American Economic Review 2025, 115(4): 1226–1257
https://doi.org/10.1257/aer.20220130

Weighted Linear Discrete Choice†

By Christopher P. Chambers, Yusufcan Masatlioglu, Paulo Natenzon, and Collin Raymond*

We introduce a new model of stochastic choice that assigns each choice option a utility, along with a salience parameter reflecting economic frictions. We characterize our model behaviorally and investigate its comparative statics properties. We show that the model generates intuitive closed-form solutions in equilibrium settings where firms can choose price, quality, and advertising. In addition, we show that the model allows for flexible substitution patterns and changes in market shares across choice sets. We demonstrate that our model can be easily identified and can outperform alternatives in demand prediction. (*JEL* D11, D21, D43, M37)

This paper introduces a *simple and tractable* yet *flexible* model of probabilistic discrete choice, which we call the weighted linear (WL) model of discrete choice. In our model, each choice option is described by two parameters: One reflects the underlying quality or utility of an item, while the second captures the ease of choosing an item as a result of prominence, or what has been termed salience in the cognitive science and marketing literatures.

The WL model adds a single parameter to describe each option compared to the widely used Luce model (also called the multinomial logit model). Although many models generalize the Luce model, we will demonstrate that our model has several advantages over the alternatives in environments where products exhibit asymmetries, especially when responding to changes in the choice set. Our formulation allows the WL model to sidestep some of the "counterintuitive implications" of commonly used models that Berry and Pakes (2007) highlight (see also Benkard and Bajari 2001). For example, many of these models predict that a superior product will obtain zero market share in the limit as a very large number of inferior products are introduced. The WL produces intuitive market share responses to the introduction of new products and can account for flexible cross-price substitution effects as well as the relationship between markups and advertising under oligopolistic competition. In addition, the WL performs well when conducting demand estimations and counterfactual predictions relative to alternative formulations of random choice.

* Chambers: Georgetown University (email: christopher.chambers@georgetown.edu). Masatlioglu: University of Maryland (email: yusufcan@umd.edu). Natenzon: WashU Olin Business School (email: pnatenzon@wustl.edu). Raymond: Cornell University (email: collinbraymond@gmail.com). Pietro Ortoleva was the coeditor for this article. We thank Ian Krajbich, Joe Mazur, Yesim Orhun, Andre Veiga, and Chenyu Yang for their helpful comments. We are especially grateful to Christopher Turansick for excellent research assistance.
† Go to https://doi.org/10.1257/aer.20220130 to visit the article page for additional materials and author disclosure statement(s).

Section I introduces the WL model. The model ranks outcomes by two orderings: a utility function u, which captures "how good" an item is, and a salience function, m, which captures "how salient" an item is. The probability of choosing x from S is equal to

$$\rho(x|S) = \underbrace{\frac{m(x)}{\sum_{y \in S} m(y)}}_{\text{Base Probability}} + \underbrace{m(x)\left[u(x) - \bar{u}_m(S)\right]}_{\text{Comparative Probability Transfer}}.$$

Here, $\bar{u}_m(S)$ is the average utility of outcomes in S weighted using m; that is, $\bar{u}_m(S) = \frac{\sum_{y \in S} m(y)u(y)}{\sum_{y \in S} m(y)}$. The probability of choosing x is the sum of two components: a base probability plus a comparative probability transfer. The first, the base probability, reflects how easy it is to think of x compared to the rest of the choice set and mirrors Luce's choice rule (Luce 1959). The comparative probability transfer, in line with Fechernian stochastic choice (Fechner 1860), reflects the difference in utility between the item and a weighted average of other items in the choice set, where the weights depend on the salience of the items. The salience of x then scales this difference. Thus, unlike Luce and many of its existing generalizations, we move the main source of choice stochasticity from utility to salience. For example, when all outcomes have the same utility, the comparative probability transfer term above vanishes, and ties in utility are broken by salience; while in the limit when all items become very salient, choice becomes deterministic (and the highest-utility item is chosen).

We link our model to an object-specific notion of salience based on prominence that is in line with intuitions in cognitive science on visual salience (e.g., Milosavljevic et al. 2012; Towal, Mormann, and Koch 2013; Janiszewski, Kuo, and Tavassoli 2013; Weingarten and Hutchinson 2017; Dai, Cone, and Moher 2020), and in marketing on brand salience (e.g., Hoyer and Brown 1990) that has also been tied to advertising (Sutherland and Galloway 1981; Moran 1990; Yi 1990; Miller and Berry 1998).[1] Consistent with these approaches, and Bordalo, Gennaioli, and Shleifer's (2022) definition, we think of salience as capturing a "bottom-up" mental process, one that affects choice in an automatic and involuntary fashion. We review some stylized facts from these literatures in Section I and how they accord with our approach. Although recent models in cognitive science have been developed to explain attention and visual salience as well as choice, our model, which abstracts away from the choice process, has the advantage of being quite tractable, even for large choice sets, as well as matching many stylized facts.

At the same time, our approach is the stochastic choice equivalent of a linear demand system, where market shares are linear in net utility differences between items, which is widely used in economics.[2] In our model, demand for a product reflects (i) a base component independent of the utility of the item and (ii) a component that reflects the difference in utility between the item and the average item scaled by some number

[1] Bordalo, Gennaioli, and Shleifer (2022) describes salient stimuli as those that attract attention through an involuntary mental process, and are typically driven by contrast, surprise, or prominence. There is a recent body of work in economics on context and attribute-based salience driven by contrast and surprise (e.g., Bordalo, Gennaioli, and Shleifer 2013), which embodies a distinct approach to ours.

[2] The approach of linear demand systems has been used extensively in applied settings (early references include Shubik and Levitan 1980; Dixit and Stiglitz 1977; Spence 1976; and Singh and Vives 1984).

that captures how easy it is to consider alternative products (i.e., a kind of friction) in the market (as the number gets bigger, the best items eventually attract the entire market). In other words, our model merges the ideas of Luce (1959) and of linear demand systems using salience. In particular, our model yields both the standard multinomial logit model and the “simple” version of linear demand systems (with all firms symmetric in salience) as special cases.

The WL model has straightforward behavioral foundations. Section II demonstrates that three simple axioms summarize all behavioral implications of the WL model. Structurally, the behavioral content of the model is characterized by a novel type of acyclicity condition. These axioms connect the (unobserved) components of the model to observed choice behavior. We show that identification can be achieved by observing choices on relatively few choice sets.

Section III compares the WL to the additive perturbed utility (APU) approach of Fudenberg, Iijima, and Strzalecki (2015b) and the random utility model (RUM). Although logically distinct from the APU, our model has a nontrivial intersection with it. Both models nest the multinomial logit model and the simple linear demand approach described above. Our model, in turn, is nested in the well-known class of RUMs (characterized by Falmagne 1978). This means our model shares a common interpretation with most existing models of random choice as representing the choice frequencies of a population of heterogeneous but otherwise standard rational agents. Drawing on the population interpretation of our model, we introduce a measure we term “rank polarization.” We define it as the sum of the proportion of consumers who rank an option as their first (best) option and as their last (worst) option among all the options, capturing how extreme are the preferences for a given item. Rank polarization subsequently plays a role in interpreting the WL model’s ability to explain empirical patterns and to predict market shares.

Section IV demonstrates the explanatory power of the WL model, showcasing its ability to accommodate several empirical regularities and highlighting the potential benefits relative to the multinomial logit as well as a wider class of discrete choice models. First, we examine substitution patterns in our model, including those derived from utility, salience, and price changes. We show how our model can allow for richer patterns of cross-price substitution relative to the logit approach. We show that in line with intuitions, items with higher salience induce higher cross-price elasticities. Second, we show that our model can accommodate natural shifts in market share when new products are introduced and that these shifts are tied to rank polarization. We show that items that are rank polarized have more similar market shares across choice sets of different sizes and that higher utility means that an item will lose relatively less market share when new products are introduced. In fact, high-utility items can even maintain nonnegligible market share when an infinite number of new products are introduced, unlike many other models of discrete choice.

Section V provides an application to oligopolistic competition with advertising. We extend the model to allow for firms to compete by influencing utility via setting price and salience via advertising (in line with evidence from marketing that shows brand salience is influenced by advertising). We show that the model can tractably capture these kinds of oligopolistic competition, delivering closed-form solutions, and that the model generates relationships between markups, the number of firms, and advertising consistent with intuitions and existing evidence.

Section VI highlights the predictive power of the WL model. We compare the performance of the WL model to some of the most well-known models in discrete choice estimation, namely classic logit, nested logit and (covariance) probit. To assess the strengths of each model in a variety of situations, we use simulated data. We estimate each model in 14,850,000 simulated datasets, covering the entire range of demand systems that come from a population of heterogeneous rational agents (i.e., the entire range of demand systems generated by RUMs). We compare the out-of-sample predictions of these models using several standard metrics, and we organize the results by the level of rank polarization (extremeness) of tastes in each simulation. Rank polarization turns out to be an important measure to track the predictive performance of the WL model versus the alternatives. The logit model becomes heavily disadvantaged once rank polarization moves away from the knife-edge case of 0.5. Nested logit improves predictions over logit somewhat, but WL and probit clearly outperform both logit and nested logit. We show probit makes slightly better predictions than WL for larger choice sets, WL has slightly better predictions than probit for intermediate-sized sets, and the picture is mixed for predictions in smaller choice sets. Hence, while the WL does not clearly dominate the probit model, and performs slightly worse in some cases, it could prove a better model to use given its smaller number of parameters, closed-form choice probabilities, computational simplicity, and easy interpretation.

Finally, Section VII concludes. The Supplemental Appendixes include proofs and discuss other relevant ideas, such as behavioral properties of the extension of our model, which allows for zero choice probabilities.

I. Model

We begin by describing our model in an abstract environment, and in later sections apply it to particular settings. Let X be a finite set of alternatives. Let $\mathcal{X}$ be the set of all probability measures on X. That is, $\rho(\cdot|X) \in \mathcal{X}$ implies $\rho(x|X) \geq 0$ and $\sum_{x\in X}\rho(x|X) = 1$. Let $\mathcal{D}$ denote the set of nonempty subsets of X. For every $S \in \mathcal{D}$, denote by $\mathcal{S}$ the elements in $\mathcal{X}$ that naturally induce probability measures on S; that is, $\rho(\cdot|S) \in \mathcal{S}$ means $\rho(x|S) \in \mathcal{X}$ and $\rho(x|S) = 0$ whenever $x \notin S$. $\rho(x|S)$ denotes the choice probability (market share) of x in S. We also denote the sum of choice probabilities in $T \subset S$ as $\rho(T|S)$. Similarly, for any real function, f on X, $f(S)$ denotes the sum of $f(x)$ for all $x \in S$. We will denote binary choices as $\rho(x,y)$ instead of $\rho(x|\{x,y\})$. A *stochastic choice* (sometimes called a stochastic choice rule or stochastic choice function) is a family $\{\rho(\cdot|S)\}_{S\in\mathcal{D}}$, where each $\rho(\cdot|S) \in \mathcal{S}$.

In our stochastic choice model, each alternative x is represented by two values: its utility $u(x)$ and its salience $m(x)$. Utility captures the value of each option, while salience captures how accessible or prominent the object is to the decision-maker (we elaborate in detail the interpretation of our salience parameter in Section IA). For any set S, we let

$$\bar{u}_m(S) \equiv \frac{\sum_{x\in S} u(x)m(x)}{m(S)}$$

denote the weighted utility average of options in S with respect to m.

DEFINITION 1: *A stochastic choice ρ is consistent with a weighted linear model on $\mathcal{D}$ if there exist functions $u : X \to \mathbb{R}$ and $m : X \to \mathbb{R}_{++}$, such that, for each $S \in \mathcal{D}$,*

$$(1) \qquad \rho(x|S) = \frac{m(x)}{m(S)} + m(x)\left[u(x) - \bar{u}_m(S)\right].$$

We also say (u,m) represents ρ, or (u,m) is a WL representation of ρ.

