricting how focus changes across menus. Hence a third contribution of our paper is the introduction of a new property on consideration sets that unites the models of Masatlioglu et al. (2011), Cherepanov et al. (2013), and Lleras et al. (2017). This new property, which we call *Conditionally Decreasing*, requires that if a new alternative is non-focal, then the set of focal alternatives may only shrink: $F(A) \subseteq B \subseteq A$ implies $F(A) \subseteq F(B)$. This property captures increasing difficulty of focus as the number of alternatives increases. We introduce a single axiom, *Focal Betweenness*, which jointly characterizes conditionally decreasing F and increasing δ, suggesting an intimate link between these behavioral properties.

We then develop several applications of the FLM in section 6. In our first application, we show that the FLM captures several features from choice overload and choice with a status quo. A decision maker displays the choice overload effect (increasing status quo bias) when adding alternatives to a menu increases her probability of "walking away" (choosing the status quo) from the choice problem. In our example the probability of choosing the default option is U-shaped in the menu size.[7] Our second application takes the FLM to experimental data on consumer choice from Kivetz et al. (2004), which documents more nuanced behavioral regularities associated with the compromise effect (e.g., a shifting compromise and multiple compromises). Our third application

[7]The ability to capture thus U-shaped choice frequency, which is consistent with empirical findings, is something lacking in deterministic models of limited consideration. This is because deterministic models cannot distinguish between what is focal and the magnitude of the bias from focality.

illustrates how the FLM can be used to correct for the role of focus (or attention) when estimating price elasticity of demand.[8] Our final results show in section 7 that the utilities of alternatives, focal sets, and magnitudes of the bias from focality can be identified from observed choice frequencies by solving simple linear programming and least squares problems. Moreover, we show the consistency of our estimators for linear utility models.

We close our paper with a discussion of related literature in section 8. The proofs are collected in Appendix A. In an online appendix, we discuss the maximum likelihood estimation for the FLM as well as an axiomatic characterization of a deterministic version of the FLM.

2 The Model

2.1 Setup

We closely follow the standard random choice setup. Let X be a finite set of alternatives and $\mathscr{A}$ be a collection of nonempty subsets of X (menus). Unless it is specifically specified, we require that $\mathscr{A}$ includes all menus with two alternatives. Let $\mathbb{R}_+$ ($\mathbb{R}_{++}$) denote the non-negative (positive) real numbers and $\mathbb{1}\{\cdot\}$ be the indicator function.

Definition 1. A function $p : X \times \mathscr{A} \to [0, 1]$ is called a *random choice rule* if for any $A \in \mathscr{A}$,

$$\sum_{a \in A} p(a, A) = 1.$$

A random choice rule p is *positive* if $p(a, A) > 0$ for all $A \in \mathscr{A}$ and $a \in A$.

Throughout this paper we focus on positive random choice rules.[9] For notational simplicity, we will denote the probability ratio of a and b as $r(a, b) \equiv \frac{p(a,ab)}{p(b,ab)}$ and the probability ratio of a and b in menu A as $r_A(a, b) \equiv \frac{p(a,A)}{p(b,A)}$. Before we introduce our axioms and representation, we first provide a brief review of the Luce model and Luce's main axiom, *Independence of Irrelevant Alternatives* (IIA).

Definition 2. A random choice rule p satisfies Luce's *independence of irrelevant alternatives (IIA) axiom at a, b in menu A* if

$$r(a, b) = r_A(a, b).$$

Moreover, p satisfies IIA if for all $A \in \mathscr{A}$ and all $a, b \in A$, p satisfies IIA at a, b in A.

Luce (1959) proves that if a positive random choice rule satisfies IIA, then it can be represented by the Luce model (also referred to as multinomial logit), which we now define.

[8] In a companion paper, Kovach and Tserenjigmid (2018), we use the FLM to study the implications of bounded rationality (due to limited consideration, salience, or regret aversion) in strategic decision making. In particular, we combine the FLM with the celebrated quantal response equilibrium of McKelvey and Palfrey (1995), thereby creating the *Focal Quantal Response Equilibrium* (F-QRE). The F-QRE can explain much of the observed heterogeneity in play across different games.

[9] It is not difficult to extend the FLM to allow for zero probabilities. Formal discussions are available upon request.

Definition 3. A random choice rule p is a *Luce Model* if there exists a $u : X \to \mathbb{R}_{++}$ such that for any $A \in \mathscr{A}$ and $x \in A$,

$$p(x, A) = \frac{u(x)}{\sum_{a \in A} u(a)}. \tag{2}$$

Luce (1959) also introduced an axiom called *Product Rule* which only applies to two-alternative menus.

Definition 4 (LPR)**.** A random choice rule p satisfies *Luce's Product Rule* (LPR) if for any $a, b, c \in X$, $r(a, c) = r(a, b) \cdot r(b, c)$.

LPR is a weakening of IIA. Luce (1959) proves that if a positive random choice rule satisfies LPR, then it can be represented by the Luce model on binary menus. Hence, LPR can be understood as IIA for binary menus.[10]

We also formally define regularity, a monotonicity property that is satisfied by any random utility model.

Definition 5. A random choice rule p satisfies *Regularity* if for any $A, B \in \mathscr{A}$ and $x \in B \subset A$, $p(x, B) \geq p(x, A)$.

2.2 Model

To define our model, we define a function that specifies focal alternatives for each menu. A mapping $F : \mathscr{A} \to 2^X \setminus \{\emptyset\}$ is a *focus function* if $F(A) \subseteq A$ for any $A \in \mathscr{A}$. A focus function $F : \mathscr{A} \to 2^X \setminus \{\emptyset\}$ is *binary* if $F(\{a, b\}) = \{a, b\}$ for any $a, b \in X$.

Definition 6. A random choice rule p is a *Focal Luce Model* (FLM) if there exist a utility function $u : X \to \mathbb{R}_{++}$, a focus function $F : \mathscr{A} \to 2^X \setminus \{\emptyset\}$, and a distortion function $\delta : \mathscr{A} \to \mathbb{R}_{++}$, such that for any $A \in \mathscr{A}$ and $a \in A$,

$$p(a, A) = \frac{u(a)\big(1 + \mathbb{1}\{a \in F(A)\}\, \delta(A)\big)}{\sum_{b \in A} u(b)\big(1 + \mathbb{1}\{b \in F(A)\}\, \delta(A)\big)}. \tag{3}$$

Moreover, a random choice rule p is a *Binary Focal Luce Model* (BFLM) if p is a FLM with respect to some binary focus function.

The FLM deviates from Luce's model in the following way. The decision maker divides A into $F(A)$ and $A \backslash F(A)$, and the choice frequencies of alternatives in $F(A)$ relative to alternatives in $A \backslash F(A)$ are distorted by $1 + \delta(A)$. Note that for any A, if $F(A) = A$, then the FLM reduces to Luce's model in menu A.

[10] LPR implies *strong stochastic transitivity*: if $r(a, b) \geq 1$ and $r(b, c) \geq 1$, then $r(a, c) \geq \max\{r(a, b), r(b, c)\}$. Hence LPR might be thought of as a stochastic counterpart to "no binary cycles" conditions which are often used in deterministic choice. For instance, see Ok et al. (2014). See He and Natenzon (2018) for discussion on how the various stochastic transitivity properties are related.

Alternatives/ Menus	$\{x, y\}$	$\{x, y, z\}$
x	0.50	0.22
y	0.50	0.57
z	—	0.21

Figure 1: Choice frequencies for 35mm Minolta Camera models in Simonson and Tversky (1992)

In the BFLM, we have $p(a, ab) = \frac{u(a)}{u(a)+u(b)}$ and $p(b, ab) = \frac{u(b)}{u(a)+u(b)}$ as in Luce's model. Thus in the BFLM there is no distortion due to focus in menus with two alternatives. This is intuitive since it is easier to choose between two alternatives than from among many, hence we should not expect focus to come into play.[11]

Since the FLM is a model of stochastic choice, it is open to various interpretations as to the source of randomness. We offer three. The first two are typical in the literature. Any FLM (u, F, δ) is equivalent to the maximization of a random, menu-dependent utility $v(x, A) = \bar{u}(x) + \bar{\delta}(A)\, \mathbb{1}\{x \in F(A)\} + \epsilon_x$, where each ϵ_x is independently distributed according to the standard extreme value type I distribution, $\bar{u}(x) = \log(u(x))$ and $\bar{\delta}(A) = \log(1 + \delta(A))$. Therefore we may interpret the FLM as (i) a model of mistakes or choice tendency of an individual decision maker or as (ii) a model of population level taste heterogeneity. In the third interpretation, the FLM may be interpreted as a model of two types of consumers, one which chooses from $F(A)$ and one which chooses from A. Each type however is a Luce DM, and thus our model captures two distinct forms of individual heterogeneity. In particular, let $p_1(a, A) \equiv \frac{u(a)}{\sum_{b \in A} u(b)}$ and $p_2(a, A) \equiv \frac{u(a) \mathbb{1}\{a \in F(A)\}}{\sum_{b \in F(A)} u(b)}$. Then there is $\lambda(A) \in [0, 1)$ such that $p(a, A) = \big(1 - \lambda(A)\big)\, p_1(a, A) + \lambda(A)\, p_2(a, A)$.[12]

2.3 Discussion

To illustrate our approach to incorporating focus, consider the following choice frequency data on the *compromise effect* from Simonson and Tversky (1992) (Figure 1).[13]

In their experiment, the alternatives are 35mm Minolta Camera models, where x is of low quality with a low price, y is of medium quality with a medium price, and z is of high quality with a high price. The most important observation here is that the introduction of z causes a

[11] There is a view that binary choices reveal the DM's true preferences or utilities. Further, an experiment by Mariotti and Manzini (2010) suggests that most violations of WARP are due to menu-effects, in line with our approach.

[12] By direct calculations, we find that $\lambda(A) = \frac{\delta(A) \sum_{a \in F(A)} u(a)}{\sum_{a \in A} u(a) + \delta(A) \sum_{a \in F(A)} u(a)}$. The assumption of a menu-dependent weight $\lambda(A)$ is quite natural in many instances. For example, it may become increasingly difficult to consider all alternatives as the total number of alternatives increases and thus the DM becomes more likely to only consider a subset of alternatives. This is captured when $\lambda(A)$ is increasing in set containment. When the magnitude of the bias is large enough, alternatives in the non-focal group will be chosen with probabilities near zero, and thus we can approximate nearly any deterministic choice function and Luce rule over consideration sets with the FLM. Lastly, notice that a menu-independent (population) weight, $\lambda(A) \equiv \lambda$, still results in a menu-dependent FLM distortion, $\delta(A) = \frac{\lambda(A)}{1-\lambda(A)} \frac{\sum_{a \in A} u(a)}{\sum_{b \in F(A)} u(b)}$. Hence a fixed distribution of types throughout the population necessitates menu-dependence of FLM parameters.

[13] The compromise effect was documented and confirmed in the experimental studies of Simonson (1989), Simonson and Tversky (1992), Herne (1997), and Chernev (2004), among many others.

violation of IIA between x and y. Specifically, y is now chosen relatively more frequently than x: $r_{\{x,y,z\}}(y,x) > r(y,x)$. The common explanation for the compromise effect is that x and z are both extreme alternatives, which decision makers do not like, and thus they prefer the compromise good, y.

The FLM can naturally capture the compromise effect choice pattern shown earlier. Suppose $F(\{x,y\}) = \{x,y\}$ and $F(\{x,y,z\}) = \{y\}$. In words, the compromise good y is focal and extreme goods x and y are nonfocal in $\{x,y,z\}$ Then the FLM choice probabilities for x and y are

$$p(x,\{x,y,z\}) = \frac{u(x)}{u(x)+u(z)+u(y)(1+\delta)} \text{ and } p(y,\{x,y,z\}) = \frac{u(y)(1+\delta)}{u(x)+u(z)+u(y)(1+\delta)}.$$

Therefore, adding the third alternative z increases the relative probability of choosing y over x: $r(y,x) = \frac{u(y)}{u(x)} < r_{\{x,y,z\}}(y,x) = \frac{u(y)(1+\delta)}{u(x)}$.

Additionally, the compromise effect example from Simonson and Tversky (1992) features a violation of regularity, $p(y,\{x,y,z\}) > p(y,\{x,y\})$, and thus their data cannot be explained by any random utility model. The FLM however can handle this, as it should be obvious that for large values of δ the FLM may violate regularity. Hence the FLM is not a special case of random utility.

3 Inferring Focal Alternatives

Before we axiomatically characterize the FLM, we discuss several ways to identify focal alternatives from choice data. We discuss how to identify focal sets by solving linear programming problems and also show the consistency of least squares estimators for linear utility models in section 7.

In our model, IIA can be violated because some alternatives are more focal than others. Hence, focal alternatives can be identified from violations of IIA as formalized below.

Proposition 1. *Suppose p is a FLM with (u,F,δ). For any $A,B \in \mathscr{A}$ and $a,b \in A \cap B$, if $r_A(a,b) > r_B(a,b)$, then either $a \in F(A)$ or $b \in F(B)$.*

In words, $r_A(a,b) > r_B(a,b)$ reveals that either a is focal in A or b is focal in B. However, it does not necessarily imply that b is nonfocal in A. Below we discuss ways to make sharper conclusions about focal alternatives. We start with the BFLM since identifying focal alternatives is simpler when the focus function is binary.