A pair (u,m) must generate nonnegative choice probabilities in (1) to be a WL representation. A simple restriction on (u,m) is necessary and sufficient.

PROPOSITION 1: *(u,m) is a WL representation for some ρ if and only if, for all $x \in X$,*

$$u(x) \geq \bar{u}_m(X) - \frac{1}{m(X)}.$$

In Section IB, we show that the expression $\bar{u}_m(X) - 1/m(X)$ above has a natural interpretation as the shadow price in the problem of maximizing the decision-maker's expected utility subject to a quadratic penalty of paying attention to make choice less random. Proposition 1 shows that a pair of functions (u,m) generates a stochastic choice according to the WL model if and only if the smallest utility among the options in the grand set X is larger than this shadow price.

Like most generalizations of the Luce model, we introduce additional parameters —in our case, a single additional value for each option—which we call salience. The first term of equation (1) reflects the fact that each item has a base probability of attracting consumers in a way proportional to its salience. The second term reflects the fact that products gain or lose consumers in proportion to the difference between their utility $u(x)$ and the salience-weighted average utility in the market $\bar{u}_m(S)$: More salient items attract more attention, and so their utility benefits (or drawbacks) are more easily noticed, and so shift choice more. These two terms allow us to naturally combine two well-known approaches that capture nondeterministic choice: the Luce model and models of linear demand.

To see this, first consider the case where the utility function is constant but salience differs across alternatives. If $u(x) = u(y)$ for all x, y, the second term in the representation disappears, and the choice probabilities are solely driven by m:

$$\rho(x|S) = \frac{m(x)}{m(S)}.$$

Clearly, this is the classical model of (Luce 1959) but with a key conceptual difference: Here, randomness in choice is driven by salience, while in the Luce model, it is driven by utility. Thus, in our model, we obtain Luce's choice probabilities when utility is constant, and the ties in utility are broken by salience. In fact, this is the only way to obtain Luce choice probabilities in our framework: It is straightforward to verify that ρ is a WL where $u(x) = u(y)$ for all $x, y \in X$ if and only if it has a Luce representation.[3]

Now consider the other extreme case, where salience remains constant: $m(x) = \bar{m}$ for all x. In this case, the first term is simply $1/|S|$; each alternative attracts attention uniformly. Then, the weighted utility average becomes the ordinary average, $\bar{u}_m(S) = \bar{u}(S)$, and we have

$$\rho(x|S) = \frac{1}{|S|} + \bar{m}\big[u(x) - \bar{u}(S)\big].$$

This is equivalent to the basic linear demand system that features prominently in many models of monopolistic competition, where market shares are assumed to be linear in the net utility of a product.[4] To see this more clearly, without loss, we define $u(x) = \bar{u} - p(x)$. If we call $p(x)$ the price of x, then we can define demands as $\rho(x|S) = 1/|S| + \bar{m}\big[\bar{p}(S) - p(x)\big]$. Salience can be interpreted as reflecting how easily consumers are able to notice the item, including its utility relative to the average utility: If all items are very salient, then utility differences are easily noticed and have a large impact on choice, while if items are not very salient, utility differences have a small impact on choice.

As mentioned, a key conceptual difference between our model and the Luce model, as well as many of its generalizations, is that in the WL model stochasticity is driven by salience rather than utility (one exception is the similar approach of Fudenberg, Iijima, and Strzalecki 2015b). Salience drives stochasticity in two ways. First, the overall level of salience in the market determines how close choices will approximate utility maximization. If all items have low salience, any given item struggles to draw attention, and agents are constrained to choose among only the few items that do come to mind (which may have low utility). In other words, when overall salience in the market is low, the importance of utility also diminishes, as the second term in (1) vanishes. Conversely, as the salience of items grows (that is, if we multiply the salience of all items by some large scalar), choice probabilities become more concentrated on high-utility items, as the second term in (1) starts to dominate. This is because agents can more easily conceive of all alternatives and choose the one with highest utility. In fact, extending the model using the microfoundation of Section IB to allow zero choice probabilities, one can show that in the limit, as the salience of all items scales to infinity, we would observe a convergence to deterministic choice of the highest-utility item.

While the total salience in the market drives the amount of stochasticity in choice, the relative salience of different items determines the shape of the stochasticity. This is most clearly seen in the two cases where choices are determined mostly by salience rather than utility: first, when all items have the same utility or are close to

[3] Our model is distinct from the nested logit where alternatives within a nest must satisfy Luce's IIA, although the idea of nests can be introduced in our framework. Kovach and Tserenjigmid (2022) provides several characterizations for the nested logit and its generalizations.

[4] Linear probability models are popular for their tractability in discrete choice estimation (e.g., Section 4.2 in Ben-Akiva and Lerman 1985) and appear in solution concepts with boundedly rational players in noncooperative games (Rosenthal 1989; Voorneveld 2006). Early examples are Shubik and Levitan (1980); Dixit and Stiglitz (1977); Spence (1976); and Singh and Vives (1984), while Choné and Linnemer (2020) provides a survey.

indifference and second, as we have seen, when the overall salience in the market is low. In both cases, choice probabilities are driven by the first term in (1), and relative salience determines choice: If one option is twice as salient as another, it is chosen twice as often.

In interpreting the parameters, thinking about the relative impact of utility and salience on choice, and linking the WL model to empirical evidence (such as that discussed in Section IA), it is useful to think about what happens to demand (i.e., the probability of being chosen) for a given object as both u and m change. First, we note that increasing either $u(x)$ or $m(x)$ results in an increase in the choice probability $\rho(x|S)$. For both of these changes, there is a direct effect of the change in the parameter, as well as an indirect effect, which emerges because of the change in $\bar{u}_m(S)$. Given that $\rho(x|S) > 0$, algebra shows that

$$\frac{\partial \rho(x|S)}{\partial u(x)} = \frac{m(x)m(S\setminus\{x\})}{m(S)} \quad \text{and} \quad \frac{\partial \rho(x|S)}{\partial m(x)} = \frac{m(S\setminus\{x\})}{m(x)m(S)}\rho(x|S).$$

Viewed as a function of u, for a given $S \in \mathcal{D}$, $\rho(x|S)$ is linear. This implies that by reducing $u(x)$, it is possible for x to be chosen with probability arbitrarily close to zero. Similarly, increasing $u(x)$ enough will eventually lead to all other items being chosen with zero probability. It is also easy to see from this that $m(x)$ and $u(x)$ are complements: Increases in utility have a larger impact on highly salient items, and vice versa.

Unlike utility u, the effect of salience m on choice probabilities is nonlinear. If $u(x) \geq \bar{u}_m(S)$, then the probability of x being chosen will go to one as m increases. If not, the probability of x being chosen converges to an upper bound strictly below one. In contrast, as $m(x)$ decreases, the probability of x being chosen falls to zero (in line with the idea, discussed in Section IA, that "unseen is unsold").

Holding the initial choice probability constant, the effect of a change in salience on choice is larger when $m(S)$ is larger, e.g., when the choice set is large. As the choice set expands, the impact of a marginal change in the utility of x on its market share approaches $m(x)$, while the impact of a marginal change in its salience approaches $\rho(x|S)/m(x)$. A more subtle consequence of the asymmetry between u and m is that larger sets tend to benefit larger utility items, as follows.

PROPOSITION 2: *$u(x) \geq u(y)$ and $\rho(x|S) \geq \rho(y|S)$ implies $\rho(x|S \cup T) \geq \rho(y|S \cup T)$.*

It is easy to find examples where a lower-utility product enjoys a higher market share in a smaller set ($u(x) > u(y)$ while $\rho(x|S) < \rho(y|S)$), but enlarging the product variety favors the higher-utility product ($\rho(x|S \cup T) > \rho(y|S \cup T)$).[5] However, Proposition 2 shows this can never happen the other way around.

[5] For instance, when $u(x) = 2$ and $u(y) = 1$ while $m(x) = 1/10$ and $m(y) = 2/10$, the market share of y is 50 percent larger than x's in a binary comparison. But once the set is expanded to include z with $u(z) = 1$ and $m(z) = 6/5$, y's market share becomes 25 percent smaller than x's.

A. *Interpreting Salience*

A novel feature of our approach is the importance of salience or market presence in determining choice probabilities. Salience, as a term, has been widely applied across disciplines to mean a variety of related concepts. Bordalo, Gennaioli, and Shleifer (2022) summarizes salience as capturing an involuntary process of attracting attention through contrast, surprise, or prominence. Moreover, some work focuses on the salience of objects, others on the salience of attributes. For example, Bordalo, Gennaioli, and Shleifer (2013) develops a model where certain aspects of an item are more or less salient depending on the choice set (and so depends on contrast). In a distinct line of work, Chetty, Looney, and Kroft (2009) and related papers consider what happens when one aspect of a good may be less salient (e.g., the posttax price, relative to the pretax price).

Our model takes a complementary approach, where salience is an inherent feature of the object itself, rather than being derived from comparisons to other objects via, for example, contrast effects (although the weight of salience in determining choice is determined by the context in our model), and can be entirely independent of utility. We think that this corresponds most clearly to a notion of prominence and relates to existing work on visual salience (in cognitive science) or brand salience (e.g., in marketing). Of course, work in both those fields also relates salience to other context-related factors, like contrast. We assume that the salience of an object is treated as given by the decision-maker, although it may be manipulable by others, through, for example, advertising. We now turn to discuss how our particular model relates to stylized evidence on the importance of salience in both cognitive science and marketing literature.

First, we focus on the relatively recent literature examining visual salience. The initial stylized fact that our model captures is fairly intuitive: Extensive evidence indicates that choice probabilities increase in salience. The best evidence comes from work on visual salience. Although the evidence in economics is relatively small (Li and Camerer 2022; Reutskaja et al. 2011), there is extensive documentation in cognitive science (e.g., Bialkova and van Trijp 2011; Milosavljevic et al. 2012; Orquin and Loose 2013; Orquin and Lagerkvist 2015; Peschel, Orquin, and Loose 2019; Bhatnagar and Orquin 2022). Moreover, the evidence indicates that salience is combined with subjective value (i.e., utility) in order to determine choice (Navalpakkam et al. 2012; Towal, Mormann, and Koch 2013). Milosavljevic et al. (2012) find that the effects of salience on choice are particularly strong when preferences are weak, and salience can be as influential as preferences in determining choice probabilities, although other studies find subjective valuations matter more (Towal, Mormann, and Koch 2013).

Second, studies have found that increasing salience has a stronger effect for high-utility items (Krajbich, Armel, and Rangel 2010; Smith and Krajbich 2019). In other words, subjective value (i.e., utility) and salience are complements, exactly what occurs in the WL model.

Third, in the WL model (although we focus on situations where all choice probabilities are positive), it is the case that when utility differences are large enough, the choice is determined by utility (so long as all items have positive salience). This is in line with findings in Li and Camerer (2022), where if the value difference between items is large enough, salience seems to have a negligible impact on choice.

Fourth, the WL model predicts that as salience increases, choice probabilities increase, but that a relatively unattractive item will never be always chosen just because of a high m. It is the case that in the literature increased salience does not lead to an item always being chosen, e.g., Li and Camerer (2022) (although there is more limited evidence on this phenomenon relative to other patterns we discuss).

Fifth, again though we focus on strictly positive probabilities, clearly as the salience of an item goes to zero, the probability that it is chosen goes to zero. This captures the adage in marketing that "unseen is unsold" (see Chandon et al. 2009 for a recent discussion of this).

Last, the WL model makes predictions about what happens to the impact of u and m as the choice set size changes (e.g., Proposition 2). Unfortunately, most work studying the relationship between visual salience and choice has focused on binary choices, although a few have focused on choice sets of sizes 3 and 4 (Towal, Mormann, and Koch 2013; Krajbich and Rangel 2011; Gluth, Spektor, and Rieskamp 2018). The one study we know of looking at relatively larger choice set sizes, Thomas, Molter, and Krajbich (2021), looks at choice sets of size 9 to 36 and finds little role for salience in choice probabilities, in contrast to the work on binary choice sets, and in line with what our model predicts. Although not explicitly testing visual salience, Reutskaja et al. (2011) considers choice sets of sizes 4, 9, and 16 and finds that item location matters more for choice probabilities going from 4 to 9; its importance then falls going from 9 to 16.

The literature on brand salience in marketing often leverages field data (with both the benefit of external validity and the cost of identification). Broadly speaking, this field's findings about the impact of salience accord with, and oftentimes are directly related to, the findings on visual salience in cognitive science. The salience of a good or brand has been conceptualized as measuring the prominence of a brand in memory (Wedel and Pieters 2000; Romaniuk and Sharp 2003; Nedungadi 1990; Zhang et al. 2021), visual, perceptual, or cognitive awareness (Wedel and Pieters 2000; Van der Lans, Pieters, and Wedel 2008; Janiszewski, Kuo, and Tavassoli 2013; Weingarten and Hutchinson 2017; Dai, Cone, and Moher 2020); how often it enters into consideration sets; or how loyal customers are to it (Romaniuk and Sharp 2004; Ehrenberg, Barnard, and Scriven 1997). Extensive work has shown that, as in our model, higher salience implies a higher chance of being chosen (Sutherland and Galloway 1981; Ehrenberg, Barnard, and Scriven 1997; Romaniuk and Sharp 2003; Vieceli and Shaw 2010). Not only that, but salience (as measured through awareness) can impact choice even in the face of utility differences such as quality and price (Macdonald and Sharp 2000; Hoyer and Brown 1990).

The marketing literature also suggests that firms can directly alter the salience of their brands via advertising (Sutherland and Galloway 1981; Moran 1990; Yi 1990; Miller and Berry 1998; Yang et al. 2024). We specifically explore the implications of this in Section IV, where we consider an equilibrium setting where firms can choose both the price and salience (via advertising) of their products.

Of course, this is not to imply that our model captures all important details about salience or even visual salience. Clearly, our model abstracts away from attribute-based notions of salience, as in Bordalo, Gennaioli, and Shleifer (2013) and the following literature. Moreover, our model is silent about important interactions between salience, directed attention, and time that models such as the attentional drift-diffusion model capture (Krajbich 2019; Krajbich, Armel, and Rangel 2010). More generally, many models of salience in cognitive science explicitly take into account choice process data, which our model neglects. Although this means our model is less rich, it means it is far more tractable for solving out for choice probabilities.

The cognitive science literature has developed a set of models that are meant to explain how salience and choice interact (e.g., the attention drift-diffusion model, Krajbich, Armel, and Rangel 2010). These models are process focused and describe not only the choice probabilities but the process by which those probabilities are reached. However, they have primarily been applied to binary and trinary (Krajbich and Rangel 2011) choice sets and are difficult to solve in closed form. In contrast, our model, while neglecting to capture the process of choice, attends to many of the same stylized facts and could be seen as a simple reduced-form alternative that can be used to understand salience.

B. *Microfoundation*

We now show that our model can be microfounded as the solution to a simple utility maximization problem subject to the cost of paying attention. In particular, the decision-maker chooses the probability with which each option is realized, as in other recent models of deliberate randomization. Motivated by the "stochastic choice as optimization" paradigm of Machina (1985) and Cerreia-Vioglio et al. (2019), we suppose that for any set S, probabilities reflect the solution to maximizing expected utility less quadratic cost:

$$\mathcal{P}(S) = \underset{\rho(\cdot|S)\in\mathcal{S}}{\arg\max} \sum_{x\in S}\left[u(x)\rho(x|S) - \frac{1}{2m(x)}\rho(x|S)^2\right]. \tag{2}$$

The objective function in (2) describes individuals who want to maximize utility but find it costly to pay attention to the utility of the options. We can understand the cost of paying attention as follows. First, suppose every option is equally salient; that is, for all x, we have $m(x) = \bar{m}$ for some fixed $\bar{m} > 0$. Then, we may write the cost in (2) as a quadratic penalty for deviating from purely random choice,

$$\frac{1}{2}\sum_{x\in S}\frac{1}{\bar{m}}\left[\rho(x|S)^2 - \left(\frac{1}{|S|}\right)^2\right];$$

that is, the cost of not paying attention at all is zero, in which case every option is chosen with the same probability, namely, $1/|S|$. Paying attention to maximize utility is costly, and the cost is quadratic. The salience level $\bar{m}$ modulates the cost of paying attention: It is easier or less costly to pay attention when all the options are more salient.

Our key assumption is that salience is always positive: $m(x) > 0$ for all x. When the options vary in their salience, it is easier to pay attention to more salient options. That is, less salient options have steeper cost:

$$\frac{1}{2}\sum_{x\in S}\frac{1}{m(x)}\left[\rho(x|S)^2 - \left(\frac{1}{|S|}\right)^2\right] = \sum_{x\in S}\frac{1}{2m(x)}\rho(x|S)^2 - K,$$

where the right-hand side of this equality is the cost in the optimization problem in (2), except for the presence of the constant term K that does not matter for optimization and can safely be omitted.

We will focus on situations where the solution to (2) features only positive probabilities. Then, the first-order conditions to (2) are

$$u(x) - \frac{1}{m(x)}\rho(x|S) = \Lambda(S) \text{ for all } x \in S,$$

where $\Lambda(S)$ is the Lagrange multiplier of the constraint that the probabilities add up to 1. Summing across the elements of S and a bit of algebra imply

$$\Lambda(S) = \frac{\sum_{y\in S} m(y)u(y) - 1}{\sum_{y\in S} m(y)}.$$

Plugging back into FOC, we get $\rho(x|S) = m(x)\left[u(x) - \Lambda(S)\right]$, which yields the WL choice probabilities in (1) after a simple rearranging of the terms. Hence, the market shares in the WL model (1) can be obtained from the optimizing behavior of a single representative agent who faces a cost of paying attention, where the cost of paying attention to an item is modulated by its salience.

This microfoundation sheds light on the asymmetric role played by utility and salience in our model. The marginal benefit of increasing the market share of an option is constant and equal to its utility parameter. The marginal cost is scaled by salience, but, unlike the marginal benefit, it is also proportional to the option's current market share. This explains the different comparative statics on u and m. For example, as we have seen, it is always possible to increase the market share of an item by raising its utility; however, raising the salience of a low-utility item will never lead the item to dominate the market.

Finally, this microfoundation allows us to reinterpret the parameter restrictions in Proposition 1. This result shows that all choice probabilities are positive if and only if $u(x) \geq \bar{u}_m(X) - 1/m(X)$. Given that $\bar{u}_m(X) - 1/m(X) = \Lambda(X)$, the restriction requires utility of all items to be larger than the shadow price (of relaxing the constraint that probabilities sum to 1) in the grand choice set X. In Supplemental Appendix B, we use the microfoundation in (2) to obtain an extended version of our model allowing for options to be chosen with zero probability.

II. Characterization, Uniqueness, and Identification

We now discuss the behavioral implications of our model.[6] The first behavioral postulate is the often-invoked notion of positivity. Positivity says that every alternative is chosen with a positive probability.[7]

[6] Readers interested in applying the model may safely skip ahead to Section III.

[7] As discussed, this assumption cannot be rejected by any finite data set, but we relax it in the Supplemental Appendix.

AXIOM 1: [*Positivity*] $\rho(x|S) > 0$ *for every* $x \in S$ *and* $S \in \mathcal{D}$.

The next behavioral postulate is a well-known property in the stochastic choice literature. It states that when the competition gets fiercer among alternatives, choice probabilities strictly decrease.

AXIOM 2: [*Strict Regularity*] $\rho(y|S) < \rho(y|S\setminus\{x\})$ *for every* $x \in S$ *and* $S \in \mathcal{D}$.

To state our next axiom, we define $d(x|S,T) := \rho(x|S) - \rho(x|T)$, where $S \neq T$ and $x \in S \cap T$. The quantity $d(x|S,T)$ is simply the change in the probability of choosing x as the choice set T changes to S.

AXIOM 3: [*Product Rule for Differences* (*PRD*)] *For any* x,y,z *and* $S_i, T_i \in \mathcal{D}$,

$$d(x|S_1,T_1)d(y|S_2,T_2)d(z|S_3,T_3) = d(z|S_1,T_1)d(x|S_2,T_2)d(y|S_3,T_3)$$

whenever the expressions are well defined.

In terms of interpretation, the property states that this product of probability differences depends only on the collection of elements with respect to which differences are taken. It does not, however, depend on how these elements are assigned to the given choice sets.

The Luce model satisfies Axiom 3. In addition, the Luce model satisfies a product rule[8] for choice probabilities:

$$\rho(x|S_1)\rho(y|S_2)\rho(z|S_3) = \rho(x|S_3)\rho(z|S_2)\rho(y|S_1)$$

whenever $x,y \in S_1$, $y,z \in S_2$ and $x,z \in S_3$. The left-hand side is the probability of a revealed preference choice cycle in the direction $x \to y \to z$, and the right-hand side is the probability of a choice cycle in the opposite direction.

It is easy to show that, if Axiom 3 holds in the full domain, then the axiom implies that an analogous condition holds for products of differences of length n, for any $n > 3$. Conversely, taking $y = z$, Axiom 3 implies a strictly weaker condition of length $n = 2$. When all differences are positive, this becomes

$$\frac{d(x|S_1,T_1)}{d(y|S_1,T_1)} = \frac{d(x|S_2,T_2)}{d(y|S_2,T_2)},$$

which is a version of Luce's independence of irrelevant alternatives (IIA) for differences. It says the ratio between differences for two options x,y only depends on x and y; it does not depend on which other alternatives were present initially in the

[8] See, for example, Ahumada and Ulkü (2018); Echenique and Saito (2019); and Horan (2021).

set with x and y nor on which other alternatives were added or removed when the set changed.

We now state our characterization. This result does not require that we observe choices from the full domain, only the menus with sizes 2 and 3.

THEOREM 1: *Suppose $\mathcal{D}$ contains all menus with sizes 2 and 3. Then a stochastic choice function ρ has a WL representation on $\mathcal{D}$ if and only if it satisfies Axioms 1–3.*

The idea of the proof for sufficiency is as follows. We first define the salience of each alternative by using the ratio of differences $d(x|S,T)/d(y|S,T)$, where S and T are menus with sizes 2 and 3. Then, instead of directly constructing the utility function, we define the "shadow values" for an optimization problem, for each set in the domain. This step helps us to define the utility function. We then show that the data can be represented by the WL model.

Theorem 1 provides two simple tests for our model. While Axiom 2 is innocuous, Axiom 3 is based on a principle similar in spirit to Luce's IIA. In our axiom, the ratio of *relative* levels is important rather than the absolute levels, as in Luce's IIA.

A. *Uniqueness*

Our model enjoys strong uniqueness properties. If (u,m) represents ρ, then $(au + b, m/a)$ also represents ρ for $a > 0$ and b. We also show that if (u,m) and (u',m') represent the same choice data, they are equivalent up to the same class of transformations. The utility function is unique up to an affine transformation, whereas the salience function is unique up to a scale transformation. The scale parameter of utility is the inverse of the scale parameter of salience.

THEOREM 2 (Uniqueness): *Let (u,m) be a WL representation of ρ. Then (u',m') is a WL representation of ρ if and only if $u' = au + b$ and $m' = m/a$ for $a > 0$.*

B. *Identification and Out-of-Sample Prediction*

We examine what can be inferred about the primitives of the model based on observed choices. This is important for understanding the underlying model and its predicted behavior, as well as for making out-of-sample predictions. We consider an analyst who observes stochastic choice data. The analyst posits that the data are generated by the weighted linear model. The analyst would like to answer what are the utility and the salience parameters for each alternative. We show in this section how this question can be answered within the framework of our model. Moreover, we would like to illustrate that we can make out-of-sample predictions (outside $\mathcal{D}$) given that our representation is unique.[9] This illustration will also help the reader to understand the Proof of Theorem 1 better.