3.1 Focal Alternatives for Binary FLM

For the BFLM, it is enough to compare $r_A(a,b)$ with $r(a,b)$ as formalized below.

Proposition 2. *Suppose p is a BFLM with (u,F,δ). Take any $A,B \in \mathscr{A}$ and $a,b \in A \cap B$.*

i) $r_A(a,b) > r(a,b)$ if and only if $a \in F(A)$ and $b \notin F(A)$.

ii) Suppose $r_A(b,c) > r(b,c)$ for some $c \in A$. If $r_A(a,b) \geq r_B(a,b)$, then either $a \in F(A)$ or $a \notin F(B)$.

iii) Suppose $r_A(b,c) > r(b,c)$ for some $c \in A \cap B$. If $r_A(a,c) \leq r_B(a,c)$, then either $a \notin F(A)$ or $a \in F(B)$.

As stated in Proposition 2.i), when choice frequencies from binary menus are observed, it is straightforward to identify focal sets. In fact, in Section 4.1 our characterization theorem for the BFLM shows that we can completely identify everything when we observe choice probabilities from all binary menus.[14] However, many instances we only have limited data. To that end, Proposition 2.ii-iii) deal with cases where not all binary choices are observed.

3.2 Inferring Focal Sets for General FLM

In the BFLM, a and b are always "equally focal" in $\{a,b\}$ since $p(a,\{a,b\}) = \frac{u(a)}{u(a)+u(b)}$. Hence, $r_A(a,b) > r(a,b)$ is equivalent to a being focal and b being nonfocal in A.

However, for the general FLM, $r_A(a,b) > r(a,b)$ may not be enough to conclude that a is more focal than b in A. For example, if a and b are equally focal in A, but b is more focal than a in $\{a,b\}$, then we can also have $r_A(a,b) > r(a,b)$. However, when $r_A(a,b) > r_B(a,b)$ and a and b are equally focal in B, we can conclude that a is more focal than b in A. Hence, we provide a definition for *revealed equally focal*.

Revealed Equally Focal. For any two alternatives $a, b \in A$, we say that a and b are *revealed equally focal in menu A*, denoted by $a \sim_A b$, if there is a menu B such that $r_A(a,b) = r_B(a,b)$.[15]

One important feature of this definition is that it does not rely on binary menus. Thus we can identify focality when binary menus are not observed, which is often the case with market data.

Proposition 3. *Suppose p is a FLM with (u, F, δ). Suppose $\delta(A') \neq \delta(A'')$ for any $A', A'' \in \mathscr{A}$. For any $A, B \in \mathscr{A}$ and $a, b \in A \cap B$,*

i) If $a \sim_B b$, then $r_A(a,b) > r_B(a,b)$ if and only if $a \in F(A)$ and $b \notin F(A)$.

ii) Take sequences $\{a_i\}_{i=1}^{n}$ and $\{A_i\}_{i=1}^{n-1}$ such that $a = a_1, b = a_n$, and $a_i \sim_{A_i} a_{i+1}$ for each $i < n$. Then $r_A(a,b) > \prod_{i=1}^{n-1} r_{A_i}(a_i, a_{i+1})$ if and only if $a \in F(A)$ and $b \notin F(A)$.

One appeal of Proposition 3 is that we can identify focal alternatives without assumptions on F. Proposition 3.i) says that if IIA is violated between a, b at A, B in favor of a and a and b are revealed equally focal in B, then a is focal in A and b is nonfocal in A. Proposition 3.ii) extends the first part to a sequence of alternatives. Indeed, when there are more restrictions on F and δ, focal alternatives can identified by many other alternatives ways.[16]

[14] More specifically, when $\mathscr{A}$ contains all binary menus, u and F are uniquely identified, and for each $A \in \mathscr{A}$, $\delta(A)$ is uniquely identified when IIA is violated at A.

[15] By symmetry, a and b are equally focal in B as well: $a \sim_B b$.

[16] One interesting possibility is to assume that $\delta(A) > \delta(\{a,b\})$ for any A and $\{a,b\} \subset A$. This assumption is justified by increasing difficulty of focus or arguments from footnote 12. In this case, for any A and $a \in A$, $r_A(a,b) \geq r(a,b)$ for any $b \in A$ if and only if $a \in F(A)$.

4 Behavioral Foundations

4.1 Behavioral Foundations of the Binary Focal Luce Model

Rather than make arbitrary assumptions about how focus is determined, we take the view that changes in focus are revealed by violations of IIA. Recall the compromise effect example in Section 2.3. In this example, adding a third alternative z to $\{x, y\}$ makes y a compromise alternative and x, z extreme alternatives and hence y will be chosen more frequently relative to x. In line with this explanation, we argue observed choice frequencies reveal that y is now focal relative to x (and z) in menu $\{x, y, z\}$. Since both x and z are non-focal (because they are both extreme), they ought to be "treated equally" and IIA should hold between them.

Our first axiom therefore disciplines how changes in focus change choice behavior. Since IIA violations reveal relative focus, our axiom requires that IIA hold among alternatives that are revealed equally focal. Hence our axiom places a restriction on when there may be an IIA violation. Put another way, suppose that for some menu A, a is revealed more focal than b and c. Then b and c are both non-focal and hence IIA must hold between them. This is precisely the content of our axiom, stated below.

Axiom 1 (Focal IIA). For any $A \in \mathscr{A}$ and $a, b, c \in A$,

$$\text{if } r_A(a, b) > r(a, b) \text{ and } r_A(a, c) > r(a, c), \text{ then } r_A(b, c) = r(b, c).$$

In terms of the compromise effect example above, Focal IIA suggests that if the compromise alternative, y is focal, then extreme alternatives, x and z, must be treated equally: $r_{\{x,y,z\}}(x, z) = r(x, z)$.

The second axiom is LPR (which is defined in Section 2.1). We now state our first characterization theorem, the proof of which is in A.4.

Theorem 1. *A positive random choice rule p satisfies Focal IIA and LPR if and only if p is a BFLM.*

Proposition 4. *Suppose there are two tuples (u, F, δ) and (u', F', δ') such that p is a BFLM with (u, F, δ) as well as p is a BFLM with (u', F', δ'). Then (i) $F = F'$, (ii) there is $\alpha > 0$ such that $u' = \alpha\, u$, and (iii) $\delta(A) = \delta'(A)$ for any $A \in \mathscr{A}$ where IIA is violated at A.*

Remark. As is common in rational choice theory, we require certain assumptions on the domain $\mathscr{A}$ to obtain the characterization result. In Theorem 1, we assume for simplicity that $\mathscr{A}$ contains all binary menus. However, there is no assumption on larger menus. Moreover, we do not actually need all binary menus. For example, it is enough to assume that there is $x^* \in X$ such that $\{x, x^*\} \in \mathscr{A}$ for any $x \neq x^*$.[17] Proposition 5 in Section 4.1 demonstrates that the FLM can be rejected when

[17] In fact, it is enough to assume that any two alternatives in a menu can be connected by a sequence of binary menus. Formally, for any $a, b \in X$, there is a sequence $a_1, \ldots, a_n$ such that $a_1 = a, a_n = b$ and $\{a_i, a_{i+1}\} \in \mathscr{A}$. This assumption can be relaxed further but the uniqueness result in Theorem 1 might be weaker in this case.

$\mathscr{A}$ contains only two menus. Theorem 3 provides a characterization theorem for the FLM without the assumption that $\mathscr{A}$ contains all binary menus.

In the FLM, $\delta(A)$ captures the magnitude of the departure from IIA, and thus tells us the extra weight given to focal alternatives in menu A. The FLM is able to separate *what is focal* from *how much focus biases choice*. Motivated by the literature on choice overload and limited attention, we impose a monotonicity property on δ so that the bias toward focal alternatives increases as the number of alternatives increases. When there are many alternatives, the DM realizes that she cannot carefully consider all alternatives and she focuses more carefully on just a few. This is captured by increasing δ, defined below.

Definition 7. We say that δ is *increasing* if $\delta(A) \geq \delta(B)$ for any $A, B \in \mathscr{A}$ with $B \subseteq A$.

Axiom 2 (Increasing Focal Impact). For any $B \subseteq A$, $a, b \in A$, $a', b' \in B$, if

$$r_A(a,b) > r(a,b) \text{ and } r_B(a',b') > r(a',b'), \text{ then } \frac{r_A(a,b)}{r(a,b)} \geq \frac{r_B(a',b')}{r(a',b')}.$$

Our next result shows that, when combined with LPR, Increasing Focal Impact actually implies Focal IIA by taking $a = a'$, $b' = c$, and $A = B$.

Theorem 2. *A positive random choice rule p is satisfies LPR and Increasing Focal Impact if and only if p is an FLM with some (u, F, δ) and δ is increasing. Moreover, F is unique, u is unique up to multiplication by a positive scalar, and $\delta(A)$ is unique if there is a violation of IIA in menu A.*

Note that we may reverse the inequality in Increasing Focal Impact to obtain a decreasing δ. Hence by imposing equality we obtain a constant, or menu-independent, δ. Although it may be desirable to assume a constant weight in some settings, there are many important examples that require a menu-dependent $\delta(A)$ (see footnote 12). In section 5, we show that consideration set models, which are intimately linked with the FLM, typically lead to increasing bias.

4.2 Behavioral Foundations of the Focal Luce Model

Without the assumption of equal focus in binary menus, we need a different way to identify when an alternative has become focal in some menu. Hence recall our definition of revealed equally focal from section 3. Given this definition, we can modify Focal IIA and Luce's Product Rule in the following way. Roughly speaking, Focal IIA requires that if a is more focal than b in A as well as a is more focal than c in A, then b and c are equally focal in A. Generalized Focal IIA requires that when a and b are not equally focal in A and a and c are not equally focal in A, then b and c are equally focal in A. Formally,

Axiom 3 (Generalized Focal IIA). For any $A \in \mathscr{A}$ and $a, b, c \in A$, if $a \nsim_A b$ and $a \nsim_A c$, then $b \sim_A c$.

Roughly speaking, LPR requires that the impact of c on b times the impact of b on a is equal to the impact of c on a. Generalized Product Rule requires a similar condition to hold only when any two of a, b, c are equally focal in some menu. Formally,

AXIOM 4 (Generalized Product Rule). For any $A, B, C \in \mathscr{A}$ and $a, b, c \in X$, if $a \sim_A b$, $b \sim_B c$, and $a \sim_C c$, then

$$r_A(a,b) \cdot r_B(b,c) \cdot r_C(c,a) = 1.$$

We also need the following richness condition on p.

Definition 8 (Focal Richness)**.** A stochastic choice function p is *focally rich* if for any $x, y \in X$, $x \sim_A y$ for some $A \in \mathscr{A}$.[18]

Focal Richness requires that for any two alternatives there are at least two menus in which they are both equally focal, i.e., they have equal probability ratios (reminiscent of IIA). It turns out that under Focal Richness, our two axioms are necessary and sufficient for the FLM representation.

Theorem 3. *Suppose a stochastic choice function p is focally rich. Then p satisfies Generalized Focal IIA and Generalized Product Rule if and only if there exist a utility function $u\colon X \to \mathbb{R}_{++}$, a focus function F, and a distortion function $\delta\colon \mathscr{A} \to \mathbb{R}_{++}$, such that p is a FLM with $(u, F, \delta)$and $\delta(A') \neq \delta(A'')$ for any $A', A'' \in \mathscr{A}$ with $F(A') \cap F(A'') \neq \emptyset$ and $A' \cap A'' \setminus F(A') \setminus F(A'') \neq \emptyset$. Moreover, F is unique, u is unique up to multiplication by a positive scalar, and $\delta(A)$ is unique if there is a violation of IIA in menu A.*

Focal Richness is required for the characterization of the FLM. However, to reject the FLM, we only need to observe choice frequencies from two menus.

Proposition 5. *Suppose p is an FLM. For any $A, B \in \mathscr{A}$ and $x, y, z, t \in A \cap B$,*

$$\text{if } \frac{r_A(x,y)}{r_B(x,y)} > \frac{r_A(y,z)}{r_B(y,z)} > \frac{r_A(z,t)}{r_B(z,t)}, \text{ then } 1 \geq \frac{r_A(z,t)}{r_B(z,t)}.$$

In subsection 6.2, we show how to apply the result in Proposition 5 by using experimental data to test the FLM.

5 A Theory of Focus: Properties of F and δ

The FLM is permissive when it comes to changing focal sets across menus. Hence, an important contribution of our paper is to show how to identify focal alternatives without assumptions on focal sets. However, it is useful to have restrictions on F and δ to make sharper predictions out of sample. In this section we develop a more structured theory of focus for the FLM by introducing a novel property on F that encompasses many models of bounded rationality.

[18] Proposition 3 illustrates that Focal Richness can be relaxed in the following way: for any $a, b \in X$, there are sequences $\{a_i\}_{i=1}^n$ and $\{A_i\}_{i=1}^{n-1}$ such that $a = a_1, b = a_n$, and $a_i \sim_{A_i} a_{i+1}$ for each $i < n$. Even under this weaker richness condition, we obtain the same uniqueness result.