[9] The parameters of the models are derived from the domain $\mathcal{D}$.

Consider four alternatives $X \equiv \{x,y,z,t\}$, and suppose that $\mathcal{D}$ consists of all sets containing at most three elements. Imagine the choice probabilities from binary and ternary sets satisfy the following conditions:

- Choices from pairs are equiprobable: For all $a,b \in X$, $\rho(a,b) = 0.50$.
- For any triple, if y,z,t are members of the triple, then they are chosen with equal probabilities.
- x is chosen with probability 0.30 from any triple.

These are our choice data on $\mathcal{D}$. Since $X \notin \mathcal{D}$, choices from the entire set X are not observed. Our goal is (i) to identify u,m values for each and (ii) to predict choice probabilities from X.

When two alternatives, a and b, are both chosen with positive probability from two sets, and the probability that they are chosen differs in both sets, then it becomes easy to identify the ratio of their salience parameters: $m(a)/m(b) = r_{S,T}(a,b)$, where both S and T include a and b. For example, we have

$$\frac{m(x)}{m(y)} = \frac{\rho(x|\{x,y\}) - \rho(x|\{x,y,z\})}{\rho(y|\{x,y\}) - \rho(y|\{x,y,z\})} = \frac{4}{3}.$$

So, by symmetry among y,z,t, we may conclude directly that $m(y) = m(z) = m(t) = (3/4)m(x)$. We may normalize up to scale, so let us suppose that $m(x) = 4$ and that $m(y) = m(z) = m(t) = 3$. Once m is identified up to scale, we can use the equality

$$u(a) - u(b) = \frac{\rho(a|S)}{m(a)} - \frac{\rho(b|S)}{m(b)}$$

to identify u. Since we may normalize u up to translation, this allows us to choose $u(x) = 0$. In so doing, it becomes apparent that $u(y) = u(z) = u(t) = 1/24$. This is the full identification of our model. With these identifications in hand, we may directly conclude that $\rho(x|\{x,y,z,t\}) = 5/26$, whereas $\rho(y|\{x,y,z,t\}) = \rho(z|\{x,y,z,t\}) = \rho(t|\{x,y,z,t\}) = 7/26$, thus affording an out-of-sample prediction.

Even outside of this particular situation, our approach allows for a very transparent identification of the two parameters. The key function introduced above, $r_{S,T}(x,y)$, identifies m up to a scale factor: $r_{S,T}(x,y) = m(x)/m(y)$. Given our identified m's, we can then identify the ranking of u's by defining $u(x) - u(y) = \frac{1}{m(x)}\rho(x|S) - \frac{1}{m(y)}\rho(y|S)$.

Our identification result is based on choice set variations observed in the data. In the Supplemental Appendix, we discuss how our model can be easily identified in choice settings with a fixed choice set (no choice set variation) but with observable attributes of outcomes. In particular, we can show that one can identify the parameters via solving a set of simple linear equations. This identification will improve the applicability of our model in different environments.

III. Relation to APU and RUM

A. *Relation to APU*

Our formulation as a solution to maximizing expected utility minus costs in (2) brings to mind the additive perturbed utility model (Fudenberg, Iijima, and Strzalecki 2015b). APU is a random choice model for which

$$\rho(x|S) \equiv \arg\max_{p \in \mathcal{S}} \sum_{x \in S} \left[u(x)p(x) - k(p(x))\right],$$

where k is some strictly convex and smooth function. Although this formulation is similar to our approach, the distinction is twofold:[10]

- In APU, the cost of probability is independent of the alternative in question; that is, k depends only on the probability of choosing each x but does not vary across alternatives.
- In APU, the cost function k need not be quadratic.

In terms of the generated choice behavior, APU and WL have a nontrivial intersection. For example, taking k to be quadratic in the APU is equivalent to assuming a constant salience function m in the WL. That is exactly the situation where we recover simple linear demands.

Another example of intersection is that APU and WL both nest the classic logit model. We note, however, that logit arises in two very different ways. APU becomes logit when the cost $k(p)$ is equal to a multiple of $p \log p$, and the choice probability of each option x is proportional to the exponential of its utility $e^{u(x)}$. In contrast, we saw that the WL produces logit choice probabilities if and only if utility is constant and ties are broken by salience, whereby the choice probability of each option x is proportional to its salience $m(x)$.

Despite having a nontrivial intersection, the APU and WL models are logically independent. In Supplemental Appendix D, we show that WL allows more flexible patterns of choice in binary comparisons. APU, like logit, satisfies a strong form of transitivity of binary comparisons, while the WL relaxes it to a moderate form of transitivity. On the other hand, Fudenberg, Iijima, and Strzalecki (2015b) show the APU can violate the Block-Marschak inequalities that characterize the random utility model (Barbera and Pattanaik 1986; Falmagne 1978). We now show that the WL is a RUM; hence, despite being the solution to the individual maximization problem (2), perhaps surprisingly, the WL also represents the choices of a heterogeneous population of standard rational individuals. In that sense, our model provides some additional flexibility without sacrificing the rationality assumed in the predominant paradigm of applied discrete choice estimation.

[10] In their working paper, Fudenberg, Iijima, and Strzalecki (2015a) weaken the first condition and consider the more general model,

$$\rho(x|S) \equiv \arg\max_{p \in \mathcal{S}} \sum_{x \in S} \left[u(x)p(x) - k(p(x), x)\right],$$

which nests our approach. Unlike our model, this model does not have a closed-form solution.

B. *Relation to RUM*

The most well-known generalization of the Luce model is the random utility model. In order to define the class of RUMs, first, let $\mathcal{R}$ be the set of all possible linear orders (rankings) on X and π be a probability distribution over rankings. $\pi(\succ)$ represents the probability of $\succ$ being realized as the preference. Given a set of available alternatives A, the probability of an alternative x being chosen is determined by the probability of a ranking for which x is at the top of A. Let $\mathcal{R}(a,A)$ be the set of rankings of X, which rank a at the top of A; that is, $\mathcal{R}(a,A) := \{\succ \in \mathcal{R}: a \succ b \text{ for all } b \in A\backslash a\}$. The RUM stochastic choice associated with π is ρ_π defined by

$$\rho_\pi(a|A) = \sum_{\succ \in \mathcal{R}(a,A)} \pi(\succ).$$

Our model is a RUM. Thus, although it is more general than Luce, it still fits within the most classic paradigm of random choice.

PROPOSITION 3: *If ρ has a WL representation, then it is a RUM.*

Let (u,m) be a WL representation of ρ. We now show how to construct a RUM representation for this ρ. First, we normalize u by subtracting $\Lambda(X)$, the shadow price for the grand set. Theorem 2 implies that $(\tilde{u},m)$ is also a WL representation of ρ, where $\tilde{u} = u - \Lambda(X)$. With this normalization, we have $\rho(x|X) = \tilde{u}(x)m(x)$ for each x, and therefore, $\sum_{x\in X}[\tilde{u}(x)m(x)] = 1$.

Next, to each ranking of alternatives, with the n alternatives enumerated as $x_1 \succ x_2 \succ \cdots \succ x_n$, our RUM representation assigns the following probability:

$$\pi(\succ) = \tilde{u}(x_1)m(x_1)\frac{m(x_2)}{m(X\backslash\{x_1\})}\frac{m(x_3)}{m(X\backslash\{x_1,x_2\})}\cdots\frac{m(x_n)}{m(x_n)}. \tag{3}$$

This construction mimics the construction of Theorem III.6 in Block and Marschak (1959). It is then standard to show that $\rho = \rho_\pi$.[11]

Rank Polarization.—Equation (3) shows that the WL model is a special (restricted) case of RUM. However, we now show the WL retains enough flexibility to allow a full range of variation in the extremeness of tastes in the RUM model. For a given RUM π, define the *rank polarization* of each option $x \in X$, denoted by $P(x,\pi)$, as the sum of the proportion of consumers who rank option x as their first (best) option and as their last (worst) option among all the options. Formally,

$$P(x,\pi) := \pi\{\succ \in \mathcal{R}: x \succ y \text{ for all } y \neq x\}$$
$$+ \pi\{\succ \in \mathcal{R}: y \succ x \text{ for all } y \neq x\}.$$

[11] This representation is not unique in general. See Turansick (2022) for details.

Since Block and Marschak (1959), we know that the full distribution π from a RUM model cannot be identified from choice behavior. However, the probability that an option x is ranked in the k-th position is identified. In particular, whenever $\rho = \rho_\pi$ is a RUM, rank polarization is uniquely recovered from choice as follows:

$$P(x,\pi) = \rho(x|X) + \sum_{A:x\in A} (-1)^{|A\setminus\{x\}|}\rho(x|A).$$

In particular, while the RUM representation is not unique, we must have $P(x,\pi) = P(x,\pi')$ whenever $\rho_\pi = \rho_{\pi'}$. Hence, for a given RUM ρ, the rank polarization measure is well defined; we may write $P(x,\rho)$ instead of $P(x,\pi)$ whenever ρ belongs to RUM.

By construction, the rank polarization of x may take any value in the range

$$\rho(x|X) \leq P(x,\pi) \leq 1, \tag{4}$$

where $\rho(x|X)$ is the market share of option x in X. For example, when $\rho(x|X) = 1/2$, one-half of the population ranks x as its first (best) option; and $P(x,\pi)$ can vary from $1/2$ to 1 according to how much probability π assigns to rankings where x is the worst (last) option. Equation (3) demonstrates that, by varying the trade-off between the utility $u(x)$ and the salience $m(x)$ of an option x, the WL model can generate the full range of rank polarization in (4). This flexibility allows the WL to accommodate several empirical regularities (Section IV) and to perform surprisingly well in out-of-sample prediction (Section VI), despite having fewer parameters than other commonly used models.

Number of Parameters.—The RUM is the reigning paradigm for discrete choice estimation in applied work. However, in its most general formulation, the RUM can sometimes prove too unruly for practical use. With n alternatives, there are $n!$ possible rankings (or types) in a population. A probability distribution over types thus has $n! - 1$ free parameters, and these cannot be identified from stochastic choice. In fact, the RUM is overparameterized, in that $n! - 1$ is larger than the effective number of market shares that need to be explained in a random choice rule (as shown by the comparison of "RUM" and "Data" in Figure 1). In applications, practitioners often employ special cases of RUM that (i) facilitate identification, (ii) provide more tractable and interpretable functional forms, and (iii) provide sharper out-of-sample predictions.

Figure 1 compares the number of free parameters in the WL model to some of the best-known special cases of RUM in the literature. Each one of these special cases of RUM attempts to provide a good balance between generality and practical use. In comparing the number of parameters of each model, it is important to keep in mind that, due to nonlinearities in the mapping from parameters to market shares, having additional parameters does not necessarily mean that a model yields useful additional flexibility. The classic logit model is the most extreme example of tractability and simplicity, with $n - 1$ free parameters. More general models like nested logit and covariance probit provide more flexibility at the expense of having more parameters,

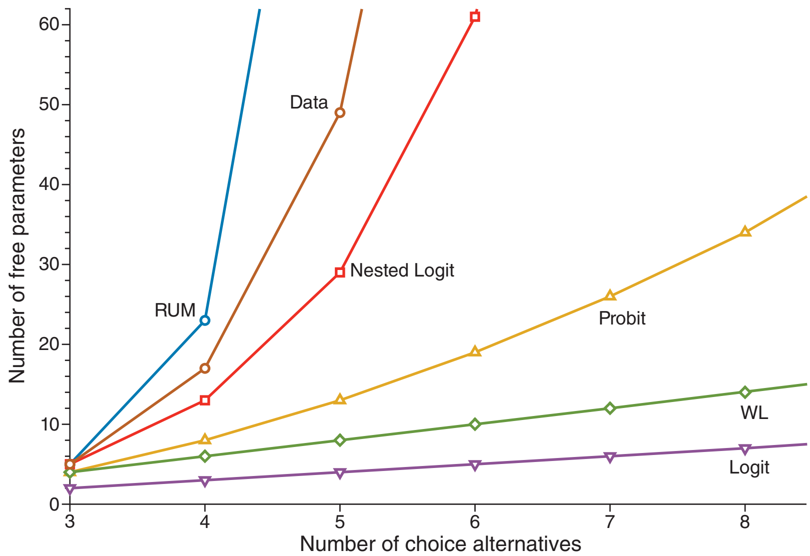

Figure 1. Comparing the WL to Other Special Cases of RUM: Number of Free Parameters as a Function of the Number of Choice Alternatives.

being less tractable, requiring more data for identification, and producing less sharp predictions out of sample.

Compared to some of these alternatives, the WL model is closer in terms of simplicity and tractability to the classic logit model. In particular, WL, like logit, has a closed-form formula for choice probabilities, and its number of parameters increases linearly with the number of alternatives. In the next section, we show the WL adds enough flexibility to accommodate a wide range of relevant empirical phenomena; in Section VI, we also show it performs extremely well in out-of-sample prediction.

C. *Relation to Other Models*

Finally, we discuss the relationship with other models. First, we consider the rational inattention model (RI) of Matejka and McKay (2015), which makes an important connection between the optimal choice of information and stochastic choice. While their state-dependent choice probabilities resemble the Luce formula, RI violates RUM. Moreover, as opposed to WL, RI does not have closed-form expression for state-independent choice probabilities and requires much richer data. The intersection between WL and RI is nonempty, however, as both models nest the classic logit model: RI becomes the classic logit in the special case of a symmetric prior, and WL nests logit in the special case of constant utility, as we have shown in Section I.

Cattaneo et al. (2020) characterizes a class of stochastic choice rules referred to as the random attention model (RAM). RAM attributes randomness in choice to attention for fixed preferences. As in RI, RAM model is more general than RUM, and it is not uniquely identified. Two parametric models on limited attention are introduced by Manzini and Mariotti (2014) and Brady and Rehbeck (2016). It is routine to show that WL is distinct from these two models.

Finally, the model of Ellis and Masatlioglu (2022) provides a characterization for a general class of deterministic models of categorization. This model generalizes the salience model of Bordalo, Gennaioli, and Shleifer (2013). These models are based on observable attributes and are deterministic. As far as we are aware, there is no stochastic version of such models.

IV. Explaining Empirical Patterns

We now turn to demonstrating the explanatory power of our model for stylized empirical facts about random choice and market demand. We highlight the explanatory power of our model relative to not just the multinomial logit approach but also many other widely used discrete choice models.

A. *Substitution Patterns*

A key question confronting models of market demand is their predictions regarding cross-outcome substitution patterns. We next turn to describing substitution patterns in our model, looking at both utility (including price) and salience changes.

Consider what happens to the demand for product x in the choice set S as the utility of another available product y increases. Not unexpectedly, the demand for x will decrease. Moreover, it decreases in a linear fashion, that is, in proportion to the ratio of $m(x)m(y)$ and $m(S)$. In other words, the more salient either x or y, the more impactful is a change in $u(y)$ on $\rho(x|S)$. This should not be surprising; changes in the utility of more salient items impact the demand for other items more, and a salient item (which is often thought about with other objects) also is impacted more by shifts in the value of competing items. Moreover, we have symmetry as

$$\frac{\partial \rho(x|S)}{\partial u(y)} = \frac{\partial \rho(y|S)}{\partial u(x)} = -\frac{m(x)m(y)}{m(S)},$$

analogous to the Slutsky substitution patterns.[12] We do not have symmetric responses for cross-m changes:

$$\frac{\partial \rho(x|S)}{\partial m(y)} = -\frac{m(x)}{m(y)m(S)}\rho(y|S).$$

Again, this should be relatively intuitive. The effect of changing y's salience should depend on the choice probability of y, $\rho(y|S)$, since items chosen with higher probability benefit more from increased salience.

These comparative statics extend in a straightforward manner to analyze the substitution patterns arising from changes in an attribute of an option x, such as price, that enter utility or salience. It is well-known that, in the multinomial logit model,

[12] Positive semidefiniteness follows as well because this matrix of substitution patterns is obviously diagonally dominant, with positive diagonal elements.

the cross-price elasticity of item x for item y does not depend on x, a condition that is at odds with both intuition and reality: We would expect cross-price effects to be much more variable. Like other alternative models (such as nested logit), our model allows for more flexible patterns than logit.

To see this, suppose that price enters utility linearly, $u(x) = \tilde{u}(x) - p(x)$, and let $m(x)$, $\tilde{u}(x)$, and the set of products, S, be fixed. Then, the cross-price elasticity of x for y is

$$\epsilon_{x,y} = \frac{m(x)}{\rho(x|S)} \frac{p(y)m(y)}{m(S)}.$$

Thus, the WL model allows for any given pair of items x and y to have a distinct cross-price elasticity, captured in a simple and easy-to-use formula. The cross-price elasticities of x to y are increasing in the price of y as well as the salience of both x and y but decreasing in the existing market share of x.

That said, our model still imposes restrictions on relationships between pairs of cross-price elasticities. In particular, the ratio of $\epsilon_{x,y}$ to $\epsilon_{x,z}$ is equal to the ratio $p(y)m(y)/[p(z)m(z)]$ and so is independent of x. We believe there is a natural interpretation of this condition. Suppose y is more salient than z. In this case, when the price of y shifts, it is more noticeable (than if the price of z shifts) and so causes greater changes in the demand for x. Of course, because y is more salient than z, it also means that changes in the price of y will also impact the demand for w more than z. Thus, increased saliency leads to an increase in cross-price substitution patterns. Similarly, the ratio of $\epsilon_{x,y}$ to $\epsilon_{z,y}$ is independent of y (with a similar interpretation). Moreover, the ratio of elasticities is independent of choice sets (S); in other words, the relative impact of y on x compared to z on x does not depend on the availability of other products.

We can use existing approaches to allow for additional flexibility in our substitution patterns. Random coefficients models, such as mixed logit, generate more realistic patterns of substitution than logit by incorporating additional information provided by a vector of observable characteristics for each object $\mathbf{x} = (x_1 \ldots, x_n)$. They usually specify a convenient utility function, such as $u(\mathbf{x}) = e^{\beta'\mathbf{x}}$, where β are randomly distributed coefficients. This formulation yields "local" substitution patterns, in the sense that changes in an attribute x_i of an object $\mathbf{x}$ will impact more intensely the market share of objects that are "close" to $\mathbf{x}$ in the space of observable characteristics. Given its tractable closed-form expression for choice probabilities, the WL model can be deployed in a setting with observable attributes and random coefficients in the same way, with the added flexibility of including a salience function, in addition to utility. The identification of the WL model with attributes is straightforward (see Supplemental Appendix C).

B. *Market Share Persistence*

In Section III, we defined rank polarization $P(x, \pi)$ as the sum of the proportions of consumer types that rank x as their first (best) and last (worst) option in the

RUM model π. Rank polarization has a clear empirical counterpart: Higher rank polarization for an option x translates into more persistence in the size of the market share for x as the set of options changes.

To see this, consider a set with three options $X = \{x, y, z\}$, and let ρ be the behavior of a population of consumers described by a RUM π. Then, the rank polarization for x can be obtained from market share data by

$$P(x, \pi) = 1 - \left[\rho(x, y) - \rho(x | \{x, y, z\})\right] - \left[\rho(x, z) - \rho(x | \{x, y, z\})\right].$$

The first term in brackets is the change in x's market share when z is removed from X; it is precisely the proportion of consumers who rank $z \succ x \succ y$. Likewise, the second term in brackets is the proportion of consumers who rank $y \succ x \succ z$. Hence, when the market share for x is persistent across sets, the terms in brackets are close to zero, and the rank polarization $P(x, \pi)$ is close to one. Thus, the equation above shows the polarization of tastes for x is directly related to the size of the market share that x gains (loses) when other options are removed (introduced).

Now consider a population where one-half of the consumers rank option x first, and the other half ranks x last. This implies that x is always chosen one-half of the time from any doubleton, e.g., $\rho(x, y) = 1/2$. However, since nobody in the population ranks option x between y and z, it retains the same market share throughout; that is,

$$\rho(x, y) = \rho(x | \{x, y, z\}) = \rho(x, z) = \frac{1}{2}.$$

This kind of market share persistence is illustrated in the famous "red bus, blue bus" example (Debreu 1960). Two distinct products, x (a car) and y (a blue bus), receive equal market share in a binary comparison. Introducing option z (a red bus) that closely resembles y does not alter the likelihood of choosing x.

Logit and other models based on a single-dimensional utility measure cannot generate nor approximate persistent market share behavior. In the case of both logit and the APU model, for example, the binary choices $\rho(x, y) = \rho(y, z) = 1/2$ immediately imply that each option must be chosen with probability $1/3$ from the large set. These models' difficulty in accommodating persistent market shares is pervasive and also holds with a larger number of options.

The WL can accommodate market share persistence not only in the example above but also in larger choice sets. Let X be any finite set of options and fix any value of $\rho(x, X)$, the market share of x in the grand set X. Let (u, m) be a WL representation of ρ. By Theorem 2, we may subtract a constant from the utility function u, if necessary, so that $\Lambda(X) = 0$ and $\rho(x, X) = u(x)m(x)$. We adjust x so that it has, in the limit, the same market share for all choice sets. We multiply the initial utility $u(x)$ by a scaling factor $k > 1$ and divide the initial salience $m(x)$ by the same factor. Equation (3) shows that in the RUM representation the probability that x is ranked first (best) remains constant, while the probability that x is last (worst) increases. As $k \to \infty$, $\rho(x|S)$ converges to $u(x)m(x)$ in every set S as desired.

C. *Introduction of New Products*

Benkard and Bajari (2001) note that almost all discrete choice models in markets with large numbers of items predict that demand for any one item must converge to zero.[13] This includes the counterintuitive prediction that introducing a large number of inferior products will drive a superior product's market share to zero. In contrast, we now show that the WL model allows for nonnegligible market shares even as the number of products grows infinitely large. In particular, the WL model accommodates the notion of dominant and inferior products.

Suppose the initial set of options S contains a superior product x and an inferior product y, with $u(x) > u(y)$. Now let the set $T_n = \{y_1, \ldots, y_n\}$ consist of n replicas[14] of the inferior y, and consider what happens when the replicas in T_n enter the market. As the number n of replicas grows large, the market share of the superior x converges to

$$\lim_{n\to\infty} \rho(x|S \cup T_n) = m(x)\left[u(x) - u(y)\right] > 0.$$

This is easy to see from the WL formula: The first term in (1) vanishes when its denominator increases without bound, while in the second term, the weighted average utility $\bar{u}_m$ in the market converges to $u(y)$ as the number of replicas grows large.

V. Application: Markups and Advertising

We now embed the WL model into an equilibrium model of price setting and advertising and show it can naturally generate intuitive results around markups, advertising, and the relationship between them (see Rosen 1974 for a seminal paper looking at price setting and demand when products have multidimensional attributes). As discussed in Section I, evidence from marketing is typically interpreted to mean that firms can directly alter the salience of their brands via advertising (Sutherland and Galloway 1981; Moran 1990; Yi 1990; Miller and Berry 1998). We show that the WL model yields tractable closed-form equilibrium solutions, which match stylized facts from the literature.

We assume each firm i offers a single item and that the item has an underlying exogenous quality $\tilde{u}_i$. The firms can compete by setting prices, p_i, and the total utility a consumer gets from item i is then $u_i = \tilde{u}_i - p_i$. However, we also allow the firms to compete on another margin: They can adjust the salience of items. In line with our interpretation of m, we suppose that increasing m_i reduces the mental friction required to purchase i. This can occur because, e.g., advertising raises the salience of an item.[15] The flexibility embodied in our model via m allows us to easily capture these novel

[13] They show this is true (focusing on the case of where there is an outside good present) in the models where, in addition to mild technical conditions, the support of the error terms is continuous and unbounded above. This not only nests the multinomial logit approach but almost all other widely used approaches, such as nested logit and random coefficients.