Motivated by the idea of cognitive load, the choice overload effect, and robust evidence that decision makers resort to simplifying heuristics in large menus (e.g., Payne (1976), Payne et al. (1993)), we propose the following monotonicity property on F: a DM can focus on relatively fewer alternatives when choosing from larger menus. For intuition, consider how adding a new alternative a changes focal alternatives in B. If the new alternative a is focal in $B \cup \{a\}$, which might occur if a is very salient or desirable, then focus might shift with the introduction of a. However, if a is not that salient or desirable, i.e., conditional on a not attracting focus, then $F(B \cup \{a\})$ should be no larger than $F(B)$ as it is increasingly difficult to focus. That is, $F(B \cup \{a\}) \subseteq F(B)$ if $a \notin F(B \cup \{a\})$. Formally,

Definition 9. A focus function F is *conditionally decreasing* if for any $A, B \in \mathscr{A}$, $F(A) \subseteq B \subset A$ implies $F(A) \subseteq F(B)$.

It turns out that our conditionally decreasing property "nests" two separate classes of consideration set models. In these two models, $C(A) = \arg\max_{\succ} F(A)$ for some filter function F with $F(A) \subseteq A$ for any $A \in \mathscr{A}$. The following two properties on F have been studied carefully in the literature.

- **Attention Filter:** (see Masatlioglu et al. (2011)) F is an *attention filter* if, for any A and $x \notin F(A)$, $F(A \setminus x) = F(A)$.
- **Competition Filter:** (see Cherepanov et al. (2013) and Lleras et al. (2017)) F is a *competition filter* if, for any A, B with $B \subset A$, $F(A) \cap B \subseteq F(B)$.

It is not difficult to see that both the attention filter and the competition filter properties are distinct yet are also conditionally decreasing.[19] Masatlioglu et al. (2011), Cherepanov et al. (2013), and Lleras et al. (2017) provide many interesting choice procedures that are captured by attention filter and competition filter: Status Quo Bias, Shortlisting/Social Preferences, Top-N choice, and Willpower/temptation. Hence, these procedures are also covered by the conditionally decreasing property.

- **Status Quo Bias:** (see Masatlioglu and Ok (2013) and Apesteguia and Ballester (2013)) A DM has a default (or status quo) option $d \in X$ and her choices are biased towards d and alternatives that are obviously better than d; i.e., $F(A) = A \cap Q(d)$ for some closed-valued self-correspondence Q on X. In this setting, $\delta(A)$ also captures a menu-dependent *intensity of status quo bias.*
- **Top-N choice:** (see Eliaz et al. (2011)) A DM only considers Top-N alternatives; i.e., $F(A) = \{x \in A : |\{y \in A : y \succ x\}| < N\}$ for some linear order $\succ$ on X.[20]

[19] We also include characterizations of the Attention Filter and Competition Filter conditions for F in an online appendix.
[20] A binary relation $\succ$ is a linear order if it is complete, transitive, and asymmetric.

- **Shortlisting/Social Preferences:** (see Manzini and Mariotti (2007)) A DM first eliminates alternatives that are obviously dominated by a binary relation P and then she maximizes her utility; i.e., $F(A) = \{x \in A : \nexists y \in A \text{ s.t. } y \, P \, x\}$. The binary relation P can also capture DM's social preferences. We include a characterization for this F in an online appendix.

We will add an example that is not discussed in Masatlioglu et al. (2011), Cherepanov et al. (2013) and Lleras et al. (2017).

- **Perception-Based Reference Dependence:** A DM notices some alternative based on its perceptual attractiveness, then focuses on alternatives which are sufficiently better than this reference alternative. Formally, for a given linear perception order P on X, the reference alternative $R(A)$ in A is the P-maximal element of A.[21] Then let $F(A) \equiv \{x \in A : v(x) \geq v(R(A)) + \theta(A)\}$ for some increasing $\theta : 2^X \to (-\infty, 0]$ (i.e., $\theta(B) \leq \theta(A)$ when $B \subset A$).

The above example is conditionally decreasing, yet it may violate both the attention filter and competition filter properties.

While the above properties capture the entire content of "increasing difficulty of focus" for deterministic choice, the FLM distinguishes what is focal from how much focus biases choice. Therefore in addition to our restrictions on F, a theory of focus must also restrict how δ changes. Just as F was motivated by the literature on choice overload and cognitive load, our restriction on δ is similarly derived. Our conditionally decreasing property restricts F when a new alternative is non-focal, hence it stands to reason that we should similarly restrict δ. That is, we also impose that δ is weakly increasing when a new alternative is non-focal, defined below.

Definition 10. A distortion function δ is *conditionally increasing* if $\delta(A) \geq \delta(B)$ for any $A, B \in \mathscr{A}$ with $F(A) \subseteq B \subset A$.

We now characterize conditionally decreasing F and conditionally increasing δ. For simplicity, we focus on the BFLM. While the above monotonicity properties on F and δ might seem behaviorally distinct, it turns out that they are simultaneously characterized by the following "focal betweenness" condition.

Axiom 5 (Focal Betweenness). For any $A \in \mathscr{A}$ and $a, b, c \in A$,

$$\text{if } r_A(a,b) > r(a,b), \text{ then } r_A(a,c) \geq r_{A \setminus \{b\}}(a,c) \geq r(a,c).$$

Focal Betweenness requires the relative probability $r_{A \setminus \{b\}}(a,c)$ to be between that of $r_A(a,c)$ and $r(a,c)$ if IIA is violated in favor of a in A.

Proposition 6. *Suppose p is a BFLM with (F, δ) such that $F(A) \neq A$ for any $|A| \geq 3$. Then p satisfies Focal Betweenness if and only if F is conditionally decreasing and δ is conditionally increasing.*

[21] This reference formation procedure is studied in Kibris et al. (2018).

Remark 1. Recall the interpretation of the FLM in which $p(a,A) = \big(1-\lambda(A)\big)\, p_1(a,A) + \lambda(A)\, p_2(a,A)$ where p_1, p_2 follow Luce rules. Then $\delta(A) = \frac{\lambda(A)}{1-\lambda(A)} \frac{\sum_{a\in A} u(a)}{\sum_{b\in F(A)} u(b)}$. Let λ be a constant. Then δ is conditionally increasing.

Remark 2. Consider a fairly known generalization of the Luce model, called *dogit*, introduced by Gaundry and Dagenais (1979). In dogit, $p(a,A) = \frac{u(a)+\theta(a)(\sum_{b\in A} u(b))}{(1+\sum_{b\in A}\theta(b))(\sum_{b\in A} u(b))}$ for any A and $a \in A$, where $\theta(a)$ is a measure of consumer loyalty to alternative a. Let X_1, X_2 be a partition of X such that $\theta(a) = \gamma\, u(a)\mathbb{1}\{a \in X_1\}$ for some $\gamma > 0$. For example, X_1 is the set of brand goods that are market leaders or consumers are loyal to, and X_2 is the set of other goods. Then we obtain an FLM with (F,δ) where $F(A) = A \cap X_1$ and $\delta(A) = \gamma\,(\sum_{b\in A} u(b))$. Notice that F is a Competition Filter and δ is increasing.

5.1 Limits of Conditionally Decreasing F

This section illustrates that the Conditionally Decreasing property is not excessively permissive. In particular, we provide two examples, one is based on the idea of Salience Theory (Bordalo et al. (2013)) and another is based on reference-dependence behavior, in which Conditionally Decreasing is violated.

Salience. Let $A = \{r_1, r_2, b_1, b_2, \ldots, b_n\}$, consist of two red alternatives and many black ones, while $B = \{r_1, r_2, b_1\}$, consists of the same two red alternatives and a single black one. Intuitively, in A, the red alternatives stand out and hence the red ones are focal: $F(A) = \{r_1, r_2\}$. By removing most of the black alternatives, B now consists of mostly red options, and focus shifts entirely to black alternatives: $F(B) = \{b_1\}$. Notice that $F(A) \subset B \subset A$ but $F(A) \not\subseteq F(B)$, violating the Conditionally Decreasing property. In general, the salience model will violate Conditionally Decreasing F, as it allows focus to change in a general, menu-dependent fashion.

General Reference Dependence. The decision maker chooses an endogenous reference alternative $R(A)$ from the menu A, and divides A into $F(A) = \{a \in A | a \succ R(A)\}$, the set of alternatives that dominate $R(A)$ by some strict partial order $\succ$, and $A \backslash F(A)$, the set of alternatives that do not dominate $R(A)$.[22] Then, the utilities of dominating alternatives in $F(A)$ will be overweighted by $1+\delta(A)$ (In other words, alternatives worse than $R(A)$ are punished by $1+\delta(A)$, for example due to loss aversion). We include a characterization of this in an online appendix.

6 Applications

This section presents several applications of the FLM. For our first application we demonstrate that the FLM can naturally accommodate choice overload and (increasing) status quo bias: the tendency to choose a default option more when the choice set is large. Our second application takes

[22]Note that the key difference between this example and the Perception-Based Reference Dependence example is how the reference $R(A)$ is determined. In the Perception example, $R(A)$ exhibits some systematic regularities across menus (is determined by the maximization of a linear oder). When the reference formation process is allowed to depend on the menu more generally, the Conditionally Decreasing property of F need not hold.

our model to two sets of experimental data on multi-attribute choice. We conclude our discussion by showing that the FLM can be used to correct for the role of attention when estimating price elasticity of demand.

6.1 Choice Overload and Status Quo Bias

There have been numerous experiments demonstrating the proclivity of decision makers to "opt out" of making a choice as the set of options becomes large. One instance of this is known as *choice overload*, a well-known behavioral phenomenon documented in both lab and field experiments. A subject experiencing choice overload is inclined to "choose not to choose" when presented with many alternatives. In one of the first papers on choice overload (Iyengar and Lepper, 2000), subjects were asked to choose among a large set of nearly identical alternatives (different flavored jams). Iyengar and Lepper found that a substantial fraction of subjects made no choice whatsoever (walked away from the display) and that the fraction of subjects walking away increased from 26% to 40% as the number of alternatives increased from 6 to 24. A similar pattern emerges in problems with a default or status quo alternative. For example, Dean (2009) finds that subjects are significantly more likely to stick with the status quo in large choice sets. Since the subjects in these two experiments violate regularity, their choice behavior cannot be explained any random utility model. We now demonstrate that FLM can generate this type of behavior.

Suppose that for menu $A_n = \{a_0, a_1, \ldots, a_n\}$, a_0 is an alternative that represents either the outside option (not making a choice) or a status quo alternative, and $a_1, \ldots, a_n$ are nearly identical alternatives. For simplicity, assume that $u(a_0) = u(a_1) = \ldots = u(a_n)$. We will now calculate the probability of choosing a_0 from A_n as a function of n.

In Luce's model, the probability $p(a_0, A_n) = \frac{1}{n+1}$ is always decreasing in n. In the FLM however this probability can be U-shaped, a property that cannot occur under any deterministic model generating choice overload. For simplicity, suppose that $F(A_n) = a_0$ and $\delta(A_n) = 0.024\,n^2 + 0.054\,n - 0.078$. Notice that $F(A_n) = a_0$ captures the idea of status quo bias. A direct calculation of the FLM choice probability gives,

$$p(a_0, A_n) = \frac{0.024\,n^2 + 0.054\,n + 0.922}{0.024\,n^2 + 1.054\,n + 0.922} = f(n).$$

It is simple to check that $f(n)$ is decreasing when $n < 6$ but increasing when $n \geq 6$. Moreover, we can match the observed frequencies exactly: $f(6) = 0.26$ and $f(24) = 0.40$. We can generate qualitatively similar behavior even when the outside option is strictly worse than all other alternatives.

6.2 Explaining Experimental Data

To demonstrate the explanatory power of the FLM, we discuss two experimental studies by Kivetz et al. (2004). In the first study, the authors ask three groups (for a total 463 subjects) to choose one of three portable PCs. The first group choose from the menu $A^1 = \{x, y, z\}$, the second group from

Alternatives/ Menus	A^1	A^2	A^3
x	0.06	—	—
y	0.50	0.18	—
z	0.44	0.51	0.24
t	—	0.31	0.47
s	—	—	0.29

Table 1: Choice frequencies presented in Empirical Application I of Kivetz et al. (2004)

Alternatives/Menus	B^1	B^2	FLM frequencies	
a	0.03	—	0.03	—
b	0.24	0.09	0.24	0.09
c	0.32	0.26	0.32	0.26
d	0.32	0.42	0.32	0.42
e	0.09	**0.13**	0.09	**0.118**
f	—	**0.09**	—	**0.102**

Table 2: Choice frequencies presented in Empirical Application II of Kivetz et al. (2004)

$A^2 = \{y, z, t\}$, and the third from $A^3 = \{z, t, s\}$. The data are summarized in Table 1. Alternatives are monotonically ordered by their speed and memory, e.g., s is the fastest but lowest memory PC while x is the slowest but highest memory PC.