[14] We use the term "replica" to mean that each y_i has the same salience and utility as y. This implies y_i is a replica of y in the behavioral sense of Faro (2023).

[15] As Bagwell (2007) points out, advertising has been seen by economists as having three approaches: persuasive, informative, and complementary. Our model is closest in spirit with the informative approach to advertising, where advertising helps raise "awareness" of a product.

considerations, like competition on salience, outside of the typical one-parameter random utility models.[16]

In order to endogenize entry in the market, we allow for n firms (we refer to the set of firms as N). We normalize the size of the market to 1. All firms face a marginal cost of production k. We will think of m_i as being proportional to the amount of advertising so that higher m_i corresponds to more advertising. We will assume that the cost of advertising m_i is $\gamma(m_i) = gm_i^2$, where g is the marginal cost of increasing advertising.

Given these assumptions, and letting $\Lambda = \frac{\sum_j m_j(\tilde{u}_j - p_j) - 1}{\sum_j m_j}$, the profit function for the firm is

$$(p_i - k)\left[m_i(\tilde{u}_i - p_i - \Lambda)\right] - gm_i^2 - hu_i^2.$$

We will focus on symmetric equilibria and so will suppose that exogenous variables are the same across all firms (i.e., that k and $\tilde{u}$ are the same for all firms) and that all firms play the same strategies (so that $p_i = p$ and $m_i = m$ for all i in equilibrium). We initially suppose that n firms exist in the market but will later consider what happens when there is entry and exit. We also suppose k is small enough so that an equilibrium with positive firms profits exists for a fixed n.[17]

To solve, we take the first-order conditions for p_i, $m_i(\tilde{u}_i - p_i - \Lambda) + M(p_i - k)m_i\left[-1 - \Lambda_p(N)\right] = 0$, and m_i, $(p_i - k)\left[(\tilde{u}_i - p_i - \Lambda) - m_i\Lambda_m\right] + 2gm_i = 0$. Given the symmetry assumption, $\Lambda = \tilde{u} - p - 1/mn$ and $\Lambda_p = -1/n$, $\Lambda_m = -1/(m^2n^2)$. Substituting back into the first-order conditions, we obtain the following proposition.[18]

PROPOSITION 4: *With n firms, the symmetric equilibrium has the following solution:* $p_i = k + \frac{1}{m_i(n-1)}$, $m_i = \frac{1}{(2gn^2)^{1/3}}$.

Our model has several interesting, and empirically relevant, implications. First, suppose that the number of firms, n, is exogenous. As n gets large, markups $p_i - k$ must go to zero, which is in contrast to the equilibrium in the multinomial logit model.[19] Thus, the WL captures the intuition that in markets with a large number of firms, no firm has market power.

Second, we can also use the results to understand firm entry. Suppose that firms face a fixed cost of entry f. Then firm profits are $\frac{g^{1/3}(n+1)}{2^{2/3}(n-1)n^{4/3}}$. As the cost of advertising grows (i.e., g increases), we see a larger number of firms (n increases). The increase in g and n jointly cause m_i to fall for all firms; implying that advertising falls. Because m_i falls and n increases, it is not immediately obvious what happens to markups. But algebra (using the fact m_i depends on n and g, and that the zero profit

[16] In fact, our model can also be extended to allow for endogenous choice of quality, $\tilde{u}$ as well, while still being able to generate closed-form solutions for the equilibrium.

[17] Therefore, in a symmetric equilibrium all products will be purchased with positive probability.

[18] The equilibrium p_i also characterizes the solution where m_i is exogenous.

[19] As Benkard and Bajari (2001) point out, this issue applies more generally to GEV and random coefficients logit models, although Probit models can avoid these implications.

condition allows us to write g as a function of m) shows that $m_i(n - 1)$ is falling in the equilibrium number of firms, implying price and markups increase.

Therefore, in equilibrium, we observe price and markups being positively correlated with the equilibrium number of firms n but negatively correlated with the degree of advertising (m_i). Intuitively, high advertising costs mean firms cannot use advertising as effectively to gain market share by increasing their salience. This in turn means that firms compete less with each other on price. This means profits increase, and more firms enter.

This result, a correlation between low prices and high levels of advertising, is reminiscent of results that emerge in completely different contexts in Robert and Stahl (1993) and Bagwell and Ramey (1994) (who find that advertising is greater when prices are lower) and which have been extensively investigated in the literature (see Bagwell 2007 for a relatively recent survey). Our result is also in line with the empirical evidence in Syverson (2019), which indicates that reductions in market frictions (i.e., increases in m_i) prompt customers to shift toward larger, lower-cost sellers, creating higher market concentration but lower markups.

The microfoundations of our model also allow us to understand the degree to which advertising changes welfare (as in a large literature beginning with Butters 1977). In our setting, advertising, by reducing the mental frictions involved in thinking about, and purchasing, a particular item, can generate consumer surplus (related mechanisms are developed in other papers, e.g., by Grossman and Shapiro 1984 in a horizontally differentiated market). Moreover, in equilibrium, because increases in advertising are associated with reductions in advertising costs, they are also correlated with reductions in markups, generating an additional channel of welfare gains. Thus, our model can directly link the amount of advertising in the market to consumer surplus via changes both in price and the mental costs associated with choice.

VI. Predicting Choice

In this section, we compare the predictions of the WL model to the predictions of three discrete choice workhorse models in the literature: the classic multinomial logit, the nested logit, and the covariance probit. These well-known models are described in any standard discrete choice estimation textbook (e.g., Train 2009), and we review them in detail in Supplemental Appendix F.

We estimate each model in 14,850,000 synthetic datasets aimed at covering the entire range of choice behavior generated by the RUM family. This is a natural domain to consider since RUM is often the de facto assumed paradigm in discrete choice estimation. Every model we use for estimation and prediction, including our own, is a special (restricted) case of RUM. Supplemental Appendix E contains a detailed description of our simulated data and the maximum likelihood estimation.

Our simulation setup has $n = 4$ choice alternatives, which yields a domain with binary, ternary, and quaternary choice problems, namely,

(i) six binary choice problems $\{1,2\}$, $\{1,3\}$, $\{1,4\}$, $\{2,3\}$, $\{2,4\}$, and $\{3,4\}$;
(ii) four ternary choice problems $\{1,2,3\}$, $\{1,2,4\}$, $\{1,3,4\}$, and $\{2,3,4\}$; and
(iii) one quaternary choice problem $\{1,2,3,4\}$.

For each simulated demand system, we compute the market shares in (i)–(iii) above, and we perform three leave-one-out prediction exercises:

- **Binary Choice Prediction:** Fit each model to the market shares observed in (ii) and (iii), and use the estimated model to predict the market shares in (i).
- **Ternary Choice Prediction:** Fit each model to the market shares observed in (i) and (iii); use the estimated model to predict market shares in (ii).
- **Quaternary Choice Prediction:** Fit each model to the market shares observed in (i) and (ii), and use the estimated model to predict market shares in (iii).

We sort the space of RUM data-generating processes by the level of rank polarization of two alternatives, which we label i and j throughout. For each fixed level of rank polarization $P(i,\pi)$ and $P(j,\pi)$, we compare the obtained empirical distributions of market share prediction errors across the models. We use two metrics to compare the market share prediction errors of each model. First, Figure 2 presents an ordinal comparison, showing the proportion of simulated datasets in which the WL model makes a smaller prediction error than each alternative model. Next, Figure 3 shows the median absolute prediction error for each model. In Supplemental Appendix G, we include additional comparisons using the root mean square error and the mean absolute error.

First, consider the direct comparison of WL and logit in the top row of Figure 2. The three panels are mostly blue, which means the WL model's market share prediction is closer to the true data-generating process than logit's prediction in more than 50 percent of the simulations with a given level of rank polarization. The WL's advantage over logit is most notable when rank polarization is either very high or very low, indicated by the darker blue color toward the corners of each panel. WL makes a better quaternary choice prediction in close to every simulation where both options i,j have high rank polarization and when both options have low rank polarization, shown by dark blue in the top-right panel of Figure 2. On the other hand, the WL's advantage tends to disappear toward the center of each panel, where the white color indicates a toss-up between logit and WL. A notable case where WL does worse than logit is for binary choice predictions in a small area close to the knife-edge case $P(i,\pi) = P(j,\pi) = 0.5$, indicated by the light yellow color in the top-left panel in Figure 2.

Figure 3 sheds further light into this comparison by plotting the mean absolute prediction errors for each model. The second row of Figure 3 shows that logit's median prediction errors increase significantly as rank polarization levels move away from 0.5. The large areas of red color indicate levels of rank polarization where logit's median prediction error surpasses 10 percentage points of market share.

Nested logit performs better than logit against the WL but not by a significant amount, and the blue color still predominates in the middle row of Figure 2. The improvement of nested logit over logit is most visible for quaternary predictions in the middle-right panel of Figure 2, and when the rank polarization of one option is fixed close to 0.5, while the other option has a rank polarization level of 0.5 or higher. The third row of Figure 3 shows the nested logit model has significantly smaller median absolute prediction errors than logit in those regions, though still larger than WL's. In particular, nested logit still exhibits areas of red color where the

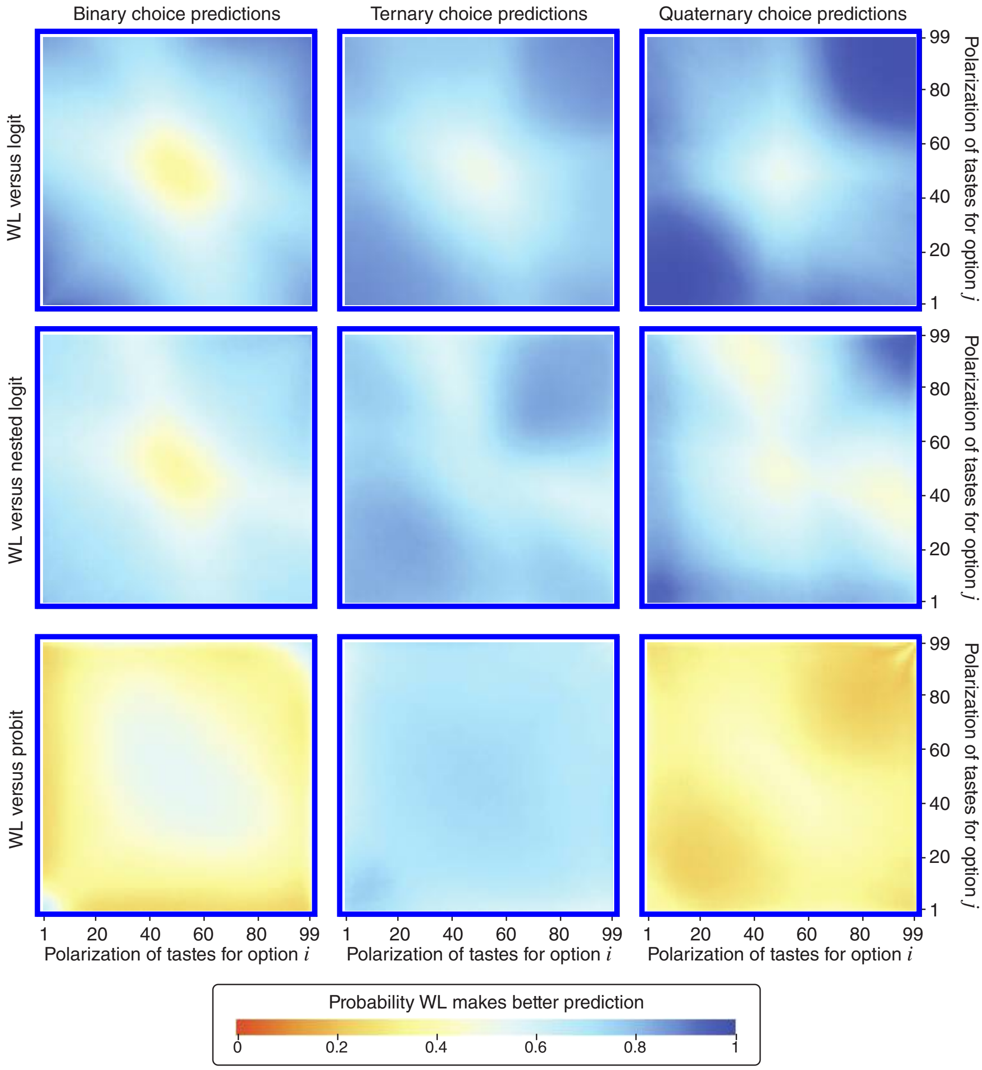

Figure 2. Leave-One-Out Prediction Comparison between the WL Model and Logit, Nested Logit, and Probit

Notes: The axes display the fixed level of rank polarization for two different options. The color scale represents the proportion of the simulated demand systems in which the WL model makes a closer market share prediction than the competing model.

median prediction error is larger than 10 percentage points of market share across the three prediction exercises.

Finally, consider the comparison between WL and probit in the bottom row of Figure 2. Probit is more likely to make a better prediction than WL across quarternary choice predictions, indicated by the yellow color in the bottom-right panel. On the other hand, WL is more likely to make a better prediction than probit across ternary predictions, indicated by the blue color in the bottom-middle panel of Figure 2. Binary choice predictions turn in favor mostly of probit, except toward the middle of the bottom-left panel where polarization levels are close to 0.5 and also for very high or very low polarization levels.

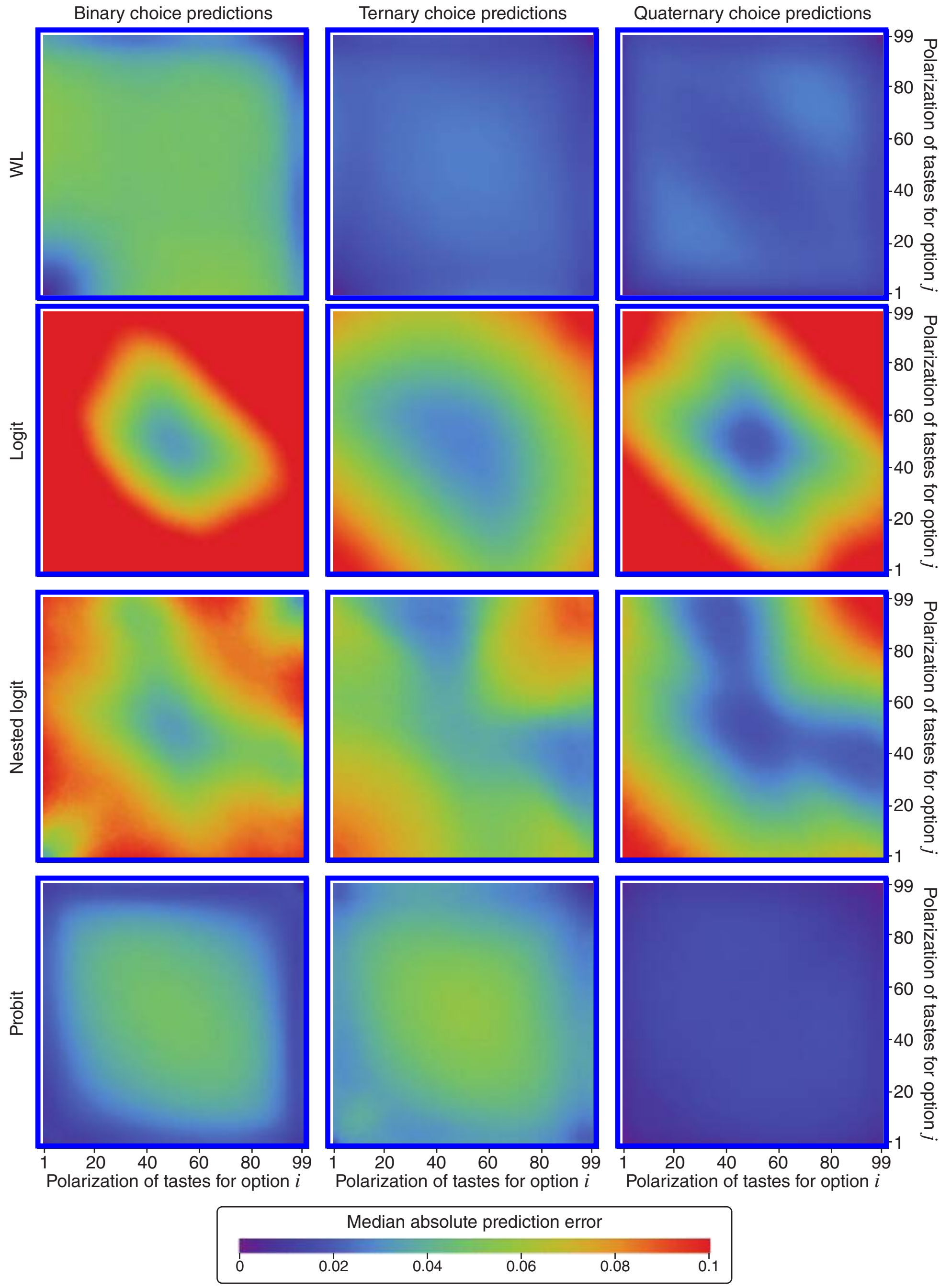

Figure 3. Median Absolute Prediction Error in Leave-One-Out Prediction for WL, Logit, Nested Logit, and Probit

Note: The axes display the fixed level of rank polarization for two different options.

Despite the large presence of yellow in the bottom row of Figure 2, Figure 3 shows that the difference in mean absolute prediction errors between WL and probit is small. The top row of Figure 3 shows WL has small prediction errors across ternary and quaternary predictions (dark blue in top-middle and top-right panels) and slightly larger prediction errors for binary choice prediction (green color in top-left panel), while probit has the lowest prediction errors for quaternary choice (darker blue in the bottom-right panel) while slightly higher errors for binary and ternary choice prediction (green color in the bottom-left and bottom-middle panels).

One advantage of using the WL over probit is its close-form choice probabilities and ease of computation. Estimating the parameters for WL, logit, and nested logit took roughly a week on a (vintage 2023) desktop computer. Estimating the covariance probit, however, took several weeks running in parallel in Washington University's RIS cluster computer and Amazon's Elastic Cloud Computing cluster. For Mathematica replication code, see Chambers et al. (2025).

In summary, Figures 2 and 3 show that the classic logit model becomes heavily disadvantaged once rank polarization moves away from 0.5. Nested logit improves prediction over logit somewhat, but the WL and the probit model clearly perform in a class of their own. While the WL does not clearly dominate the probit model, and performs slightly worse in some cases, it could prove a better model to use given its smaller number of parameters, closed-form choice probabilities, much lower computational burden, and easy interpretation.

VII. Conclusion

This paper has introduced a new model of stochastic choice: the weighted linear model of discrete choice. It represents a generalization of the classic model of Luce, and, as such, provides more explanatory power. The choice probabilities of any given product depend on two dimensions, the utility of an item and the salience of an item. The model sits at the intersection of the classic models of random utility and models of deliberate stochastic choice.

The weighted linear approach is closely related to well-known models in which demand is linear in the utility differences between products. The importance of salience for choice has been recognized by researchers (not just in economics but also in marketing, cognitive science, and psychology). Our model presents one way of incorporating such considerations. At the same time, it is quite tractable. The model lends itself naturally to describing consumers who experience some friction in being able to choose the best item (whether they be physical or attentional frictions). When used in models of strategic firm interactions, it can generate intuitive closed-form solutions that shed light on advertising, quality choice, and the number of firms serving a market.

Our hope is that our model can be useful to better understand market interactions between firms and consumers. In particular, the fact that our model allows for intuitive empirical patterns, such as flexible patterns of cross-price substitution patterns, or the existence of dominant market shares in large markets, can lead to new insights in many markets where these kinds of behavior need to be captured. We also think that empirical work geared toward understanding which kind of product attributes affect utility versus salience (e.g., does advertising increase the perceived utility of an item versus changing the cost of choosing it) could help shed useful insights into the structure of consumer choice.

REFERENCES

Ahumada, Alonso, and Levent Ülkü. 2018. "Luce Rule with Limited Consideration." *Mathematical Social Sciences* 93: 52–56.

Allen, Roy, and John Rehbeck. 2023. "Revealed Stochastic Choice with Attributes." *Economic Theory* 75: 91–112.

Armstrong, Mark, and John Vickers. 2015. "Which Demand Systems Can Be Generated by Discrete Choice?" *Journal of Economic Theory* 158 (A): 293–307.

Bagwell, Kyle. 2007. "The Economic Analysis of Advertising." In *Handbook of Industrial Organization*, Vol. 3, edited by Mark Armstrong and Robert Porter, 1701–1844. Elsevier.

Bagwell, Kyle, and Garey Ramey. 1994. "Coordination Economies, Advertising, and Search Behavior in Retail Markets." *American Economic Review* 84 (3): 498–517.

Barberá, Salvador, and Prasanta K. Pattanaik. 1986. "Falmagne and the Rationalizability of Stochastic Choices in Terms of Random Orderings." *Econometrica* 54 (3): 707–15.

Ben-Akiva, Moshe, and Steven R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand.* MIT Press.

Benkard, Lanier, and Patrick Bajari. 2001. "Discrete Choice Models as Structural Models of Demand: Some Economic Implications of Common Approaches." Unpublished.

Berry, Steven T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *RAND Journal of Economics* 242–62.

Berry, Steven, and Ariel Pakes. 2007. "The Pure Characteristics Demand Model." *International Economic Review* 48 (4): 1193–1225.

Bhatnagar, Roopali, and Jacob L. Orquin. 2022. "A Meta-analysis on the Effect of Visual Attention on Choice." *Journal of Experimental Psychology: General*. 151 (10): 2265–83.

Bialkova, Svetlana, and Hans C. M. van Trijp. 2011. "An Efficient Methodology for Assessing Attention to and Effect of Nutrition Information Displayed Front-of-Pack." *Food Quality and Preference* 22 (6): 592–601.

Block, H. D., and Jacob Marschak. 1959. "Random Orderings and Stochastic Theories of Response." Unpublished.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2013. "Salience and Consumer Choice." *Journal of Political Economy* 121 (5): 803–43.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2022. "Salience." *Annual Review of Economics* 14: 521–44.

Brady, Richard L., and John Rehbeck. 2016. "Menu-Dependent Stochastic Feasibility." *Econometrica* 84 (3): 1203–23.

Butters, Gerard R. 1977. "Equilibrium Distributions of Sales and Advertising Prices." *Review of Economic Studies* 44 (3): 465–91.

Cattaneo, Matias D., Xinwei Ma, Yusufcan Masatlioglu, and Elchin Suleymanov. 2020. "A Random Attention Model." *Journal of Political Economy* 128 (7): 2796–2836.

Cerreia-Vioglio, Simone, David Dillenberger, Pietro Ortoleva, and Gil Riella. 2019. "Deliberately Stochastic." *American Economic Review* 109 (7): 2425–45.

Chambers, Christopher P., Yusufcan Masatlioglu, Paulo Natenzon, and Collin Raymond. 2025. *Data and Code for: "Weighted Linear Discrete Choice."* Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. https://doi.org/10.3886/E214961V1.

Chandon, Pierre, J. Wesley Hutchinson, Eric T. Bradlow, and Scott H. Young. 2009. "Does In-Store Marketing Work? Effects of the Number and Position of Shelf Facings on Brand Attention and Evaluation at the Point of Purchase." *Journal of Marketing* 73 (6): 1–17.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review* 99 (4): 1145–77.