Notice that Table 1 lends further support for the compromise effect. That is, the compromise alternatives (e.g., y in A^1 and z in A^2) are chosen relatively more frequently than extreme alternatives (e.g, x and z in A^1 and y and t in A^2). However, this experiment features varying compromise alternatives (i.e., y in A^1 to z in A^2 to t in A^3) and therefore it imposes additional challenges to existing models that explain the compromise effect.

From these three menus, we can observe at most two violations of IIA, since $A^1 \cap A^2 = \{y, z\}$ and $A^2 \cap A^3 = \{z, t\}$. With only two violations of IIA the FLM can perfectly replicate the frequencies in Table 1. Surprisingly, we can replicate Table 1 by the FLM with a constant δ. Suppose that $F(A^1) = \{y\}$, $F(A^2) = \{z\}$, and $F(A^3) = \{t\}$, which is consistent with the compromise effect.[23] The FLM with a constant δ predicts that $\frac{p(y,A^1)}{p(z,A^1)} / \frac{p(y,A^2)}{p(z,A^2)} = \frac{p(t,A^3)}{p(z,A^3)} / \frac{p(t,A^2)}{p(z,A^2)}$. It turns out that the observed frequencies are consistent with this prediction: $3.2197 = \frac{0.50}{0.44} / \frac{0.18}{0.51} \approx \frac{0.47}{0.24} / \frac{0.31}{0.51} = 3.2217$. In particular, when $u(x) = 1$, $u(y) = 4.64$, $u(z) = 7.33$, $u(t) = 8$, and $u(s) = 8.86$, the FLM with a constant weight $\delta = 0.795$ replicates frequencies in Table 1.

In the second study of Kivetz et al. (2004), the authors ask two groups (for a total 198 subjects) to choose one of five portable PCs. As shown in Table 2, the first group chose from the menu $B^1 = \{a, b, c, d, e\}$, and the second group chose from $B^2 = \{b, c, d, e, f\}$. As in the previous study, alternatives are monotonically ordered by their speed and memory: a is the fastest but with the lowest memory while f is the slowest but with the highest memory. Similar to the first study, Table 2 is consistent with the compromise effect. Moreover, it also imposes additional challenges

[23]By directly comparing relative probabilities (i.e., $r_{A^1}(y, z) = \frac{0.5}{0.44} > r_{A^2}(y, z) = \frac{0.18}{0.51}$ and $r_{A^2}(z, t) = \frac{0.51}{0.31} > r_{A^3}(z, t) = \frac{0.24}{0.47}$) we can easily conclude that $F(A^1) \neq z$, $F(A^2) \ni z$, and $F(A^3) \neq z$.

to existing models since each menu has multiple compromise alternatives (e.g., b and c in B^1 and c and d in B^2).

Since $B^1 \cap B^2 = \{b, c, d, e\}$, there are at most six possible violations of IIA. In this case, the FLM cannot replicate the frequencies in Table 2 exactly, as there are 6 violations of IIA. In fact, we can reject the FLM with just two menus. Recalling Proposition 2, in the FLM, $\frac{r_{B_1}(b,c)}{r_{B_2}(b,c)} > \frac{r_{B_1}(c,d)}{r_{B_2}(c,d)} > \frac{r_{B_1}(d,e)}{r_{B_2}(d,e)}$ implies $1 \geq \frac{r_{B_1}(d,e)}{r_{B_2}(d,e)}$. Yet by direct calculation we see that, $\frac{r_{B_1}(d,e)}{r_{B_2}(d,e)} \approx 1.1$, and hence the data is not matched exactly by FLM.[24]

However, the FLM can generate frequencies very similar to the observed frequencies. Consistent with the compromise effect, suppose that $F(B^1) = \{b, c\}$ and $F(B^2) = \{c, d, e\}$ (these focal sets can be partially identified from the data by looking at violations of IIA and their magnitudes). Then when $u(a) = 1$, $u(b) = 4.95$, $u(c) = 6.6$, $u(d) = 10.67$, $u(e) = 3$, and $u(f) = 5.61$, the FLM with $\delta(B^1) = \frac{8}{13}$ and $\delta(B^2) = \frac{7}{6}$ generates frequencies in the right-hand side of Table 2. The only difference is that in the experimental data IIA is violated at d, e (i.e., $\frac{0.32}{0.09} > \frac{0.42}{0.13}$), but the FLM generated data satisfies IIA at d, e (i.e., $\frac{0.32}{0.09} \approx \frac{0.42}{0.118}$).

6.3 Estimating Price Elasticity of Demand: An example

In this section we illustrate the importance of correcting for focus (or attention) when estimating utility parameters and elasticities, which are of great importance to the analysis of policy.[25] For example, suppose we want to estimate price elasticity of demand. When a firm offers a discount, demand for the good increases and this change is used to estimate elasticity. However, this discount may also affect how the decision maker focuses on goods. In particular, a discount may affect demand through two channels. First, demand increases due to the classical law of demand. Second, if the sale good becomes more focal than other goods, then demand may increase for non-instrumental reasons. If the effect of this change in focus is not accounted for, then the estimate of elasticity will be biased.[26]

For simplicity, assume each good $a = (x, p) \in X \subset \mathbb{R}^2_{++}$ has two attributes, where x represents the characteristics of a good and p is the price of a good. Suppose that a researcher has a logit demand model $p((x, p), A) = \frac{\exp(\alpha\, x - \beta\, p)}{\sum_{(x', p') \in A} \exp(\alpha\, x' - \beta\, p')}$ in mind and is interested in finding α, β.

Suppose there are n goods in the market and the researcher observes choice frequencies from two menus $A = \{a_1 = (x_1, p_1), a_2 = (x_2, p_2), \ldots, (x_n, p_n)\}$ and $B = \{a'_1 = (x_1, p'_1), a_2 = (x_2, p_2), \ldots, a_n = (x_n, p_n)\}$. Here $a'_1 \in B$ is a discounted version of a_1; i.e., $p_1 > p'_1$. It is reasonable to think that the discount for good 1 changes consumers' focus (e.g., focal sets in A and B). Below we illustrate that if the researcher does not account for changes in focus due to the discount, she may overestimate price elasticity of demand (or β). Intuitively, the demand for good 1 increases when price p_1 decreases to

[24] We take this as evidence in favor of the FLM for two reasons. First, the fact that it is not exact shows that the FLM can be falsified with only two menus. Second, the difference is so small in this case as to be insignificant. A switch by one subject (out of 100) would make the fit essentially exact.

[25] While the multinomial logit is used extensively in practice, it is well-known to provide "bizarre" elasticity estimates (Berry (1994) and Berry et al. (1995)).

[26] Interestingly, using data on the U.S. personal computer market Goeree (2008) shows that estimated demand curves are biased towards being too elastic if limited information (or attention) of consumers is ignored.

p_1'. However, the demand increase is not just due to the price decrease but also because of a change in focus. Therefore β will be overestimated.[27] To illustrate, assume that initially $F(A) = A$, all goods are equally focal, but after the price change, $F(B) = \{a_1'\}$. For simplicity, let us consider the FLM with a constant δ. Let $y_i^A \equiv \log\big(r_A(a_i, a_n)\big), y_1^B \equiv \log\big(r_B(a_1', a_n)\big),$ and $y_j^B \equiv \log\big(r_B(a_j, a_n)\big)$ for any $n-1 \geq i \geq 1$ and $n-1 \geq j \geq 2$.

Let us consider two simple strategies that the researcher can use to identify α, β. First, the research can simply exploit the price change, $p_1 \to p_1'$. Notice that for a logit model with a linear utility,

$$y_1^B - y_1^A = \big(\alpha(x_1 - x_n) - \beta(p_1' - p_n)\big) - \big(\alpha(x_1 - x_n) - \beta(p_1 - p_n)\big) = \beta(p_1 - p_1').$$

Therefore, $\hat\beta = \frac{y_1^B - y_1^A}{p_1 - p_1'}$ is a natural estimate for β. However, if a_1' becomes very salient because of the discount (i.e., $F(B) = \{a_1'\}$), the above overestimates β, i.e., $\hat\beta = \frac{y_1^B - y_1^A}{p_1 - p_1'} = \beta + \frac{\delta}{p_1 - p_1'}$.[28]

The second strategy is to run an OLS regression for the linear model $y = \alpha x - \beta p + \epsilon$.[29] In this case we obtain $\hat\beta = \beta + \epsilon^T q + \delta\, r$ where q and r are some functions of x and p (see the appendix for details). Just as in the first strategy, the researcher obtains a biased estimate. Indeed, this bias can be corrected by estimating the FLM as in section 7.

7 Estimation and Identifying Focal Sets

In sections 3 and 6.2, we directly identified focal sets from violations of IIA and their magnitudes. When focal sets are given it is simple to calculate the utilities of alternatives. However, large data sets require a systematic way to identify focal sets. In this section, we show that the FLM parameters (i.e., utilities, focal sets, and biases) can be estimated as easily as the Luce model. Moreover, we show the consistency of our estimators for linear utility models.

Suppose we have the following data: $\big\{A^l = \{a_0^l, a_1^l, \ldots, a_{m_l}^l\}, P^l = (p_0^l, \ldots, p_{m_l}^l)\big\}_{l=1}^L$ with $(p_0^l, \ldots, p_{m_l}^l) \in (0,1)^{m_l}$ and $\sum_{t=0}^{m_l} p_t^l = 1$ for each $l \in \{1, 2, \ldots, L\}$. Here L is the number of menus and p_k^l is the probability that alternative a_k^l is chosen from menu A^l. According to our model,

$$p_t^l = \frac{\exp\big(\bar u(a_t^l) + \bar\delta(A^l)\, \mathbb{1}\{a_t^l \in F(A^l)\}\big)}{\sum_{s=0}^{m_l} \exp\big(\bar u(a_s^l) + \bar\delta(A^l)\, \mathbb{1}\{a_s^l \in F(A^l)\}\big)}.^{30}$$

For simplicity, we assume that there is an alternative a_0 such that $a_0^l = a_0$ for all $l \leq L$. This assumption is satisfied in the two experimental studies of Kivetz et al. (2004) and is very natural

[27] We wish to remark that in principle a price change in good a might cause some other good, say a', to become focal. In this case there would be a change in demand for a' without a change in its own price. This should only further illustrate the importance of accounting for focus when estimating utility parameters.

[28] It may be reasonable to think that consumers' focus may not change when $p_1 - p_1' < \Delta$ for some $\Delta > 0$. Then the magnitude of the bias is 0 when $p_1 - p_1' < \Delta$ and is decreasing in $p_1 - p_1'$ when $p_1 - p_1' \geq \Delta$.

[29] Here $y = (y_1^A - y_n^A, \ldots, y_{n-1}^A - y_n^A, y_1^B - y_n^B, \ldots, y_{n-1}^B - y_n^B)$, $x = (x_1 - x_n, \ldots, x_{n-1} - x_n, x_1 - x_n, \ldots, x_{n-1} - x_n)$, and $p = (p_1 - p_n, \ldots, p_{n-1} - p_n, p_1' - p_n, \ldots, p_{n-1} - p_n)$.

[30] Notice that we obtain the FLM by setting $u(x) \equiv \exp(\bar u(x))$ and $\delta(A) \equiv \exp(\bar\delta(A)) - 1$.

when a_0 is the outside option (or walking away). The assumption also becomes redundant for linear utility models and can be relaxed easily. Under this assumption, we have

$$\log(p_t^l/p_0^l) = \bar{u}(a_t^l) - \bar{u}(a_0) + \bar{\delta}(A^l)\big(\mathbb{1}\{a_t^l \in F(A^l)\} - \mathbb{1}\{a_0 \in F(A^l)\}\big).$$

Let $y_t^l \equiv \log(\frac{p_t^l}{p_0^l})$, $v(a_t^l) = \bar{u}(a_t^l) - \bar{u}(a_0)$, $d_t^l = \big|\mathbb{1}\{a_t^l \in F(A^l)\} - \mathbb{1}\{a_0 \in F(A^l)\}\big|$, and $\delta_l \equiv \bar{\delta}(A^l)\,\mathrm{sign}\big(\frac{1}{2} - \mathbb{1}\{a_0 \in F(A^l)\}\big)$. Then, $\delta_l = -\bar{\delta}(A^l)$ and $d_t^l = 1$ iff $a_t^l \notin F(A^l)$ when $a_0 \in F(A^l)$, but $\delta^l = \bar{\delta}(A^l)$ and $d_t^l = 1$ iff $a_t^l \in F(A^l)$ when $a_0 \notin F(A^l)$.

Now notice that we have arrived at the following simple linear programming problem: for given $\{(y_1^l, \ldots, y_{m_l}^l)\}_{l=1}^L$, we need to find

$$\{(v(a_1^l), \ldots, v(a_{m_l}^l)), \delta_l, (d_1^l, \ldots, d_{m_l}^l)\}_{l=1}^L \in \prod_{l=1}^L \mathbb{R}^{m^l+1} \times \{0,1\}^{m^l} \text{ such that}$$

for any $l, l' \leq L$, $t \leq m_l$, and $t' \leq m_{l'}$,

$$y_t^l = v(a_t^l) + \delta_l\, d_t^l \text{ and } v(a_t^l) = v(a_{t'}^{l'}) \text{ whenever } a_t^l = a_{t'}^{l'} \qquad \text{(LPP)}.$$

As demonstrated in section 6.2, if the observed menus do not intersect or only have a few intersections, then the FLM can perfectly fit the observed frequencies.[31] If there are many violations of IIA (as in Table 2), then the data rejects the FLM. In this case, we look for the best fitting FLM by solving the following least squares problem.

$$\begin{aligned}
&(4) \qquad && \min \quad \textstyle\sum_{l=1}^L \sum_{t=1}^{m_l} \left(y_t^l - v(a_t^l) - \delta_l\, d_t^l \right)^2 && \text{(LSP)} \\
&(5) && \text{subject to} \quad \{(v(a_1^l), \ldots, v(a_{m_l}^l)), \delta_l, (d_1^l, \ldots, d_{m_l}^l)\}_{l=1}^L \in \textstyle\prod_{l=1}^L \mathbb{R}^{m^l+1} \times \{0,1\}^{m^l} && \\
&(6) && \text{and} \quad v(a_t^l) = v(a_{t'}^{l'}) \text{ whenever } a_t^l = a_{t'}^{l'}. &&
\end{aligned}$$

When attributes of alternatives are observed and a linear utility model is assumed, we can run a linear minimization problem that is very similar to OLS. For example, suppose that $a_t^l = (a_{1,t}^l, \ldots, a_{n,t}^l) \in \mathbb{R}^n$ for each l, t. Let $x_t^l = a_t^l - a_0$ and $v(a_t^l) = \sum_{i=1}^n \alpha_i\, x_{i,t}^l$. For linear utility models, constraint (6) is immediately satisfied. Therefore, we obtain the following linear minimization problem.

$$\begin{aligned}
&\min \quad \textstyle\sum_{l=1}^L \sum_{t=1}^{m_l} \left(y_t^l - \sum_{i=1}^n \alpha_i\, x_{i,t}^l - \delta_l\, d_t^l \right)^2 \qquad \text{(LMP)} \\
&\text{subject to} \quad (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n,\ (\delta_1, \ldots, \delta_L) \in \mathbb{R}^L, (d_1^l, \ldots, d_{m_l}^l) \in \{0,1\}^{m_l}.
\end{aligned}$$

With linear utility we do not need the observed menus to intersect to estimate parameters or reject

[31] In the online appendix, the maximum likelihood estimation is also considered. If there are different FLM parameters that can fit observed frequencies, then one can pick parameters that can maximize the likelihood function for FLM.

the FLM. Moreover, LMP provides consistent estimates for $\alpha, (\delta_l), (d_t^l)$. More formally,

Proposition 7. *Let $(\hat{\alpha}, (\hat{\delta}_l), (\hat{d}_t^l))$ be the solution to the LMP. Suppose that $\delta_l \neq 0$ for each l. Then $(\hat{\alpha}, (\hat{\delta}_l), (\hat{d}_t^l)) \to^P (\alpha, (\delta_l), (d_t^l))$ as the number of observations for each menu converges to infinity.*[32]

The assumption $\delta_l \neq 0$ means that there is at least one violation of IIA in A^l. The proof of Proposition 7 is in the online appendix. We conclude this section with two remarks.

1. Once we obtain $\{\hat{\delta}_l, (\hat{d}_1^l, \ldots, \hat{d}_{m^l}^l)\}_{l=1}^L$ from the data, we can identify focal sets. In particular, if $\hat{\delta}_l > 0$, then $a_0 \notin F(A^l)$ and $a_t^l \in F(A^l)$ whenever $\hat{d}_t^l = 1$. If $\delta_l < 0$, then $a_0 \in F(A^l)$ and $a_t^l \in F(A^l)$ whenever $\hat{d}_t^l = 1$. Lastly, if $\hat{\delta}_l = 0$, then $F(A^l) = A^l$.

2. We can test the nested logit version of the FLM (or the FLM with menu-independent categorization) by (4)-(6) with the following additional constraint: if $a_t^l = a_{t'}^k$ and $a_t^{l'} = a_{t'}^{k'}$, then $d_t^l = d_t^{l'}$ whenever $d_{t'}^k = d_{t'}^{k'}$.

8 Related Literature

As we have demonstrated in the previous sections, the FLM can be consistent with the compromise effect, choice overload, and status quo bias.[33] In addition, the FLM is also able to capture more nuanced behavioral regularities associated with these phenomena, such as increasing status quo bias and a shifting, multiple compromises. However, the main focus of the paper is to provide a simple axiomatic model that behaviorally distinguishes focal alternatives from the bias towards these alternatives in a way that is easily applicable to data. We will therefore restrict our attention to the axiomatic literature that is directly related to Luce's model and consideration set models.

The benchmark economic model of rational behavior for random choice is the random utility model (Falmagne (1978), McFadden and Richter (1990), and Barbera and Pattanaik (1986)). In the random utility model, the utility of a given alternative x is $u(x)+\epsilon_x$, where $u(x)$ is the deterministic part and ϵ_x is the random part, and a decision maker chooses the alternative with the highest utility.

Random utility models always satisfy regularity while the FLM allows for its violation (e.g., choice overload). Therefore, the FLM is not a special case of the random utility model. However, when the magnitude of the bias is small enough, the FLM is a special case of the random utility model. Moreover, the FLM generalizes the nested logit with two nests, which is a random utility model under parametric assumptions.

Two important generalizations of the Luce model are the Elimination-By-Aspects model of Tversky (1972) and the Attribute Rule of Gul et al. (2014). Both models feature a procedure akin

[32] Suppose we observe $c_{a_i^l}^k \in \{0,1\}$ for each $l \leq L$, $k \leq N^l$ such that $\sum_{i=0}^{m_l} c_{a_i^l}^k = 1$. Here $c_{a_i^l}^k = 1$ means that consumer k (or a consumer visits the menu the k-th time) chooses a_i^l from menu $A^l = \{a_0^l, a_l^1, \ldots, a_{m_l}^l\}$. Here, N^l is the number of observations from menu A^l. Therefore, $p_i^l = \frac{\sum_{k=1}^{N^l} c_{a_i^l}^k}{N^l}$. Proposition 7 requires N^l to become large enough for each l.

[33] Some recent bounded rationality models (of deterministic or random choice) can explain all or subsets of these behavioral phenomena. For example, see Bordalo et al. (2013), De Clippel and Eliaz (2012), Echenique et al. (2018), Frick (2016), Fudenberg et al. (2015), Masatlioglu et al. (2012), Natenzon (2019), Ok et al. (2014), and Tserenjigmid (2019).

to sequential Luce rules and deal with the duplicates problem. Further, objective attributes play a key role in the choice procedure. Both models are special cases of random utility and thus cannot explain much of the literature on choice overload or the compromise effect like the FLM. Further, the FLM is motivated by menu-effects, rather than how similar alternatives influence choice.

Our model can be interpreted as a stochastic version of consideration set models such as Manzini and Mariotti (2007) and Masatlioglu et al. (2012). However, the new property of Conditionally Decreasing that we introduce is more general than what is used in those papers. Further, in these papers and other deterministic choice models of limited consideration the agent must choose from the consideration set. In contrast, we identify both the focal set and the magnitude of the bias, F and δ.

Among the stochastic models of consideration sets, such as Manzini and Mariotti (2014), the main difference between existing models and ours is the source of randomness. In our model, the main source of randomness is utility shocks and the consideration set for a given menu is fixed.[34] In Manzini and Mariotti (2014) tastes are fixed and thus the source of randomness is random attention. In particular, every possible subset is the consideration set with positive probability. Brady and Rehbeck (2016), Aguiar (2017), Zhang (2016), Kovach (2016), and Kovach and Ülkü (2018) provide generalizations of (or closely related models to) Manzini and Mariotti (2014) in which the probability that a given alternative is in the consideration set is menu-dependent. The primary distinction between those models and ours is that we identify a single focal set, while they provide a distribution over possible consideration sets in which each possible subset has positive probability.

Echenique et al. (2018) provides a generalization of Luce's model, called the Perception-Adjusted Luce Model (PALM), in which a decision maker looks at alternatives according to an exogenous order, and chooses each alternative with the Luce probability. Therefore, similar to Manzini and Mariotti (2014), the decision maker may consider a subset of the original menu, but the subset is random.

Marley (1991) proposes a stochastic choice model that has a nontrivial intersection with the FLM. He introduces the following stochastic choice model that captures a decision-making procedure which involves a sequence of binary comparisons:

$$p(x, A) = \frac{\prod_{y \in A \setminus \{x\}} (\pi(x, y))^{\frac{1}{\theta(A)}}}{\sum_{z \in A} \prod_{t \in A \setminus \{z\}} (\pi(z, t))^{\frac{1}{\theta(A)}}}.$$

Ravid and Steverson (2018) provide an elegant axiomatic characterization for the case $\theta = 1$. Suppose $\theta = 1$ and X_1, X_2 is a partition of X and $\pi(x, y) = \frac{1}{u(y)}$ and $\pi(x, z) = \frac{1}{u(z)\,\lambda(z)}$ for any $x, y \in X_i$ and $z \in X_j$. Then we obtain a FLM with (u, F, δ) such that $F(A) = A \cap X_i$ and $\delta(A) = \frac{\prod_{a \in A \cap X_i} \lambda(a)}{\prod_{b \in A \cap X_j} \lambda(b)} - 1$ when $\prod_{a \in A \cap X_i} \lambda(a) > \prod_{b \in A \cap X_j} \lambda(b)$. Note that when $\lambda \geq 1$, F

[34] There is an interpretation in which our DM has stochastic consideration, but there are only two possible realizations, A or $F(A)$.

is conditionally decreasing and δ is conditionally increasing (otherwise they are not conditionally monotonic).

References

Aguiar, V. (2017): "Random Categorization and Bounded Rationality," *Economics Letters*.

Apesteguia, J. and M. A. Ballester (2013): "Choice by sequential procedures," *Games and Economic Behavior*, 77, 90 – 99.

Armel, K. C., A. Beaumel, and A. Rangel (2008): "Biasing simple choices by manipulating relative visual attention," *Judgment and Decision Making*, 3, 396–403.

Bagwell, K. (2007): "The Economic Analysis of Advertising," in *Handbook of Industrial Organization*, ed. by M. Armstrong and R. Porter, Elsevier, vol. 3.

Barbera, S. and P. K. Pattanaik (1986): "Falmagne and the rationalizability of stochastic choices in terms of random orderings," *Econometrica: Journal of the Econometric Society*, 707–715.

Berry, S. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *The Rand Journal of Economics*, 25, 242–262.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Proces in Market Equilibrium," *Econometrica*, 63, 841–890.

Bordalo, P., N. Gennaioli, and A. Shleifer (2013): "Salience and Consumer choice," *Journal of Political Economy*, 121, 803–843.

Brady, R. L. and J. Rehbeck (2016): "Menu-Dependent Stochastic Feasibility," *Econometrica*, 84, 1203–1223.

Cherepanov, V., T. Feddersen, and A. Sandroni (2013): "Rationalization," *Theoretical Economics*, 8, 775–800.

Chernev, A. (2004): "Extremeness Aversion and Attribute-Balance Effects in Choice," *Journal of Consumer Research*, 31, 249–263.

De Clippel, G. and K. Eliaz (2012): "Reason-based choice: A bargaining rationale for the attraction and compromise effects," *Theoretical Economics*, 7, 125–162.

Dean, M. (2009): "Status Quo Bias in Large and Small Choice Sets." Mimeo New York University.

Echenique, F., K. Saito, and G. Tserenjigmid (2018): "The perception-adjusted Luce model," *Mathematical Social Sciences*, 93, 67–76.

Eliaz, K., M. Richter, and A. Rubinstein (2011): "Choosing the two finalists," *Economic Theory*, 46, 211–219.

Falmagne, J.-C. (1978): "A representation theorem for finite random scale systems," *Journal of Mathematical Psychology*, 18, 52–72.

Frick, M. (2016): "Monotone threshold representations," *Theoretical Economics*, 11, 757–772.

Fudenberg, D., R. Iijima, and T. Strzalecki (2015): "Stochastic choice and revealed perturbed utility," *Econometrica*, 83, 2371–2409.

Gaundry, M. J. and M. G. Dagenais (1979): "The dogit model," *Transportation Research Part B: Methodological*, 13, 105–111.

Goeree, M. S. (2008): "Limited information and advertising in the US personal computer industry," *Econometrica*, 76, 1017–1074.

Gossner, O., J. Steiner, and C. Stewart (2019): "Attention Please!" *Working Paper*.

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

Herne, K. (1997): "Decoy alternatives in policy choices: Asymmetric domination and compromise effects," *European Journal of Political Economy*, 13, 575–589.

Iyengar, S. S. and M. R. Lepper (2000): "When choice is demotivating: Can one desire too much of a good thing?" *Journal of personality and social psychology*, 79, 995.

Kibris, O., Y. Masatlioglu, and E. Suleymanov (2018): "A Theory of Reference Formation," Working Paper.

Kivetz, R., O. Netzer, and V. Srinivasan (2004): "Alternative models for capturing the compromise effect," *Journal of marketing research*, 41, 237–257.