Choné, Philippe, and Laurent Linnemer. 2020. "Linear Demand Systems for Differentiated Goods: Overview and User's Guide." *International Journal of Industrial Organization* 73: 102663.

Dai, Jane, Jeremy Cone, and Jeff Moher. 2020. "Perceptual Salience Influences Food Choices Independently of Health and Taste Preferences." *Cognitive Research: Principles and Implications* 5 (1): 1–13.

Debreu, Gerard. 1960. Review of *Individual Choice Behavior: A Theoretical Analysis,* by R. Duncan Luce. *American Economic Review* 50 (1): 186–88.

Dixit, Avinash K., and Joseph E. Stiglitz. 1977. "Monopolistic Competition and Optimum Product Diversity." *American Economic Review* 67 (3): 297–308.

Echenique, Federico, and Kota Saito. 2019. "General Luce Model." *Economic Theory* 68: 811–26.

Ehrenberg, Andrew, Neil Barnard, and John Scriven. 1997. "Differentiation or Salience." *Journal of Advertising Research* 37 (6): 7–15.

Ellis, Andrew, and Yusufcan Masatlioglu. 2022. "Choice with Endogenous Categorization." *Review of Economic Studies* 89 (1): 240–78

Falmagne, J. C. 1978. "A Representation Theorem for Finite Random Scale Systems." *Journal of Mathematical Psychology* 18 (1): 52–72.

Faro, José Heleno. 2023. "The Luce Model with Replicas." *Journal of Economic Theory* 208: 105596.

Fechner, Gustav T. 1860. *Elemente der Psychophysik*. Vol. 2. Breitkopf & Härtel.

Fudenberg, Drew, Ryota Iijima, and Tomasz Strzalecki. 2015a. "Stochastic Choice and Revealed Perturbed Utility." Unpublished.

Fudenberg, Drew, Ryota Iijima, and Tomasz Strzalecki. 2015b. "Stochastic Choice and Revealed Perturbed Utility." *Econometrica* 83 (6): 2371–2409.

Gluth, Sebastian, Mikhail S. Spektor, and Jörg Rieskamp. 2018. "Value-Based Attentional Capture Affects Multi-alternative Decision Making." *Elife*: e39659.

Grossman, Gene M., and Carl Shapiro. 1984. "Informative Advertising with Differentiated Products." *Review of Economic Studies* 51 (1): 63–81.

Halff, Henry M. 1976. "Choice Theories for Differentially Comparable Alternatives." *Journal of Mathematical Psychology* 14 (3): 244–46.

He, Junnan, and Paulo Natenzon. 2024. "Moderate Utility." *American Economic Review: Insights* 6 (2): 176–95.

Horan, Sean. 2021. "Stochastic Semi-orders." *Journal of Economic Theory* 192: 105171.

Hoyer, Wayne D., and Steven P. Brown. 1990. "Effects of Brand Awareness on Choice for a Common, Repeat-Purchase Product." *Journal of Consumer Research* 17 (2): 141–48.

Jaffe, Sonia, and Scott Duke Kominers. 2012. "Discrete Choice Cannot Generate Demand That Is Additively Separable in Own Price." *Economics Letters* 116 (1): 129–32.

Jaffe, Sonia, and E. Glen Weyl. 2010. "Linear Demand Systems Are Inconsistent with Discrete Choice." *BE Journal of Theoretical Economics* 10 (1): 0000102202193517041729.

Janiszewski, Chris, Andrew Kuo, and Nader T. Tavassoli. 2013. "The Influence of Selective Attention and Inattention to Products on Subsequent Choice." *Journal of Consumer Research* 39 (6): 1258–74.

Kovach, Matthew, and Gerelt Tserenjigmid. 2022. "Behavioral Foundations of Nested Stochastic Choice and Nested Logit." *Journal of Political Economy* 130 (9): 2411–61.

Krajbich, Ian. 2019. "Accounting for Attention in Sequential Sampling Models of Decision Making." *Current Opinion in Psychology* 29: 6–11.

Krajbich, Ian, Carrie Armel, and Antonio Rangel. 2010. "Visual Fixations and the Computation and Comparison of Value in Simple Choice." *Nature Neuroscience* 13: 1292–98.

Krajbich, Ian, and Antonio Rangel. 2011. "Multialternative Drift-Diffusion Model Predicts the Relationship between Visual Fixations and Choice in Value-Based Decisions." *PNAS* 108 (33): 13852–57.

Li, Xiaomin, and Colin F. Camerer. 2022. "Predictable Effects of Visual Salience in Experimental Decisions and Games." *Quarterly Journal of Economics* 137 (3): 1849–1900.

Luce, R. Duncan. 1959. *Individual Choice Behavior*. John Wiley & Sons.

Macdonald, Emma K., and Byron M. Sharp. 2000. "Brand Awareness Effects on Consumer Decision Making for a Common, Repeat Purchase Product: A Replication." *Journal of Business Research* 48 (1): 5–15.

Machina, Mark J. 1985. "Stochastic Choice Functions Generated from Deterministic Preferences over Lotteries." *Economic Journal* 95 (379): 575–94.

Mangasarian, Olvi L. 1994. *Nonlinear Programming*. SIAM.

Manzini, Paola, and Marco Mariotti. 2014. "Stochastic Choice and Consideration Sets." *Econometrica* 82 (3): 1153–76.

Matějka, Filip, and Alisdair McKay. 2015. "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model." *American Economic Review* 105 (1): 272–98.

Mellers, Barbara A., Shi-jie Chang, Michael H. Birnbaum, and Lisa D. Ordóñez. 1992. "Preferences, Prices, and Ratings in Risky Decision Making." *Journal of Experimental Psychology: Human Perception and Performance* 18 (2): 347–61.

Miller, S., and L. Berry. 1998. "Brand Salience versus Brand Image: Two Theories of Advertising Effectiveness." *Journal of Advertising Research* 38 (5): 77–78.

Milosavljevic, Milica, Vidhya Navalpakkam, Christof Koch, and Antonio Rangel. 2012. "Relative Visual Saliency Differences Induce Sizable Bias in Consumer Choice." *Journal of Consumer Psychology* 22 (1): 67–74.

Moran, W. T. 1990. "Brand Presence and the Perceptual Frame." *Journal of Advertising Research* 30 (5): 9–16.

Navalpakkam, Vidhya, Ravi Kumar, Lihong Li, and D. Sivakumar. 2012. "Attention and Selection in Online Choice Tasks." *International Conference on User Modeling, Adaptation, and Personalization*: 200–211.

Nedungadi, Prakash. 1990. "Recall and Consumer Consideration Sets: Influencing Choice without Altering Brand Evaluations." *Journal of Consumer Research* 17 (3): 263–76.

Orquin, Jacob L., and Carl Johan Lagerkvist. 2015. "Effects of Salience Are Both Short- and Long-Lived." *Acta Psychologica* 160: 69–76.

Orquin, Jacob L., and Simone Mueller Loose. 2013. "Attention and Choice: A Review on Eye Movements in Decision Making." *Acta Psychologica* 144 (1): 190–206.

Peschel, Anne O., Jacob L. Orquin, and Simone Mueller Loose. 2019. "Increasing Consumers' Attention Capture and Food Choice through Bottom-Up Effects." *Appetite* 132: 1–7.

Reutskaja, Elena, Rosemarie Nagel, Colin F. Camerer, and Antonio Rangel. 2011. "Search Dynamics in Consumer Choice under Time Pressure: An Eye-Tracking Study." *American Economic Review* 101 (2): 900–926.

Rieskamp, Jörg, Jerome R. Busemeyer, and Barbara A. Mellers. 2006. "Extending the Bounds of Rationality: Evidence and Theories of Preferential Choice." *Journal of Economic Literature* 44 (3): 631–61.

Robert, Jacques, and Dale O. Stahl II. 1993. "Informative Price Advertising in a Sequential Search Model." *Econometrica* 61 (3): 657–86.

Romaniuk, Jenni, and Byron Sharp. 2003. "Brand Salience and Customer Defection in Subscription Markets." *Journal of Marketing Management* 19 (1-2): 25–44.

Romaniuk, Jenni, and Byron Sharp. 2004. "Conceptualizing and Measuring Brand Salience." *Marketing Theory* 4 (4): 327–42.

Rosen, Sherwin. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy* 82 (1): 34–55.

Rosenthal, R. W. 1989. "A Bounded-Rationality Approach to the Study of Noncooperative Games." *International Journal of Game Theory* 18: 273–92.

Shubik, Martin, and Richard Levitan. 1980. *Market Structure and Behavior*. Harvard University Press.

Singh, Nirvikar, and Xavier Vives. 1984. "Price and Quantity Competition in a Differentiated Duopoly." *Rand Journal of Economics* 15 (4): 546–54.

Smith, Stephanie M., and Ian Krajbich. 2019. "Gaze Amplifies Value in Decision Making." *Psychological Science* 30 (1): 116–28.

Spence, Michael. 1976. "Product Selection, Fixed Costs, and Monopolistic Competition." *Review of Economic Studies* 43 (2): 217–35.

Sutherland, Max, and John Galloway. 1981. "Role of Advertising: Persuasion or Agenda Setting?" *Journal of Advertising Research* 21 (5): 25–29.

Syverson, Chad. 2019. "Macroeconomics and Market Power: Context, Implications, and Open Questions." *Journal of Economic Perspectives* 33 (3): 23–43.

Thomas, Armin W., Felix Molter, and Ian Krajbich. 2021. "Uncovering the Computational Mechanisms Underlying Many-Alternative Choice." *Elife*: e57012.

Thurstone, L. L. 1927. "A Law of Comparative Judgment." *Psychological Review* 34 (4): 273–86.

Towal, R. Blythe, Milica Mormann, and Christof Koch. 2013. "Simultaneous Modeling of Visual Saliency and Value Computation Improves Predictions of Economic Choice." *PNAS* 110 (40): E3858–67.

Train, Kenneth E. 2009. *Discrete Choice Methods with Simulation*. Cambridge University Press.

Turansick, Christopher. 2022. "Identification in the Random Utility Model." *Journal of Economic Theory* 203: 105489.

Van der Lans, Ralf, Rik Pieters, and Michel Wedel. 2008. "Research Note—Competitive Brand Salience." *Marketing Science* 27 (5): 922–31.

Vieceli, Julian, and Robin N. Shaw. 2010. "Brand Salience for Fast-Moving Consumer Goods: An Empirically Based Model." *Journal of Marketing Management* 26 (13-14): 1218–38.

Voorneveld, Mark. 2006. "Probabilistic Choice in Games: Properties of Rosenthal's t-Solutions." *International Journal of Game Theory* 34: 105–21.

Wedel, Michel, and Rik Pieters. 2000. "Eye Fixations on Advertisements and Memory for Brands: A Model and Findings." *Marketing Science* 19 (4): 297–312.

Wegge, L. L. 1968. "The Demand Curves from a Quadratic Utility Indicator." *Review of Economic Studies* 35 (2): 209–24.

Weingarten, Evan, and J. Wesley Hutchinson. 2017. "Perceptual and Cognitive Salience and Their Effects on Product Valuations." Unpublished.

Yang, Xiaozhi, Chris Retzler, Ian Krajbich, Roger Ratcliff, and Marios G. Philiastides. 2024. "Attention to Brand Labels Affects, and Is Affected by, Evaluations of Product Attractiveness." *Frontiers in Behavioral Economics* 2: 1274815.

Yi, Youjae. 1990. "The Effects of Contextual Priming in Print Advertisements." *Journal of Consumer Research* 17 (2): 215–22.

Zhang, Zhihao, ShichunWang, Maxwell Good, Siyana Hristova, Andrew S. Kayser, and Ming Hsu. 2021. "Retrieval-Constrained Valuation: Toward Prediction of Open-Ended Decisions." *PNAS* 118 (20): 2022685118.