Kovach, M. (2016): "Thinking Inside the Box: Status Quo Bias and Stochastic Consideration," Working paper.

Kovach, M. and G. Tserenjigmid (2018): "Bounded Rationality in Games: The Focal Quantal Response Equilibrium," *Working Paper*.

Kovach, M. and L. Ülkü (2018): "Satisficing with a Variable Threshold," *Working Paper*.

Lleras, J. S., Y. Masatlioglu, D. Nakajima, and E. Y. Ozbay (2017): "When more is less: Limited consideration," *Journal of Economic Theory*, 170, 70–85.

Luce, R. D. (1959): *Individual choice behavior: A theoretical analysis*, John Wiley and sons.

Manzini, P. and M. Mariotti (2007): "Sequentially rationalizable choice," *The American Economic Review*, 1824–1839.

——— (2014): "Stochastic choice and consideration sets," *Econometrica*, 82, 1153–1176.

Mariotti, M. and P. Manzini (2010): "Revealed Preferences and Boundedly Rational Choice Procedures: an Experiment," *IZA Discussion Paper No. 2341*.

Marley, A. (1991): "Context dependent probabilistic choice models based on measures of binary advantage," *Mathematical Social Sciences*, 21, 201–231.

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

Masatlioglu, Y., D. Nakajima, and E. Ozdenoren (2011): "Revealed willpower," Tech. rep., mimeo.

Masatlioglu, Y. and E. A. Ok (2013): "A canonical model of choice with initial endowments," *Review of Economic Studies*, 81, 851–883.

McFadden, D. and M. K. Richter (1990): "Stochastic rationality and revealed stochastic preference," *Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz, Westview Press: Boulder, CO*, 161–186.

McKelvey, R. D. and T. R. Palfrey (1995): "Quantal response equilibria for normal form games," *Games and economic behavior*, 10, 6–38.

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

Ok, E. A., P. Ortoleva, and G. Riella (2014): "Revealed (p) reference theory," *The American Economic Review.*

Payne, J. W. (1976): "Task Complexity and Contingent Processing in Decision Making: An informaiton Seach and Protocol Analysis," *Organizaitonal Behavior and Human Performance*, 366–387.

Payne, J. W., J. R. Bettman, and E. J. johnson (1993): *The Adaptive Decision Maker*, Cambridge University Press.

Ravid, D. and K. Steverson (2018): "Focus, Then Compare," *Working Paper.*

Reutskaja, E., R. Nagel, C. F. Camerer, and A. Rangel (2011): "Search Dynamics in Consumer Choice under Time Pressure: An Eye-Tracking Study," *American Economic Review*, 101, 900–926.

Rubinstein, A. (2016): "A typology of players: Between instinctive and contemplative," *The Quarterly Journal of Economics*, 131, 859–890.

Schelling, T. (1960): *The Strategy of Conflict*, Harvard University Press.

Simonson, I. (1989): "Choice based on reasons: The case of attraction and compromise effects," *Journal of consumer research*, 158–174.

Simonson, I. and A. Tversky (1992): "Choice in context: tradeoff contrast and extremeness aversion." *Journal of marketing research.*

Tserenjigmid, G. (2019): "Choosing with the worst in mind: A reference-dependent model," *Journal of Economic Behavior & Organization*, 157, 631–652.

Tversky, A. (1972): "Elimination by Aspects: A Theory of Choice," *Psychological review.*

Zhang, J. (2016): "Stochastic Choice with Categorization," *Working Paper.*

A Proofs

A.1 Proof of Proposition 1

Take any A, B and $a, b \in A \cap B$. Then $\frac{r_A(a,b)}{r_B(a,b)} = \frac{1+\delta(A)\mathbb{1}\{a\in F(A)\}}{1+\delta(A)\mathbb{1}\{b\in F(A)\}} / \frac{1+\delta(B)\mathbb{1}\{a\in F(B)\}}{1+\delta(B)\mathbb{1}\{b\in F(B)\}} > 1$ iff $\frac{1+\delta(A)\mathbb{1}\{a\in F(A)\}}{1+\delta(A)\mathbb{1}\{b\in F(A)\}} > \frac{1+\delta(B)\mathbb{1}\{a\in F(B)\}}{1+\delta(B)\mathbb{1}\{b\in F(B)\}}$. By way of contradiction, suppose that $a \notin F(A)$ and $b \notin F(B)$. Then $\frac{r_A(a,b)}{r_B(a,b)} > 1$ iff $\frac{1}{1+\delta(A)\mathbb{1}\{b\in F(A)\}} > 1 + \delta(B)\mathbb{1}\{a \in F(B)\}$, a contradiction.

A.2 Proof of Proposition 2

Take any A, B and $a, b \in A \cap B$.

Proposition 2.i). $\frac{r_A(a,b)}{r(a,b)} = \frac{1+\delta(A)\mathbb{1}\{a\in F(A)\}}{1+\delta(A)\mathbb{1}\{b\in F(A)\}} > 1$ iff $\delta(A)\mathbb{1}\{a \in F(A)\} > 0$ and $\delta(A)\mathbb{1}\{b \in F(A)\} = 0$; i.e., $a \in F(A)$ and $b \notin F(A)$.

Proposition 2.ii). By i), $r_A(b,c) > r(b,c)$ implies that $b \in F(A)$. Then $r_A(a,b) \geq r_B(a,b)$ iff

$$\frac{1+\delta(A)\mathbb{1}\{a \in F(A)\}}{1+\delta(A)} \geq \frac{1+\delta(B)\mathbb{1}\{a \in F(B)\}}{1+\delta(B)\mathbb{1}\{b \in F(B)\}}.$$

By way of contradiction, suppose $a \notin F(A)$ and $a \in F(B)$. Then $r_A(a,b) \geq r_B(a,b)$ iff $\frac{1}{1+\delta(A)} \geq \frac{1+\delta(B)}{1+\delta(B)\mathbb{1}\{b\in F(B)\}}$, a contradiction.

Proposition 2.iii). By i), $r_A(b,c) > r(b,c)$ implies that $c \notin F(A)$. Then $r_A(a,c) \leq r_B(a,c)$ iff

$$1+\delta(A)\mathbb{1}\{a \in F(A)\} \leq \frac{1+\delta(B)\mathbb{1}\{a \in F(B)\}}{1+\delta(B)\mathbb{1}\{c \in F(B)\}}.$$

By way of contradiction, suppose $a \in F(A)$ and $a \notin F(B)$. Then $r_A(a,c) \leq r_B(a,c)$ iff $1+\delta(A) \leq \frac{1}{1+\delta(B)\mathbb{1}\{c\in F(B)\}}$, a contradiction.

A.3 Proof of Proposition 3

Take any A, B and $a, b \in A \cap B$.

Proposition 3.i). Suppose $r_A(a,b) > r_B(a,b)$ and $a \sim_B b$. By the definition, $a \sim_B b$ implies that there is $C \in \mathscr{A}$ such that $r_B(a,b) = r_C(a,b)$. Equivalently, $\frac{1+\delta(B)\mathbb{1}\{a\in F(B)\}}{1+\delta(B)\mathbb{1}\{b\in F(B)\}} = \frac{1+\delta(C)\mathbb{1}\{a\in F(C)\}}{1+\delta(C)\mathbb{1}\{b\in F(C)\}}$. If $\frac{1+\delta(B)\mathbb{1}\{a\in F(B)\}}{1+\delta(B)\mathbb{1}\{b\in F(B)\}} = \frac{1+\delta(C)\mathbb{1}\{a\in F(C)\}}{1+\delta(C)\mathbb{1}\{b\in F(C)\}} > 1$, then we need to have $1+\delta(B) = 1+\delta(C)$, which contradicts the assumption on δ. Similarly, we cannot have $\frac{1+\delta(B)\mathbb{1}\{a\in F(B)\}}{1+\delta(B)\mathbb{1}\{b\in F(B)\}} = \frac{1+\delta(C)\mathbb{1}\{a\in F(C)\}}{1+\delta(C)\mathbb{1}\{b\in F(C)\}} < 1$. Hence, $\frac{1+\delta(B)\mathbb{1}\{a\in F(B)\}}{1+\delta(B)\mathbb{1}\{b\in F(B)\}} = \frac{1+\delta(C)\mathbb{1}\{a\in F(C)\}}{1+\delta(C)\mathbb{1}\{b\in F(C)\}} = 1$; i.e., $a, b \in F(B)$. Then $\frac{r_A(a,b)}{r_B(a,b)} = \frac{1+\delta(A)\mathbb{1}\{a\in F(A)\}}{1+\delta(A)\mathbb{1}\{b\in F(A)\}} > 1$ iff $\delta(A)\mathbb{1}\{a \in F(A)\} > 0$ and $\delta(A)\mathbb{1}\{b \in F(A)\} = 0$; i.e., $a \in F(A)$ and $b \notin F(A)$.

Proposition 3.ii). Suppose there are sequences $\{a_i\}_{i=1}^n$ and $\{A_i\}_{i=1}^{n-1}$ such that $a = a_1, b = a_n$, and $a_i \sim_{A_i} a_{i+1}$ for each $i < n$. By the proof of Proposition 3.i), $a_i \sim_{A_i} a_{i+1}$ implies $a_i, a_{i+1} \in F(A_i)$. Hence, $\prod_{i=1}^{n-1} r_{A_i}(a_i, a_{i+1}) = \prod_{i=1}^{n-1} \frac{u(a_i)}{u(a_{i+1})} = \frac{u(a_1)}{u(a_n)} = \frac{u(a)}{u(b)}$. Therefore, $r_A(a,b) > \prod_{i=1}^{n-1} r_{A_i}(a_i, a_{i+1})$ iff $\frac{1+\delta(A)\mathbb{1}\{a\in F(A)\}}{1+\delta(A)\mathbb{1}\{b\in F(A)\}} > 1$ iff $\delta(A)\mathbb{1}\{a \in F(A)\} > 0$ and $\delta(A)\mathbb{1}\{b \in F(A)\} = 0$; i.e., $a \in F(A)$ and $b \notin F(A)$.

A.4 Proof of Theorem 1

Since the necessity part is straightforward, we only prove the sufficiency part of Theorem 1. We prove by three steps.

Step 1. Let us fix a some alternative $a^* \in X$. We then construct $u : X \to \mathbb{R}_{++}$ in the following way: $u(a) \equiv r(a, a^*)$ for any $a \in X$. Note that for any $a, b \in X \setminus a^*$, $r(a,b) = \frac{u(a)}{u(b)}$ since Luce's Product

Rule implies that $r(a,b) = r(a,a^*) \cdot r(a^*,b) = u(a) \cdot \frac{1}{u(b)}$. Since $\frac{p(a,ab)}{p(b,ab)} = \frac{u(a)}{u(b)}$ and $p(a,ab)+p(b,ab) = 1$, we have $p(a,ab) = \frac{u(a)}{u(a)+u(b)}$.

Step 2. Take any $A \in \mathscr{A}$ with $|A| \geq 3$. We shall derive the BFLM representation for A.

Fact 1. For any $a, a', b, b' \in A$, if $r_A(a', a) = r(a', a)$ and $r_A(b', b) = r(b', b)$, then $r_A(a, b) > r(a, b)$ iff $r_A(a', b') > r(a', b')$.

Proof of Fact 1: Suppose $r_A(a', a) = r(a', a)$ and $r_A(b', b) = r(b', b)$. By Luce's Product Rule,

$$r_A(a', b') = r_A(a', a) \cdot r_A(a, b) \cdot r_A(b, b') > r(a', a) \cdot r(a, b) \cdot r(b, b') = r(a', b') \text{ iff } r_A(a, b) > r(a, b).$$

Fact 2. For any $A \in \mathscr{A}$ and $a, b \in A$, if $r_A(a, b) > r(a, b)$, then either $r_A(a, c) = r(a, c)$ or $r_A(b, c) = r(b, c)$ for any $c \in A$.

Proof of Fact 2: By way of contradiction, assume $r_A(a, c) \neq r(a, c)$ and $r_A(b, c) \neq r(b, c)$. Consider three cases.

Case 1. $r_A(a, c) > r(a, c)$. By Focal IIA, $r_A(a, b) > r(a, b)$ and $r_A(a, c) > r(a, c)$ imply $r_A(b, c) = r(b, c)$, a contradiction.

Case 2. $r_A(c, a) > r(c, a)$ and $r_A(c, b) > r(c, b)$. By Focal IIA, $r_A(c, a) > r(c, a)$ and $r_A(c, b) > r(c, b)$ imply $r_A(a, b) = r(a, b)$, a contradiction.

Case 3. $r_A(c, a) > r(c, a)$ and $r_A(b, c) > r(b, c)$. Then $r_A(a, b) > r(a, b)$, $r_A(b, c) > r(b, c)$, and $r_A(c, a) > r(a, c)$ imply $1 = r_A(a, b) \cdot r_A(b, c) \cdot r_A(c, a) > r(a, b) \cdot r(b, c) \cdot r(c, a)$, a contradiction to Luce's Product Rule.

Fact 3. There is $F(A) \subseteq A$ such that for any $a, a' \in F(A)$ and $b, b' \in N(A) = A \setminus F(A)$,

$$r_A(a, a') = r(a, a'),\ r_A(b, b') = r(b, b'),\ \text{and } r_A(a, b) > r(a, b).$$

Proof of Fact 3: When IIA is not violated in A, we set $F(A) = A$. Since IIA is satisfied and $r_A(a, a') = r(a, a') = \frac{u(a)}{u(a')}$, we also have $p(a, A) = \frac{u(a)}{\sum_{b \in A} u(b)}$ by Step 1. Suppose now there is a, b such that $r_A(a, b) > r(a, b)$. By Fact 2, $r_A(a, c) = r(a, c)$ or $r_A(b, c) = r(b, c)$ for any $c \in A$. Let $F(A) = \{c \in A : r_A(a, c) = r(a, c)\}$ and $N(A) = \{c \in A : r_A(b, c) = r(b, c)\}$. Note that $N(A) = A \setminus F(A)$. Moreover, by Fact 1, $r_A(a', b') > r(a', b')$ for any $a' \in F(A)$ and $b' \in N(A)$.

Step 3. Suppose that $|F(A)|, |N(A)| \geq 1$. Let $F(A) = \{a_1, \ldots, a_n\}$ and $N(A) = \{b_1, \ldots, b_m\}$.

By the construction of u, $r_A(a_i, a_1) = r(a_i, a_1) = \frac{u(a_i)}{u(a_1)}$. Therefore, $p(a_i, A) = \frac{u(a_i)}{u(a_1)} \times p(a_1, A)$. Similarly, $p(b_j, A) = \frac{u(b_j)}{u(b_1)} \times p(b_1, A)$. Then we have

$$1 = \sum_{i=1}^{n} p(a_i, A) + \sum_{j=1}^{m} p(b_j, A) = \frac{\sum_{i=1}^{n} u(a_i)}{u(a_1)} \cdot p(a_1, A) + \frac{\sum_{j=1}^{m} u(b_j)}{u(b_1)} \cdot p(b_1, A). \tag{7}$$

Note that for any $i \leq n$ and $j \leq m$, $\frac{r_A(a_1,b_1)}{r(a_1,b_1)} = \frac{r_A(a_i,b_j)}{r(a_i,b_j)}$ since $\frac{r_A(a_1,b_1)}{r_A(a_i,b_j)} = \frac{r_A(a_1,a_i)}{r_A(b_1,b_j)} = \frac{r(a_1,a_i)}{r(b_1,b_j)} = \frac{r(a_1,b_1)}{r(a_i,b_j)}$ by Step 1. Therefore, let $1 + \delta(A) \equiv \frac{r_A(a_1,b_1)}{r(a_1,b_1)}$. Finally, we have

$$
\begin{aligned}
p(b_1, A) = \frac{p(b_1, A)}{1} &= \frac{p(b_1, A)}{\frac{\sum_{i=1}^n u(a_i)}{u(a_1)} \cdot p(a_1, A) + \frac{\sum_{j=1}^m u(b_j)}{u(b_1)} \cdot p(b_1, A)} \text{ , by (7),} \\
&= \frac{u(b_1)}{\big(\sum_{i=1}^n u(a_i)\big) \cdot \frac{u(b_1)}{u(a_1)} \cdot \frac{p(a_1,A)}{p(b_1,A)} + \sum_{j=1}^m u(b_j)} \\
&= \frac{u(b_1)}{\big(\sum_{i=1}^n u(a_i)\big) \cdot \big(1 + \delta(A)\big) + \sum_{j=1}^m u(b_j)} \text{ , by the definition of } \delta(A).
\end{aligned}
$$

Since $1 = \sum_{i=1}^n p(a_i, A) + \sum_{j=1}^m p(b_j, A)$, we also have $p(a_1, A) = \frac{u(a_1)\big(1+\delta(A)\big)}{\big(\sum_{i=1}^n u(a_i)\big) \cdot \big(1+\delta(A)\big) + \sum_{j=1}^m u(b_j)}$.

Uniqueness (Proposition 4). Since the BFLM coincides with Luce's model on binary menus, the uniqueness of u is straightforward. The uniqueness of F follows form Proposition 1. Lastly, $\frac{r_A(a,b)}{r(a,b)} = 1 + \delta(A)$ since for any A and $a, b \in A$ with $r_A(a,b) > r(a,b)$, $\delta(A)$ is unique when IIA is violated in A.

A.5 Proof of Theorem 2

We only prove the sufficiency part of Theorem 2. We first prove that Increasing Focal Impact and Luce's Product Rule implies Focal IIA. Take any $A \in \mathscr{A}$ and $a, b, c \in A$ such that $r_A(a,b) > r(a,b)$ and $r_A(a,c) > r(a,c)$. By Increasing Focal Impact, $\frac{r_A(a,b)}{r(a,b)} = \frac{r_A(a,c)}{r(a,c)}$. By Luce's Product Rule, $r(b,c) = \frac{r(a,c)}{r(a,b)} = \frac{r_A(a,c)}{r_A(a,b)} = r_A(b,c)$.

Since Focal IIA and Luce's Product Rule are satisfied, by Theorem 1, p is a BFLM with some (u, F, δ). Take two menus $A, B \in \mathscr{A}$ such that IIA is violated at each of them and $B \subset A$. There are $a, b \in A$ and $a', b' \in B$ such that $r_A(a,b) > r(a,b)$ and $r_B(a', b) > r(a', b')$. Then by Increasing Focal Impact, we have $1 + \delta(A) = \frac{r_A(a,b)}{r(a,b)} \geq \frac{r_B(a',b')}{r(a',b')} = 1 + \delta(B)$.

The uniqueness part of Theorem 2 is implied by Proposition 4.

A.6 Proof of Theorem 3

Sufficiency. Suppose p satisfies Generalized Focal IIA and Generalized Product Rule and it is focally rich. By Focal Richness, there is $x^* \in X$ such that for any $y \in X$, there is A such that $x^* \sim_A y$. For each y, let us fix a menu $A(y)$ such that $x^* \sim_{A(y)} y$. Then let $u(y) \equiv r_{A(y)}(y, x^*)$ for each $y \in X$.

Step 1. For any $A \in \mathscr{A}$ and $a, b \in A$ with $a \sim_A b$, $r_A(a,b) = \frac{u(a)}{u(b)}$.

By Generalized Product Rule $a \sim_A b$, $b \sim_{A(b)} x^*$, and $x^* \sim_{A(a)} a$ imply $r_A(a,b) r_{A(b)}(b, x^*) r_{A(a)}(x^*, a) = 1$. Hence, by the definition of u, $r_A(a,b) = \frac{r_{A(x)}(a,x^*)}{r_{A(b)}(b,x^*)} = \frac{u(a)}{u(b)}$.

Step 2. Take any $A \in \mathscr{A}$. By Generalized Focal IIA, there is a partition A_1, A_2 of A such that for any $a, b \in A_i$, $a \sim_A b$.

When $A_j = \emptyset$ for some $j \in \{1,2\}$, then we obtain $a \sim_A b$ for any a, b. By Step 1, $r_A(a,b) = \frac{u(a)}{u(b)}$ for any $a, b \in A$. Therefore, $p(a,A) = \frac{u(a)}{\sum_{b\in A} u(b)}$. In this case, let $F(A) \equiv A$.

Step 3: Suppose that $|A_1|, |A_2| \geq 1$. Let $A_1 = \{a_1, \ldots, a_n\}$ and $A_2 = \{b_1, \ldots, b_m\}$.

By Step 1, $r_A(a_i, a_1) = \frac{u(a_i)}{u(a_1)}$ for any $a_i \in A_1$. Therefore, $p(a_i, A) = \frac{u(a_i)}{u(a_1)} \times p(a_1, A)$. Similarly, $p(b_j, A) = \frac{u(b_j)}{u(b_1)} \times p(b_1, A)$ for any $b_j \in A_2$. Then we have

$$1 = \sum_{i=1}^{n} p(a_i, A) + \sum_{j=1}^{m} p(b_j, A) = \frac{\sum_{i=1}^{n} u(a_i)}{u(a_1)} \cdot p(a_1, A) + \frac{\sum_{j=1}^{m} u(b_j)}{u(b_1)} \cdot p(b_1, A). \tag{8}$$

Note that for any $i \leq n$ and $j \leq m$, $\frac{r_A(a_1,b_1)}{\frac{u(a_1)}{u(b_1)}} = \frac{r_A(a_i,b_j)}{\frac{u(a_i)}{u(b_j)}}$ since $\frac{r_A(a_1,b_1)}{r_A(a_i,b_j)} = \frac{r_A(a_1,a_i)}{r_A(b_1,b_j)} = \frac{\frac{u(a_1)}{u(a_i)}}{\frac{u(b_1)}{u(b_j)}} = \frac{\frac{u(a_1)}{u(b_1)}}{\frac{u(a_i)}{u(b_j)}}$ by Step 1. Without loss of generality, let $r_A(a_1, b_1) > \frac{u(a_1)}{u(b_1)}$.[35] Then let $1+\delta(A) \equiv \frac{r_A(a_1,b_1)}{\frac{u(a_1)}{u(b_1)}}$. Finally, we have

$$\begin{aligned}
p(b_1, A) = \frac{p(b_1, A)}{1} &= \frac{p(b_1, A)}{\frac{\sum_{i=1}^{n} u(a_i)}{u(a_1)} \cdot p(a_1, A) + \frac{\sum_{j=1}^{m} u(b_j)}{u(b_1)} \cdot p(b_1, A)}, \text{ by (8)}, \\
&= \frac{u(b_1)}{\left(\sum_{i=1}^{n} u(a_i)\right) \cdot \frac{u(b_1)}{u(a_1)} \cdot \frac{p(a_1,A)}{p(b_1,A)} + \sum_{j=1}^{m} u(b_j)} \\
&= \frac{u(b_1)}{\left(\sum_{i=1}^{n} u(a_i)\right) \cdot \big(1+\delta(A)\big) + \sum_{j=1}^{m} u(b_j)}, \text{ by the definition of } \delta(A).
\end{aligned}$$

Since $1 = \sum_{i=1}^{n} p(a_i, A) + \sum_{j=1}^{m} p(b_j, A)$, we also have $p(a_1, A) = \frac{u(a_1)\big(1+\delta(A)\big)}{\left(\sum_{i=1}^{n} u(a_i)\right) \cdot \big(1+\delta(A)\big) + \sum_{j=1}^{m} u(b_j)}$. In this case, let $F(A) \equiv \{a_1, \ldots, a_n\}$.

Step 4. $\delta(A) \neq \delta(B)$ for any $A, B \in \mathscr{A}$ with $F(A) \cap F(B) \neq \emptyset$ and $A \cap B \setminus F(A) \setminus F(B) \neq \emptyset$.

Let $x \in F(A) \cap F(B)$ and $y \in A \cap B \setminus F(A) \setminus F(B)$. Then $r_A(x,y) = \frac{u(x)}{u(y)}(1+\delta(A))$ and $r_B(x,y) = \frac{u(x)}{u(y)}(1+\delta(B))$. By way of contradiction, suppose $\delta(A) = \delta(B)$. Then $r_A(x,y) = r_B(x,y)$; i.e., $x \sim_A y$. However, by Step 1, $r_A(a,b) = \frac{u(a)}{u(b)} = \frac{u(a)}{u(b)}(1+\delta(A))$, which contradicts $\delta(A) > 0$.

Necessity. Suppose p is an FLM with (F, δ) such that $\delta(A) \neq \delta(B)$ for any $A, B \in \mathscr{A}$ with $F(A) \cap F(B) \neq \emptyset$ and $A \cap B \setminus F(A) \setminus F(B) \neq \emptyset$. Let us first prove the following useful fact.

Fact 4. $a \sim_A b$ if and only if either $a, b \in F(A)$ or $a, b \notin F(A)$.

Proof of Fact 4. First, suppose $a, b \in F(A)$ or $a, b \notin F(A)$. By way of contradiction, suppose $a \not\sim_A b$. In either case we have $r_A(a,b) = \frac{u(a)}{u(b)}$. By the definition of $\sim_A$, $a \not\sim_A b$ implies $r_A(a,b) = \frac{u(a)}{u(b)} \neq r_B(a,b)$ for any $B \neq A$ with $a, b \in B$. In other words, for any $B \neq A$ with $a, b \in B$, either $a \in F(B)$ and $b \notin F(B)$ or $a \notin F(B)$ and $b \in F(B)$. By Focal Richness, there are $B, C \neq A$ such that $r_B(a,b) = r_C(a,b)$. Hence, $\delta(B) = \delta(C)$, which contradicts the original hypothesis.

[35] Note that we are back to Step 2 if $r_A(a_1, b_1) = \frac{u(a_1)}{u(b_1)}$.

Second, suppose $a \sim_A b$. By way of contradiction, suppose either $a \in F(A)$ and $b \notin F(A)$ or $a \notin F(A)$ and $b \in F(A)$. Without loss of generality, suppose $a \in F(A)$ and $b \notin F(A)$. By the definition of $\sim_A$, $a \sim_A b$ implies $r_A(a,b) = \frac{u(a)}{u(b)}(1+\delta(A)) = r_B(a,b)$ for some $B \neq A$ with $a, b \in B$. Since $r_B(a,b) > \frac{u(a)}{u(b)}$, we have $r_A(a,b) = \frac{u(a)}{u(b)}(1+\delta(A)) = r_B(a,b) = \frac{u(a)}{u(b)}(1+\delta(B))$. Hence, $\delta(B) = \delta(C)$, which contradicts the original hypothesis.

Generalized Focal IIA. Take any $A \in \mathscr{A}$ and x, y, z with $x \not\sim_A y$ and $y \not\sim_A z$. By Fact 4, $x \not\sim_A y$ implies either $x \in F(A)$ and $y \notin F(A)$ or $x \notin F(A)$ and $y \in F(A)$. Similarly, $y \not\sim_A z$ implies either $y \in F(A)$ and $z \notin F(A)$ or $y \notin F(A)$ and $z \in F(A)$. By combining these two conditions, we can show that either $x, z \in F(A)$ or $x, z \notin F(A)$. By Fact 4, $x \sim_A z$.

Generalized Product Rule. Take any $A, B, C \in \mathscr{A}$ and $x, y, z \in X$ such that $x \sim_A$, $y \sim_B z$, and $z \sim_C x$. By Fact 4, we have either $x, y \in F(A)$ or $x, y \notin F(A)$, either $y, z \in F(B)$ or $y, z \notin F(B)$, and either $z, x \in F(C)$ or $z, x \notin F(C)$, respectively. In other words, we have $r_A(x,y) = \frac{u(x)}{u(y)}$, $r_B(y,z) = \frac{u(y)}{u(z)}$, and $r_C(z,x) = \frac{u(z)}{u(x)}$. Hence, $r_A(x,y)\, r_B(y,z)\, r_C(z,x) = 1$.

A.7 Proof of Proposition 5

By way of contradiction, suppose $\frac{r_A(x,y)}{r_B(x,y)} > \frac{r_A(y,z)}{r_B(y,z)} > \frac{r_A(z,t)}{r_B(z,t)} > 1$. By Proposition 1, $\frac{r_A(x,y)}{r_B(x,y)} > 1$ implies either $x \in F(A)$ or $y \in F(B)$, $\frac{r_A(y,z)}{r_B(y,z)} > 1$ implies either $y \in F(A)$ or $z \in F(B)$, and $\frac{r_A(z,t)}{r_B(z,t)} > 1$ implies either $z \in F(A)$ or $t \in F(B)$. Hence, it is enough to consider the following five cases.

Case 1. $x, y, z \in F(A)$. In this case,

$1+\delta(B) \geq \frac{r_A(x,y)}{r_B(x,y)} = \frac{1+\delta(B)\mathbb{1}\{y \in F(B)\}}{1+\delta(B)\mathbb{1}\{x \in F(B)\}} > \frac{r_A(y,z)}{r_B(y,z)} = \frac{1+\delta(B)\mathbb{1}\{z \in F(B)\}}{1+\delta(B)\mathbb{1}\{y \in F(B)\}} > 1$. Since $\frac{1+\delta(B)\mathbb{1}\{z \in F(B)\}}{1+\delta(B)\mathbb{1}\{y \in F(B)\}} \in \{1+\delta(B), 1, \frac{1}{1+\delta(B)}\}$, we obtain a contradiction.

Case 2. $x, y \in F(A)$ and $z \notin F(A)$. Note that $z \notin F(A)$ implies $t \in F(B)$. Then

$1+\delta(B) \geq \frac{r_A(x,y)}{r_B(x,y)} = \frac{1+\delta(B)\mathbb{1}\{y \in F(B)\}}{1+\delta(B)\mathbb{1}\{x \in F(B)\}} > \frac{r_A(y,z)}{r_B(y,z)} = (1+\delta(A))\frac{1+\delta(B)\mathbb{1}\{z \in F(B)\}}{1+\delta(B)\mathbb{1}\{y \in F(B)\}} > \frac{r_A(z,t)}{r_B(z,t)} = \frac{1}{1+\delta(A)\mathbb{1}\{t \in F(A)\}} \cdot \frac{1+\delta(B)}{1+\delta(B)\mathbb{1}\{z \in F(B)\}} > 1$. Note that $\frac{r_A(x,y)}{r_B(x,y)} = \frac{1+\delta(B)\mathbb{1}\{y \in F(B)\}}{1+\delta(B)\mathbb{1}\{x \in F(B)\}} > 1$ implies $y \in F(B)$ and $x \notin F(B)$. Moreover, $1+\delta(B) > \frac{r_A(y,z)}{r_B(y,z)} = (1+\delta(A))\frac{1+\delta(B)\mathbb{1}\{z \in F(B)\}}{1+\delta(B)} > 1$ implies that either $z \in F(B)$ and $\delta(B) > \delta(A)$ or $z \notin F(B)$ and $\delta(B) < \delta(A)$. Finally, $1+\delta(B) > \frac{r_A(z,t)}{r_B(z,t)} = \frac{1}{1+\delta(A)\mathbb{1}\{t \in F(A)\}} \cdot \frac{1+\delta(B)}{1+\delta(B)\mathbb{1}\{z \in F(B)\}} > 1$ implies $z \notin F(B)$ and $t \in F(A)$. Combining the last two implications, we have $\frac{r_A(y,z)}{r_B(y,z)} = \frac{1+\delta(A)}{1+\delta(B)} > \frac{r_A(z,t)}{r_B(z,t)} = \frac{1+\delta(B)}{1+\delta(A)} > 1$, a contradiction.

Case 3. $x, z \in F(A)$ and $y \notin F(A)$. Note that $y \notin F(A)$ implies $z \in F(B)$. Then

$\frac{r_A(x,y)}{r_B(x,y)} = (1+\delta(A))\frac{1+\delta(B)\mathbb{1}\{y \in F(B)\}}{1+\delta(B)\mathbb{1}\{x \in F(B)\}} > \frac{r_A(y,z)}{r_B(y,z)} = \frac{1}{(1+\delta(A))}\frac{1+\delta(B)\mathbb{1}\{z \in F(B)\}}{1+\delta(B)\mathbb{1}\{y \in F(B)\}} > \frac{r_A(z,t)}{r_B(z,t)} = \frac{1+\delta(A)}{1+\delta(A)\mathbb{1}\{t \in F(A)\}} \cdot \frac{1+\delta(B)\mathbb{1}\{t \in F(B)\}}{1+\delta(B)} > 1$. Note that $\frac{r_A(y,z)}{r_B(y,z)} = \frac{1}{(1+\delta(A))}\frac{1+\delta(B)\mathbb{1}\{z \in F(B)\}}{1+\delta(B)\mathbb{1}\{y \in F(B)\}} > 1$ implies $y \notin F(B)$, $z \in F(B)$, and $\delta(B) > \delta(A)$. Then, we have $\frac{1+\delta(B)}{1+\delta(A)} > \frac{r_A(z,t)}{r_B(z,t)} = \frac{1+\delta(A)}{1+\delta(A)\mathbb{1}\{t \in F(A)\}} \cdot \frac{1+\delta(B)\mathbb{1}\{t \in F(B)\}}{1+\delta(B)} > 1$, which implies $t \in F(B)$ and $t \notin F(A)$. Finally, we obtain $\frac{r_A(x,y)}{r_B(x,y)} = \frac{1+\delta(A)}{1+\delta(B)\mathbb{1}\{x \in F(B)\}} > \frac{r_A(z,t)}{r_B(z,t)} = 1+\delta(A)$, a contradiction.

Case 4. $x \in F(A)$ and $y, z \notin F(A)$. Note that $y, z \notin F(A)$ implies $z, t \in F(B)$. Then, we have $\frac{r_A(z,t)}{r_B(z,t)} = \frac{1}{1+\delta(A)\mathbb{1}\{t \in F(A)\}} > 1$, a contradiction.

Case 5. $x \notin F(A)$. Note that $x \notin F(A)$ implies $y \in F(B)$. Moreover, by Proposition 1, $\frac{r_A(x,z)}{r_B(x,z)} = \frac{r_A(x,y)}{r_B(x,y)} \cdot \frac{r_A(y,z)}{r_B(y,z)} > 1$ implies $z \in F(B)$. Similarly, $\frac{r_A(x,t)}{r_B(x,t)} = \frac{r_A(x,y)}{r_B(x,y)} \cdot \frac{r_A(y,z)}{r_B(y,z)} \cdot \frac{r_A(z,t)}{r_B(z,t)} > 1$ implies $t \in F(B)$. Then, we have $1 + \delta(A) \geq \frac{r_A(y,z)}{r_B(y,z)} = \frac{1+\delta(A)\mathbb{1}\{y \in F(A)\}}{1+\delta(A)\mathbb{1}\{z \in F(A)\}} > \frac{r_A(z,t)}{r_B(z,t)} = \frac{1+\delta(A)\mathbb{1}\{z \in F(A)\}}{1+\delta(A)\mathbb{1}\{t \in F(A)\}} > 1$. Since $\frac{1+\delta(A)\mathbb{1}\{z \in F(A)\}}{1+\delta(A)\mathbb{1}\{t \in F(A)\}} \in \{1 + \delta(A), 1, \frac{1}{1+\delta(A)}\}$, we obtain a contradiction.

A.8 Proof of Proposition 6

Suppose p is a BFLM with (F, δ) and satisfies Focal Betweenness. Take any A and $a, b \in A$ with $a \in F(A)$ and $b \notin F(A)$. Since $r_A(a,b) > r(a,b)$, by Focal Betweenness, we have $r_A(a,c) \geq r_{A\setminus\{b\}}(a,c) \geq r(a,c)$ for any $c \in A\setminus\{b\}$. Equivalently, $\frac{1+\delta(A)}{1+\delta(A)\mathbf{1}\{c \in F(A)\}} \geq \frac{1+\delta(A\setminus\{b\})\mathbf{1}\{a \in F(A\setminus\{b\})\}}{1+\delta(A\setminus\{c\})\mathbf{1}\{c \in F(A\setminus\{b\})\}} \geq 1$. First, it must be $a \in F(A \setminus \{b\})$, otherwise we will have $\frac{1+\delta(A\setminus\{b\})\mathbf{1}\{a \in F(A\setminus\{b\})\}}{1+\delta(A\setminus\{c\})\mathbf{1}\{c \in F(A\setminus\{b\})\}} < 1$ for any $c \in F(A\setminus\{b\})$. Therefore, $\frac{1+\delta(A)}{1+\delta(A)\mathbf{1}\{c \in F(A)\}} \geq \frac{1+\delta(A\setminus\{b\})}{1+\delta(A\setminus\{c\})\mathbf{1}\{c \in F(A\setminus\{b\})\}} \geq 1$. Second, for any $c \in F(A)$, we will have $c \in F(A \setminus \{b\})$. Otherwise, we will have $1 = \frac{1+\delta(A)}{1+\delta(A)\mathbf{1}\{c \in F(A)\}} < \frac{1+\delta(A\setminus\{b\})}{1+\delta(A\setminus\{c\})\mathbf{1}\{c \in F(A\setminus\{b\})\}}$. Hence, $F(A) \subseteq F(A \setminus \{b\})$ for any $b \notin F(A)$, consequently F is conditionally decreasing. Take any $c \in A \setminus \{b\}$ such that $c \notin F(A \setminus \{b\})$. By the previous argument, it must be that $c \notin F(A)$. Then we will obtain $\delta(A) \geq \delta(A \setminus \{b\})$ for any $b \notin F(A)$, consequently δ is conditionally increasing.

A.9 Supplement to Section 6.3

Let us derive the OLS estimate $\hat{\beta}$ of β that we have mentioned in Section 6.3. To obtain $\hat{\beta}$ we run the following OLS regression:

$$\min_{\alpha,\, \beta} \sum_{i \leq n,\; l \in \{A,B\}} (y_1^l - \alpha\,(x_i - x_n) + \beta\,\bar{p}_i^l)^2,$$

where $\bar{p}_1^A = p_1 - p_n$, $\bar{p}_1^B = p_1' - p_n$, and for each $2 \leq j \leq n$, $\bar{p}_j^l = p_j - p_n$. Then this regression returns

$$\hat{\beta} = \beta + \epsilon^T \, \frac{x^T(x^T p) - p^T(x^T x)}{x^T x\, p^T p - (x^T p)^2} + \delta\, \frac{(x_1 - x_n)\, x^T p - p_1^B x^T x}{x^T x\, p^T p - (x^T p)^2},$$

where $y = \alpha\, x - \beta\, p + \epsilon$, and $x = (x_1 - x_n, \ldots, x_{n-1} - x_n, x_1 - x_n, \ldots, x_{n-1} - x_n)$, $p = (\bar{p}_1^A, \ldots, \bar{p}_{n-1}^A, \bar{p}_1^B, \ldots, \bar{p}_{n-1}^B)$, and $y = (y_1^A - y_n^A, \ldots, y_{n-1}^A - y_n^A, y_1^B - y_n^B, \ldots, y_{n-1}^B - y_n^B)$. Note that the bias $\delta\, \frac{(x_1 - x_n)\, x^T p - p_1^B x^T x}{x^T x\, p^T p - (x^T p)^2}$ is equal to $\frac{\delta}{p_1 - p_1'}$ when $n = 2$.